Machine Learning for UAV-Aided ITS: A Review With Comparative Study

Akbar Telikani, *Graduate Student Member, IEEE*, Arupa Sarkar, Bo Du, and Jun Shen, *Senior Member, IEEE*

***Abstract*— Unmanned Aerial Vehicles (UAVs) have immense potential to enhance Intelligent Transport Systems (ITS) by aiding in real-time traffic monitoring, emergency response, and infrastructure inspection, leading to rich data collection, lower response times, and efficient urban mobility management. Machine learning (ML) is a crucial component in UAV-assisted ITS as it processes UAV-captured data in both the perception layer and decision layers of intelligent components for vehicle/pedestrian detection, trajectory optimization, and resource allocation. Importantly, the integration of UAVs and cutting-edge deep learning (DL) techniques is fostering an exciting synergy, equipping UAVs with unparalleled intelligence and autonomy, particularly, for the perception layer of UAVs. Despite these enhancements, their usefulness for detection and traffic extraction tasks remains largely unexplored. The contributions of this paper are divided into two main aspects: (1) UAVs in different ITS application scenarios that are empowered by ML technologies are reviewed. (2) A thorough survey aiming to explore a quantitative understanding of widely used DL models via a series of experiments and comparisons is presented. Four DL models, namely Convolution Neural Network (CNN), regions with CNN (R-CNN), Faster R-CNN, and You Only Look Once (YOLO)), in combination with different backbones, are designed and employed on five aerial datasets. Finally, we present a discussion of the remaining challenges and future works.**

***Index Terms*— Unmanned aerial vehicles (UAVs), intelligent transportation systems (ITS), machine learning, deep learning.**

I. Introduction

THE substantial rise in vehicle numbers in recent decades, despite infrastructure advancements, has rendered existing transportation solutions inadequate for burgeoning traffic issues. Intelligent Transportation Systems (ITS) aims to address escalating traffic challenges, boost traffic efficiency, and aid in the evolution of intelligent road networks [1]. A complete transportation system automation requires more than just automating vehicles. It entails automating various aspects like field support teams, traffic police, road surveys, and rescue teams. Achieving this automation can be facilitated by employing smart and dependable Unmanned Aerial Vehicles (UAVs) to efficiently automate these components [2]. The benefits of UAVs, such as 3D manoeuvrability, excellent mobility, autonomous functionality, and robust communication and processing abilities, make them pivotal in the ITS context [3].

UAVs enable real-time monitoring and data transmission in ITS. Recording videos and images from roads and transmitting the gathered data to ground centres or cloud centres are the initial responsibilities of UAVs herein [4]. Also, during emergency events when ground vehicles are unable to communicate directly with base stations, UAVs can serve as temporary base stations in wireless communication systems [5]. For example, in disastrous situations such as floods or earthquakes (e.g., Hurricane Irma in 2017 in US [6]), UAVs can assist vehicles on highways by providing coverage in areas where vehicles are unable to access infrastructure-based Roadside Units (RSUs) [7]. In such critical situations, having access to road information, such as maps indicating unaffected highways and traffic congestion, can be immensely beneficial for people seeking to avoid disaster-stricken areas and for rescue teams to navigate effectively.

Moreover, real-time data analysis and decision-making performed by UAVs further empower ITS to respond promptly to changing conditions, optimize resources, and enhance overall transportation efficiency, safety, and effectiveness. Additionally, UAVs can serve as a valuable complementary solution to enhance road safety by detecting incident scenes, recognizing traffic violations, and identifying road hazards [2]. One of the advantages of using UAVs is their ability to reduce routing overhead in vehicle-to-vehicle communication, as nodes can directly communicate with these mobile RSUs. Furthermore, UAVs offer flexibility in the communication infrastructure by extending coverage to fill gaps or balancing network load in densely populated areas, all while being cost-effective [8].

Data analysis and information extraction contribute to the effective operation of ITS, especially in applications such as automated intersection management, traffic control, and autonomous cars [9]. Due to its capability to model intricate relationships in complex and dynamic network environments, machine learning (ML) has inherently received significant attention for improving ITS-related tasks or even challenging problems with UAVs [10]. Applying ML to UAVs facilitates intelligent service delivery by harnessing extensive data collected from ground vehicles, road infrastructures, or the UAVs themselves. For example, advanced sensor systems

Manuscript received 18 December 2023; revised 3 May 2024; accepted 24 June 2024. This work was supported by Australian Research Council Linkage Project under Grant LP230100083. The Associate Editor for this article was D. Sun. *(Corresponding author: Jun Shen.)*

Akbar Telikani, Arupa Sarkar, and Jun Shen are with the School of Computing and Information Technology, University of Wollongong, Wollongong, NSW 2522, Australia (e-mail: at952@uowmail.edu.au; as2019@uowmail.edu.au; jshen@uow.edu.au).

Bo Du is with the Department of Business Strategy and Innovation, Griffith University, Brisbane, QLD 4122, Australia (e-mail: bo.du@griffith.edu.au).

Digital Object Identifier 10.1109/TITS.2024.3422039

in autonomous (self-driving) vehicles, such as LiDAR and cameras, enable real-time data collection and sharing with UAVs for optimizing traffic flow and reducing congestion. This empowers ML algorithms deployed on UAVs to analyze traffic conditions dynamically and make decisions in real-time. The use of deep learning (DL) models in developing detection and classification algorithms for traffic conditions analysis has yielded attention [10]. On the other hand, ML techniques can be employed to improve the efficiency of UAVs in successful missions for monitoring and data collection. For example, the reinforcement learning (RL) approach can be used to control UAV trajectories with a set of actions (e.g., traveled distances) to provide coverage services [11].

A. Related Works

Literature in the UAV-aided ITS domain is limited to a few survey studies on the applications of UAVs in ITS. The survey presented in [12] reviewed research activities in the area of UAVs used for traffic monitoring presented by the year 2013. Menouar et al. [2] investigated the opportunities and challenges associated with UAV-enabled ITS in the context of smart cities. Much more recently, Outay et al. [13] conducted a review focusing on the application of UAVs in three distinct areas within the transportation sector: road safety, traffic monitoring and management, and highway infrastructure management. Gupta et al. [14] had an overview of 27 research papers published between 2015 and 2021 while also pinpointing some of the challenges ahead for the future of transportation systems. Saboor et al. [15] explored the integration of wireless technologies in UAV-supported ITS and highlighted open research questions in this domain. They also investigated key technological components of UAVs for ITS, such as edge/fog computing, artificial intelligence, and 5G/6G communications. A much newer review of DL-based approaches for object detection in traffic congestion analysis from aerial images/videos was provided in [16].

Some studies tried to compare the performance of ML models for drone data analysis in ITS applications. Liu et al. [17] evaluated the performance of four different models for vehicle detection: (1) the Faster Region-Based Convolutional Neural Network (Faster R-CNN), (2) the third release of You Only Look Once model (YOLOv3), (3) a Support Vector Machine (SVM) model using Histogram of Oriented Gradients (HOG) features, and (4) a visual background extractor (ViBe) algorithm [18]. Their finding proved that YOLOv3 and Faster R-CNN performed best as compared to the other two approaches. Additionally, despite better recall and precision for the Faster R-CNN than the YOLOv3, it required excessive hardware support. On the other hand, Benjdira et al. [19] compared Faster R-CNN and YOLOv3 for car detection and showed that YOLOv3 exhibited superior sensitivity and processing time compared to Faster R-CNN. Moreover, a comparison between Faster R-CNN, YOLOv3 and YOLOv4 on two datasets was provided in [20]. Through the application of diverse metrics, it was demonstrated that the performance difference between YOLOv4 and YOLOv3 on the datasets was statistically insignificant concerning average precision. Iftikhar et al. [16] compared one-stage detectors (YOLOv3, YOLOv4_Drone [21], and YOLOD [22]) and two-stage detectors (Faster R-CNN and Cascade R-CNN) for target detection based on UAVs. The findings demonstrated that YOLOv4_Drone outperformed other models. In [23], the comparison between Faster R-CNN, SSD, YOLOv3, and YOLOv4 on the AU-AIR dataset demonstrated that the YOLOv4 outperformed all the other models.

Despite existing review studies covering some aspects of UAVs in ITS, a systematic exploration of ML models applied to UAV-aided ITS has not yet been discussed in the literature. Furthermore, the existing comparative studies on UAV-aided ITS are limited with a narrow focus on specific models and metrics. They lack comprehensive model coverage, exhibiting limited evaluation metrics, dataset specificity, and insufficient exploration of hardware demands. Especially, none of the above studies provided a thorough survey of ML techniques for UAV-aided ITS from the perspective of the perception layer.

B. Our Contributions

To fill the above-mentioned gaps, we offer a detailed analysis of the intersections between ML and UAV-assisted ITS. Furthermore, this survey includes a comparative examination of popular DL techniques, illustrating the practical implementation of ML methodologies within the perception layer of UAV-assisted ITS for detection and classification tasks.

Hence, the main contributions of this paper lie in the following aspects:

- This research highlights the applications of ML techniques, such as conventional ML, deep learning, and reinforcement learning, in the context of UAV-aided ITS, with the conduction of a comprehensive examination focusing on various tasks such as vehicle, pedestrian, and cyclist detection, trajectory optimization, transmission throughput enhancement, and efficient resource management.
- This study examines a broad range of DL models, utilizing diverse datasets, incorporating a comprehensive set of metrics, assessing hardware considerations, and exploring models' adaptability to environmental changes and dynamic traffic scenarios. We explore various DL architectures by combining CNN, R-CNN, Faster R-CNN, and YOLO with different backbone architectures (e.g., ResNet, VGG16, and Inception) to investigate them on the perception layer of UAVs when being employed in ITS environments.
- Six commonly ITS-related tasks are investigated in this comparative study, including vehicle detection, vehicle classification, pedestrian detection, cyclist detection, speed estimation, and vehicle counting. Five benchmarking drone datasets are used for evaluation and the results are compared in terms of performance (i.e., recall, precision, F1-Score, false alarm rate) and efficiency (i.e., frame per second and inference time).
- Open problems and potential challenges in deploying ML techniques within UAV-assisted ITS are discussed, which

should be considered as future research directions to promote further development of UAV-assisted ITS, with insights gained from the above comparisons.

C. Paper Structure

The remainder of this study is organized as follows: Section II presents the background of the ML approaches that are being used in ITS, which are classified into three categories: clustering, deep learning, and reinforcement learning. The intersections between UAVs and ITS are explored and the potential contributions of ML techniques in improving UAV-aided ITS are investigated in Section III. Section IV provides an evaluation of DL models on different drone datasets for the perception layer of UAV-aided ITS. Open issues and future research directions are presented in Section V. Section VI concludes the paper.

II. Machine Learning

In the field concerned by this paper, ML algorithms analyze vast and complex datasets, enabling accurate traffic prediction, congestion management, and optimal routing. ML algorithms facilitate real-time decision-making, adjusting traffic signals, and controlling infrastructure to optimize traffic flow and reduce congestion by analyzing information gathered from different components. ML-powered predictive maintenance ensures the efficient operation of vehicles and transport infrastructure, minimizing downtime and enhancing safety [24]. Moreover, ML empowers advanced driver assistance systems and autonomous vehicles, enhancing navigation, collision avoidance, and adaptive cruise control. It also aids in anomaly detection, by identifying unusual traffic patterns and potential hazards. By utilizing historical data, ML enables trend analysis and long-term planning for infrastructure improvements and urban development [25].

A. Clustering

Clustering is a methodology for the grouping of data into clusters or collections based on the similarities among the features and characteristics of data points without prior knowledge of specific patterns within the data [26]. Clustering analysis techniques can be broadly categorized into two main groups: partitional and hierarchical. Partitional clustering methods segment a dataset into distinct groups by evaluating fitness measures through a set number of iterations. These algorithms, like K-Means, offer advantages in terms of simplicity and computational efficiency. These methods face two key issues: (1) the sensitivity to initial seed points, which can result in suboptimal outcomes, and (2) the need to decide the number of clusters before running the algorithm, which can be a challenging task. Hierarchical clustering employs a tree structure to illustrate relationships among cluster sets. It can be implemented through divisive (merging small clusters into larger ones) or cumulative (splitting large clusters into smaller ones) methods [27]. Hierarchical clustering is advantageous pre-specifying the number of clusters is not required. However, it has limitations, as each element can belong to only one cluster, making it less suitable for overlapping clusters [28].

B. Deep Learning

Deep learning, referred to as representation learning and originating from neural networks, employs multiple layers to create computational models that capture abstract features of the data [29]. Throughout these layers, transformations among neurons are facilitated using graph technologies. CNN stands out as a prominent network in the realm of deep learning and has demonstrated remarkable advancements, especially in computer vision applications [30]. A CNN model is composed of three fundamental layers: convolutional, pooling, and fully connected layers [31]. Feature maps are extracted using convolution kernels in convolution layers. Neurons in these maps are connected to neurons in the previous layer and apply a nonlinear activation function to the results. Computer vision successes, such as advancements in facial recognition, autonomous vehicles, self-checkout supermarkets, and intelligent medical treatments, powered by CNN, have revolutionized various aspects of modern life [30].

When employing a classic CNN model for detection problems in computer vision, it is tasked with processing multiple regions simultaneously, which leads to a significant increase in time complexity. To address this challenge, Region-based CNN (R-CNN) [32] uses a designated method to identify the object's location in an image by focusing on specific regions, known as proposal regions, which are generated through a greedy algorithm [33]. The process of generating region proposals in R-CNN and running CNNs on these proposals for object detection is computationally intensive, resulting in slower inference times compared to other object detection methods [34]. Fast R-CNN [34] incorporates the Region of Interest (ROI) pooling layer. Additionally, it introduced the concept of shared convolution features between object classification and bounding box regression tasks, resulting in increased speed and efficiency for the R-CNN network.

Fast R-CNN relies on the selective search algorithm or a similar external region proposal method to generate potential object regions within an image. This external region proposal step adds complexity to the overall process, and the region proposal step is not fully integrated into the network, limiting the potential for real-time applications. In tackling this challenge, Faster R-CNN, as described by Ren et al. [35], employs a region proposal network (RPN) to extract a RoI. Subsequently, an RoIPool layer is utilized to compute features from these proposals, facilitating the inference of bounding box coordinates and the classification of the object. The above-mentioned detection techniques, R-CNN [32], Fast R-CNN [34], and Faster R-CNN [35], fall into two-stage approaches. In this particular category, pre-selected region proposals are utilized, and subsequent object classification is performed by using CNNs.

Conversely, in a single-stage approach, the model directly outputs both the category probability and the positional coordinates of the objects. You Only Look Once (YOLO) [36] and its variants contribute significantly to this field. YOLO [36] partitions an image into multiple grid cells and discards those with the lowest probability of containing an object. Subsequently, the boxes with the highest probability of enclosing an object are merged to enhance the localization precision [33]. To handle the large number of localization errors introduced by the YOLO, YOLOv2 [37] incorporates anchor boxes rather than the fully-connected layer to improve bounding box predictions. YOLOv3 [38] replaces the mean squared error, logistic regression, and Darknet-53 in the YOLOv2 with the cross-entropy loss function, softmax function, and Darknet-19, respectively. YOLOv4 [39] uses multiple anchors for a single ground truth rather than one anchor utilized in the previous versions of YOLO. Moreover, it adapted Intersection over Union loss function.

C. Reinforcement Learning

Reinforcement learning (RL) is a machine learning approach for addressing problems involving decision-making in uncertain environments. In RL, an autonomous agent, whether a person, animal, robot, or software program, interacts with the environment to maximize cumulative rewards over time. These rewards are earned gradually through a sequence of actions [28]. To identify the optimal decision-making strategy or policy, the agent engages in exploration across pertinent state and action spaces, experimenting with different sequences of state-action-next state transitions. The average length of these sequences is referred to as the task horizon. In scenarios where the task has a prolonged horizon and encompasses extensive state and action spaces, the exploration space becomes expansive. This poses a challenge, as standard Reinforcement Learning (RL) algorithms may exhibit sub-optimal performance on tasks with extended horizons unless advanced exploration techniques are employed, as highlighted by Pateria et al. [40] recently.

III. Empowering Machine Learning in UAV-Assisted ITS

We organize research works on the UAV-aided ITS into four main categories: conventional ML models in III-A, DL models in III-B, RL models in III-C, and Hybrid models in III-D.

A. Conventional ML-Based Models

A typical detection strategy for vehicle detection is to extract features using, for example, bootstrapping, HOG, and KLT tracker, and then apply classifiers for vehicle detection and classification. The use of Bayesian networks in [41], [42], and [43], Support Vector Machine (SVM) in [44], [45], [46], [47], and [48], a combination of Fuzzy logic classifier and the Genetic Algorithm [49] are some example of this contribution. Garcia and Xiao [50] presented the multi-object measurement model characterising a drone camera as a direction-of-arrival sensor. Then, the Poisson multi-Bernoulli mixture filter, which is a Bayesian multi-object tracking algorithm, was used to estimate the set of vehicle trajectories.

The utilization of clustering methods assist in feature clustering and background removal, leading to more accurate and effective vehicle detection and tracking. Ke et al. [51] employed an adaptive DBSCAN algorithm to filter redundant lines extracted using modified Canny edge detection. The K-means method has been applied in [52], [53], [54], [55], and [56] to cluster features and separate vehicle- and background-related features. Choi and Yang [57] utilized the Mean-shift clustering algorithm to extract initial candidate vehicles and employed the geometric shape of vehicles to obtain the shape of vehicles.

Object detection algorithms are known for their robustness in handling image noise, background motions, and complex scenes, making them well-suited for vehicle detection in UAV videos. The Haar cascade classifier [58], a popular AdaBoost-based approach for vehicle detection as well as backgrounds, such as roads, trees, and pedestrians [59], [60], [61]. Findings presented in [62] the Haar cascade classifier is sensitive to object orientations and cannot work when the orientations of vehicles are unknown in UAV images. A simple strategy is to rotate each UAV image to align the roads with the horizontal direction. Then, the original Haar cascade classifier can be directly applied [63], [64], [65], [66], [67], [68].

The methods developed in [63], [66], and [67] first extracted the interest features and then employed an SVM classifier for vehicle detection and classification. In [63], Each video image was rotated nine times (each time 20 degrees) to cover 180 degrees. Leitloff et al. [64] used a combination of Adaboost and SVM for vehicle detection and shape-based matching for vehicle tracking. In [65], SVM classification combined with the correlation measure was used as the similarity measure to evaluate an object in 36 possible directions before detecting it as the car. Xu et al. [69] developed an adaptive switching strategy between the HOG+SVM model and the Haar cascade classifier according to that the HOG+SVM performed better than the Haar cascade classifier when a large number of vehicles were detected. In this case, the detection process can be adjusted according to the change in detection speed in handling more complex scenarios.

Ensemble learning is a technique that involves training and combining multiple weak classifiers to create a robust and accurate classification model [70]. Adaboost technique was used in [71], [72], [73], and [74] for car detection. Tuermer et al. [75] employed a combination of an AdaBoost classifier and Histograms of Oriented Gradients (HOG) features to extract cars from images. More recently, Mahajan et al. [76] utilized the Savitzky-Golay filter [77], XGBoost with adaptive regularization, and Gaussian Filter to remove the noise and anomalies from the dataset.

When facing restrictions on the maximum flight range of UAVs for extensive traffic surveillance, a feasible solution is to divide the monitored area into smaller sections and assign one UAV to each subsection for efficient coverage. The K-means algorithm in combination with optimization approaches has been developed for the placement of UAVs for traffic surveillance [78], [79]. Liu et al. [78] utilized the K-means algorithm for region partitioning and formulated the UAV traffic monitoring as a travelling salesman problem. The simulated annealing algorithm was employed to solve this problem. In the method proposed by Ghazzai et al. [79], the

k-means++ approach selected initial centroids of pre-defined points that were selected using measures, such as vehicle density, event/incident frequency, or vehicle speed. Then, the Particle Swarm Optimization approach was employed to select UAVs with the highest coverage efficiency for each cluster. Top-hat and Bottom-hat transformations were used in [55] for vehicle detection. The optical flow and the k-means techniques were utilized to eliminate background regions. The KLT tracker was applied to vehicle tracking.

B. Deep Learning-Based Models

1) Convolutional Neural Network (CNN): Kyrkou et al. [80] employed the Haar Cascade algorithm and CNN for vehicle detection. Qu et al. [81] introduced a multi-scale spatial pyramid model to down-sample images into feature vectors of fixed length, when the size of images input into the CNN is changed. To handle the large variation in UAV imagery, a deep convolutional feature extractor was employed to discover features from both the UAV imagery and the satellite imagery [82]. Following this, an adversarial framework uses a discriminator to force the extractor to learn indistinguishable features.

Shi et al. [83] converted the vehicle detection task into a multitask learning problem by employing a CNN model to directly predict high-level vehicle features. For high-resolution UAV images with numerous small vehicles, Li et al. [84] designed an image cropping strategy to focus on finer details. They employed ResNet101 [85] as the backbone for feature extraction and addressed the challenge of small vehicles by utilizing low-level features that lack sufficient semantic information.

More recently, Kim et al. [86] employed ResNet101 to extract low-level feature maps from input images and then DeepLabv3 [87] was used for traffic density prediction. In the work presented in [88], pre-trained CNN models like ResNet50 and Mobilenetv2 were employed to extract both vehicle position and type. This extracted information was then utilized to estimate traffic parameters, including flow rate, average speed, and density. Li et al. [89] trained a CNN framework to simultaneously detect and count vehicles in drone-based images. A scale-adaptive strategy was proposed to select anchors and apply circular flow to guide feature extraction.

Despite the success of CNN models in image classification, object detection poses a more intricate challenge that demands sophisticated methods for resolution [34]. Recognizing this complexity, R-CNN [32] adopts Selective Search [90] to generate bottom-up region proposals that are likely to contain objects. Youssef and Elshenawy [91] introduced a technique that combined feature pyramid networks with a Cascade R-CNN architecture for vehicle counting and tracking in the aerial video streams. Deng et al. [92] developed an R-CNN method to simultaneously extract the location and attributes of vehicles. They utilized cropped image blocks for training and implemented data augmentation techniques to mitigate overfitting.

One of the major drawbacks of the R-CNN model is its high computational cost due to performing the CNN feature extraction separately for each region proposal [31]. The Faster R-CNN framework [35] introduced Region Proposal Networks (RPNs) for speeding up the R-CNN model for region proposal generation. The RPN shares full-image convolutional features with the detection network, allowing for almost cost-free generation of region proposals. Building on the impressive performance demonstrated in object detection, the Faster R-CNN model has attracted attention for aerial vehicle detection [93], [94], [95], [96]. Wang et al. [94] changed the anchor size of the Faster R-CNN model and increased the anchor frame of the corresponding smaller scale. To address the poor performance of RPN in Faster R-CNN in locating small-sized vehicles, a hyper RPN was utilized in [95] to extract vehicle-like targets using hierarchical feature maps. A cascade of boosted classifiers was then replaced with the classifier following the RPN. Kim et al. [96] compared the Aggregated Channel Feature classifier and the Faster R-CNN algorithm for vehicle detection. A parallel RPN for Faster R-CNN with a density-based label assignment strategy was developed in [97] to address the challenges of small target detection and complex scenes.

Mask R-CNN [98], an extension of Faster R-CNN, improved upon the original by including a parallel branch dedicated to predicting object masks, running concurrently with the branch responsible for bounding box recognition. A vehicle detection and recognition model based on the Mask R-CNN approach was developed in [99].

In a CNN model, adding more layers can lead to the vanishing gradient problem, making it difficult to train deep networks effectively. VGG16/VGGNet [100] is an improvement of the CNN model for addressing the vanishing gradient problem by using smaller convolution filters (3×3) and stacking numerous layers. Tang et al. [10] added extra convolution layers to the VGG-16 model [100] for vehicle detection. The ResNet [85] is another enhancement that introduced residual blocks in the CNN model, that allow gradients to flow more easily through the network. The authors of [101] and [102] utilized the ResNet for the feature extraction. The SegNet [103] is another improvement of the CNN model that utilizes an encoder-decoder structure in which max pooling and sub-sampling were used to feature map resolution reduction. Furthermore, the fully connected layers of VGG16 were eliminated to reduce the size and complexity of the encoder network, making it more streamlined and manageable. Audebert et al. [104] trained the SegNet model on the VEDAI dataset [105] and transferred their knowledge to the Potsdam and Christchurch datasets.

2) YOLO (You Only Look Once): A combination of Multi-access Edge Computing (MEC) and blockchain technology was introduced in [106] to secure YOLO-based vehicle detection by using an authentication mechanism. The use of YOLOv2 [37] and YOLOv3 were introduced in [107] and [108], respectively. Feng et al. [109] applied YOLOv3 with an image registration method based on ShiTomasi corner detection. Benjdira et al. [110] used YOLOv3 [38] for vehicle detection, the DeepSORT algorithm [111] for vehicle tracking, and the Ray Casting algorithm to count the number of vehicles. Li et al. [112] used YOLOv3 [38] to detect the vehicle regarding multi-vehicle speed estimation. The process of estimating vehicle speed depends on creating a mapping relationship between pixel distance and actual distance in the real world. Chen et al. [113] utilized the YOLOv3 network for vehicle detection and a similarity measurement for vehicle tracking. The Image registration and mapping relationship were used for speed estimation.

YOLOv4-Tiny [114], a simplified version of YOLOv4 for scenarios with limited computational resources, uses CSPDarknet53-tiny instead of CSPDarknet53 as the backbone. In [115] and [116], YOLOv4-Tiny was employed for vehicle detection to handle small objects detection in aerial images with large variations. Some improvements of YOLOv5 for vehicle detection have been developed by replacing specific blocks [117] or adding the attention mechanism [118]. Feng et al. [119] employed YOLOv5 with embedded coordinate attention to enhance long-range dependency capture. In addition, a bidirectional feature fusion module was used to aggregate multi-scale features for improved vehicle detection. In [120], YOLOv7 [121] was used for vehicle detection and the SORT technique [122] for tracking vehicles.

3) Other Models: Despite of the plausible task-wise performance of CNN-based vehicle detection techniques, they face some problems, such as the need for a large dataset for training to avoid the over-fitting problem, and plenty of time to fine-tune many hyper-parameters. To address these challenges, Ma et al. [123] utilised a deep forest model [124], which required fewer hyper-parameters. Also, data augmentation was applied to generate more training samples when there is a small set of a dataset. To reduce the computation load, they introduced the approximate radial gradient transform using a lookup table. The conventional data augmentation approach typically involves applying basic transformations such as rotation, scaling, and flipping to the training data. However, this method has limitations in effectively capturing the essence of feature distribution and providing diverse data samples. To address this issue, Zheng et al. [125] introduced an approach that utilised vehicle synthesis generative adversarial networks to enhance the training data and significantly improve vehicle detection performance. Their proposed framework consisted of one generator and two discriminators, working in tandem to generate realistic vehicle samples while learning the background context concurrently. The framework introduced in [126], the road was segmented using DL models. Then, vehicles were detected using YOLO-based models. Finally, Hungarian algorithm [127] and Kalman Filter [128] were employed for vehicle tracking.

Ensuring privacy in the exchange of data between UAVs and components of ITS may face challenges due to the stringent regulations governing data privacy. Lim et al. [129] employed the federated learning approach to enable privacy-preserving collaborative ML across a federation of independent Drones-as-a-Service providers for the development of Internet of vehicles applications.

C. Reinforcement Learning-Based Models

Deploying networks in UAV-aided ITS involves strategically positioning UAVs for optimal coverage and communication with ground infrastructure. Topology selection considers scalability and reliability. Communication protocols, such as Wi-Fi or LTE, are chosen based on data requirements to ensure effective data relay. Jiang et al. [130] used multi-agent RL MARL for the placement of UAVs for a rural highway with sparse traffic. The RL algorithm employed for this problem was Deep Independent Q-Learning with a modified observation function and sharing of policies. Liu et al. [131] utilized intention information and Multi-Agent RL for optimizing UAV number, scheduling, and task assignment strategies in multiple cooperative UAVs communication.

Robust security measures are necessary to mitigate the risks associated with sensitive information or the compromise of UAV control systems. By prioritizing security, stakeholders can instil trust, protect public safety, and ensure the smooth and secure functioning of UAV-assisted ITS and communications. Papers in this domain cover authentication protocols, threat detection, and privacy-preserving approaches. An RL-based UAV relay scheme against jamming attacks, deriving the Nash equilibrium of the UAV anti-jamming transmission game, was developed in [132]. Besides, the policy hill climbing algorithm was used to choose the UAV relay policy without knowing the specific UAV-ground channel model and attack model.

Due to the limited local computation resources in UAVs and vehicles, vehicular computation requires offloading the computation tasks with time-delay sensitive and complex demands to other intelligent devices once the data is sensed and collected collaboratively. Intelligent devices will work on behalf of the vehicle to execute computation tasks and return the results to the requesting vehicle [133]. In [134], drones equipped with MEC servers were utilized to enhance resource utilization, optimizing both spectrum slicing ratios and workload partitions through neural networks and RL methods. By using states and actions as input and output parameters, policy-gradient algorithms, such as actor-critic, DDPG, and proximal policy optimization. Previous studies have primarily focused on MEC-mounted UAVs individually, without considering diverse resource demands and heterogeneous QoS requirements in vehicular networks. In [135] and [136], the macro eNodeB and UAV are both mounted with MEC servers. Peng and Shen [135] assumed a controller installed at the eNodeB to execute the optimization problem, treated as an agent in a DDPG-based solution due to its non-convex nature and high computational complexity. Building on the last work, Peng and Shen [136] formulated the resource allocation challenge within MEC servers as a distributed optimization problem, tackled using a multi-agent DDPG approach. Each MEC server acted as an individual agent, making real-time resource allocation decisions autonomously based on a pre-trained offline multi-agent DDPG model.

Typically, formulating UAV-to-Vehicular Network communication as a Markov Decision Process (MDP) provides a systematic framework. This approach, utilizing RL or dynamic programming, enables intelligent decision-making by UAVs, considering channel conditions, vehicle mobility, and energy constraints. The aim is to optimize communication objectives, ensuring efficient data transfer and reliable connectivity in dynamic Vehicular Networks [137], [138], [139], [140]. In [141], the cooperation between UAVs, RSUs, and vehicles was formulated as MDP to maximize the number of served vehicles. DDPG with an enhanced reward function was applied in [142] for path planning. The method in [143] and [144] leveraged detailed city maps and complex Q-Learning algorithms for dynamic trajectory planning. Unlike focusing on optimal UAV placement, this approach tracked moving users throughout their journey. Beam-related information was incorporated as inputs to the RL model to enhance trajectory.

In complex and uncertain real-world road environments, MDPs can improve trajectory planning to ensure effective UAVs navigation. Samir et al. [9], [11], [145] formulated the trajectory planning and resource allocation of UAVs as an MDP problem. They used RL algorithms, such as the Q-learning and Deep Deterministic Policy Gradient (DDGP) [146], to solve the original underlying problem.

Proximal Policy Optimization and DDQN algorithms were exploited in [147] and [148] to solve the problem of content delivery to vehicles on the road and maximizing the UAV coverage. Wu et al. [149] combined the genetic algorithm and the DQN algorithm to identify the optimal hovering position for each assigned mission area to a UAV. Deng et al. [5] modeled the resource allocation problem as a 3-partite graph matching problem [150] and used the Q-Learning algorithm to control the trajectory of UAV. Qi et al. [151] transformed the optimization task of resource allocation into a nonlinear mixed-integer programming problem and addressed it using a combined approach of the Hungarian algorithm and DDQN.

D. Hybrid Models

Hybrid ML models leverage the strengths of different algorithms to enhance accuracy, robustness, and adaptability. Ghasemi et al. [161], [162] introduced ensemble learning for vehicle detection utilizing pre-trained Faster R-CNN models from the ImageNet dataset. The first study involved developing ensemble models comprising 15 learners, leveraging three Faster R-CNN-based models (InceptionV3, ResNet50, and GoogleNet) as feature extractors, along with five transfer classifiers (KNN, SVM, MLP, C4.5 Decision Tree, and Naïve Bayes) [161]. In a subsequent study, the authors trained Faster R-CNN with ResNet50 and incorporated six base learners [162]. The ensemble's performance was optimized by dynamically adjusting the weights assigned to individual base learners and the final decision threshold using the Genetic algorithm. A vehicle detection system was introduced in [165]. During the initial phase, a cascade detection algorithm was utilized to discard the majority of background elements and identify patterns corresponding to man-made objects. Subsequently, in the second stage, these patterns underwent refinement through the application of image classification techniques, including KNN, SVM, decision trees, and random forest.

A combination of K-means clustering, the soft Non-Maximum Suppression (NMS) algorithm [166], and the YOLO algorithm were developed in [155] and [156] for vehicle detection. In [156], the K-means++ technique was used to cluster the label boxes of vehicle targets in the training dataset. On the other hand, using the Soft-NMS algorithm, appropriate detection boxes are suppressed. In the final step, the YOLOv3 model was used for vehicle detection. In [155], the anchor aspect and the optimal anchor width and height dimension were determined using the K-means algorithm. Combining the SVM classifier with other ML models, like CNNs, has been taken into account for vehicle detection because SVMs excel at classification tasks especially in handling structured data, while CNNs excel at feature extraction from raw data [157]. Elmikaty and Stathaki [159], [160] first utilized a Gaussian Mixture Model (GMM) classifier to identify regions of interest and evaluated the likelihood of the detection windows to a target for reducing search areas. Then, a SVM classifier was used for car classification. The work [158] proposed an approach for UAV trajectory optimization for content delivery to Vehicular Networks. Using K-means clustering, vehicles are grouped based on cellular throughput, location, and content preferences. A heuristics determines the minimum number of UAVs needed for vehicle throughput. Content delivery and trajectory design are jointly optimized using time-based graph decomposition. PSO algorithm optimizes content placement and a CNN-based learning scheme aids in online decision-making.

E. Summary of Literature

Table I summarizes ML techniques applied to UAV-assisted ITS, with diverse models across applications. Notably, DL dominates, contributing to vehicle detection, tracking, road detection, and resource management. Particularly, YOLO specializes in vehicle detection. K-Means technique aids in monitoring and optimization. RL and deep RL mostly focus on transmission, trajectory estimation, and energy-efficient resource management. The majority of models focus on Vehicle Detection/Tracking (62 papers, 71%), indicating a strong emphasis on enhancing vehicle-related tasks in UAV-assisted ITS applications. Among the surveyed 88 papers, DL techniques show a dominant proportion with 39 papers (45%). According to our findings and the contributions of DL techniques in UAV-aided ITS, particularly for the perception layer of UAV-aided ITS and tasks in this layer, i.e., vehicle detection and tracking, road detection, pedestrian detection, and traffic parameter extraction, the next section provides a comparative study of the crossover between DL and the perception layer.

IV. Comparative Study

While many studies have applied ML models for drone data analysis in ITS applications, the lack of standardization in evaluation metrics, diverse datasets, and specific focus areas poses difficulties in drawing universal conclusions. Particularly, recent advancements in DL algorithms have made them suitable for visual data in the perception layer of UAV-aided ITS. A comparative study in this area contributes to the development of robust and efficient solutions for ITS enhanced by UAV technology. Detection tasks, including identifying vehicles, pedestrians, and cyclists, as well as vehicle classification,

TABLE I

A SUMMARY OF TECHNIQUES FOR ML TECHNIQUES FOR UAV-ASSISTED ITS (VD: VEHICLE DETECTION, VT: VEHICLE TRACKING, RD: ROAD DETECTION, RM: RESOURCE MANAGEMENT, FL: FEDERATED LEARNING, BN: BAYESIAN NETWORKS, EL: ENSEMBLE LEARNING)

	Applications								**Total**
ML model	**Monitoring**	**Transmission**	**VD/VT**	**RD**	**PD**	**TE**	**Trajectory**	**RM**	
Clustering	[78]	—	[51], [54], [55], [56], [152] [57]	—	—	[52], [53]	[79]	—	**10**
SVM	—	—	[44], [46], [47], [48], [63] [65], [66], [67], [69]	—	[45]	[153]	—	—	**12**
RL	—	[130], [133] [135], [136] [137], [140]	—	—	—	—	[5], [142] [143], [144] [145], [147]	[134], [154]	**14**
DRL	—	[131], [138] [139], [141]	—	—	—	—	[9], [11] [148], [149]	[151]	**9**
DL	—	—	[10], [80], [81], [82], [83] [86], [88], [91], [92], [93] [94], [95], [96], [97], [99] [101], [102], [104], [123] [125]	[126]	—	—	—	—	**21**
YOLO	—	—	[106], [107], [108], [109] [112], [113], [117], [120] [126], [155]	—	—	—	—	—	**10**
FL	—	[129]	—	—	—	—	—	—	**1**
Hybrid	—	—	K-Means + YOLOv4 [155], [156] CNN+SVM [157] GMM+SVM [159], [160] EL [161], [162] Adaboost + SVM [64]	—	—	—	K-means + CNN [158]	—	**9**
Adaboost	—	—	[59], [60], [61], [68], [71] [72], [73], [74], [75]	—	—	—	—	—	**9**
BN	[50]	—	[41], [42], [43]	[163]	—	—	—	—	**5**
Other ML	—	—	[49], [76], [164], [165]	—	—	—	—	—	**4**
Total	**2**	**11**	**70**	**2**	**1**	**3**	**12**	**3**	**<u>104</u>**

TABLE II

DATASETS FOR EXPERIMENTS (VD: VEHICLE DETECTION, VC: VEHICLE CLASSIFICATION, PD: PEDESTRIAN DETECTION, CD: CYCLIST DETECTION, SE: SPEED ESTIMATION, VCN: VEHICLE COUNTING)

Dataset	**Dataset size**	**Type**	**Resolution**	**Training Set**	**Testing Set**	VD	VC	PD	CD	SE	VCn
Stanford Drone	60	Video	1424×1088	10 Videos	1 Video (13335 Frames)	✓	—	✓	✓	—	✓
UAV-ROD	1576	Image	1280 (1920×1080) 296 (2720×1530)	913 233	367 60	✓	—	✓	—	✓	✓
UCAS-AOD	510	Images	1280×659	400	110	✓	—	—	—	—	✓
FHY-XD-UAV-DATA	12,300	Images	1280×960	11,000	1300	✓	—	—	—	✓	✓
PSU Car Dataset	270	Images	1920×1280	218	52	✓	✓	✓	—	✓	✓

help real-time monitoring, incident detection, safety improvement, and data-driven decision-making. The integration of detection and classification capabilities enhances the overall intelligence and adaptability of ITS, making it more responsive to the evolving dynamics of urban mobility. However, these analyses cannot be applicable alone in traffic management and planning. Traffic engineers find aggregated traffic flow parameters more valuable than instantaneous results. In our study, we extend the analysis by computing aggregated traffic flow parameters derived from frame-based information. This approach allows for a more comprehensive understanding of traffic dynamics over time, providing insights that are particularly relevant for traffic management and planning purposes.

To address these challenges, this section provides a comprehensive synthesis to distill findings and insights to guide the development of effective ML solutions tailored to drone data in diverse ITS scenarios. Building upon the investigation from previous investigations, which highlighted the efficacy of both CNN-based and YOLO-based techniques for detection and classification tasks within the perception layer of UAV-aided ITS, we have prioritized their utilization in our research. In response to their significance, we have meticulously crafted 14 DL models, harnessing the capabilities of baseline CNN, R-CNN, and Faster R-CNN in tandem with three backbone architectures of ResNet, VGG16, and Inception, and five versions of YOLO networks. These models are trained and tested on five aerial datasets using multiple evaluation metrics. We have conducted experiments in two groups: detection/classification tasks and traffic extraction tasks.

Section IV-A describes the characteristics of the datasets used in this study. Evaluation benchmarks are defined in Section IV-B. Section IV-C provides the experimental setup and implementation details. Section IV-D presents the main experimental results and our main findings.

A. Datasets

There are public and private datasets that can be used for the evaluation of DL models in the perception layer of UAV-aided ITS. We have selected five public datasets, which are frequently used in the UAV-aided ITS field. (1) ***The UAV-ROD dataset***[1] comprises 1577 low-altitude images captured at heights ranging from 30m to 80m. The original images are sampled into two resolutions: 1920 × 1080 and 2720 × 1530. (2) ***The UCAS-AOD dataset***[2] is a high-resolution aerial-object detection dataset collected from Google Earth. It consists of 1510 images, including 1000 images depicting planes and 510 images featuring cars. (3) ***The PSU Car Dataset***[3] is compiled from two sources: 123 images captured by a 3DR SOLO drone equipped with a GoPro Hero 4 camera, and the remaining images sourced from an open dataset of aerial images accessible on Github.[4] (4) ***The FHY-XD-UAV-DATA dataset***[5] [112] was collected using UAV DJI-MATRICE 100 with a Point Grey monocular camera in five scenes. This dataset includes 20,448 images with a resolution of 1280 × 960. (5) ***The Stanford Drone Dataset*** [167] consists of videos recorded in eight unique scenes.

Some major characteristics of these datasets are summarised in Table II. Based on the unique characteristics of the datasets and the objects they contain, we developed custom models tailored to specific ITS applications for each dataset. For instance, the UAV-ROD dataset consists of continuous, frame-by-frame recorded images, making it well-suited for speed estimation tasks. On the other hand, the UCAS-AOD dataset is composed of individually captured images and is not suitable for the task of estimating speed.

B. Evaluation Criteria

To assess the developed models for the perception layer of UAVs for ITS, several widely utilized evaluation criteria are used. We categorize these criteria into two groups: (1) *performance measures*, including the precision-recall curve (PRC), precision, recall, F1-score, and False alarm rate (FAR); and (2) *efficiency criteria*, including Frames Per Second (FPS) and inference time.

Precision-Recall Curve (PRC): In vehicle detection, the PRC is a useful tool for assessing and fine-tuning the model's performance, especially in scenarios where the goals of ensuring safety and minimizing false alarms are critical.

Recall: This measure reflects the fraction of correctly identified actual vehicles present in images or video frames.

Precision: This measures the system's accuracy in terms of correctly identifying vehicles within images or video frames.

F1-score: This is defined as the harmonic mean of precision and recall.

FAR: The false alarm rate in the context of vehicle detection refers to the rate at which a system or model incorrectly identifies non-vehicle objects as vehicles.

FPS: This criterion shows how many frames are processed per second.

Inference time: Measuring the processing speed in millisecond per image.

C. Experimental Setup

All experiments were conducted on a machine equipped with an Intel CPU featuring 8 cores at 2.2 GHz and 32 GB RAM. The operating system used was CentOS 7 Linux, and the machine was equipped with a Tesla P100-PCIE-16 GB GPU, boasting 16 GB of GPU memory. The development environment was built on Python-3.9.7 and Tensorflow-2.9.1.

To fine-tune hyperparameter values and model architectures for the DL models, we've employed a trial-and-error approach. For all implemented models, we use the *Adam* as the training optimizer, setting the initial learning rate to 0.001. The batch size is set to 16 and the epoch to 200. Mean Squared Error (MSE) and cross-entropy (CE) are considered as the localization loss and classification loss, respectively. For all YOLO-based models, the confidence and IoU thresholds are set to 0.5 and 0.3 in all experiments. The anchor ratio and scale values for the Faster R-CNN and YOLO-based models are set to 0.5 and 32, respectively.

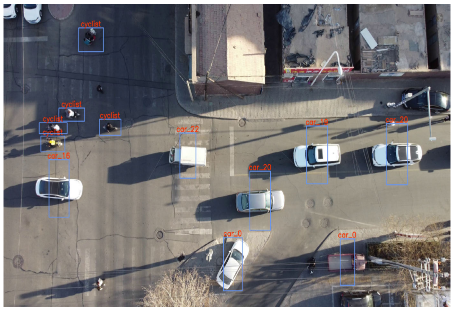
Labelling was necessary for some datasets for performing ITS-related tasks because these datasets either do not have annotations (e.g., the FHY-XD-UAV-DATA dataset) at all or only have partial annotations (e.g., UAV-ROD, PSU car Dataset, and Stanford dataset). For this reason, we used the LabelImg tool[6] to label vehicles in the datasets to train the network models.

D. Experimental Results

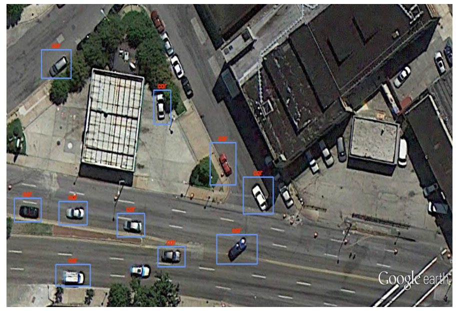
To evaluate the DL models' performance, the results are divided into two parts: performance on detection tasks and performance on traffic extraction tasks. Fig. 1 showcases representative samples of the detection and speed estimation tasks performed by the Faster R-CNN+ResNet model on aerial images utilized. From this figure, it can be seen that this model has more false detections and missing detections on the UCAS-AOD dataset, FHY-XD-UAV-DATA dataset, and Stanford Dataset, demonstrating the struggle with the accurate localization of small objects.

1) Results for Detection Tasks: In this study, four detection-related tasks, including vehicle detection, vehicle classification, pedestrian detection, and cyclist detection, are implemented using various experiments. Since all of the metrics are the same for evaluating all of these tasks, Table III compares the average performance of the designed models on the datasets for the detection tasks. In all datasets, YOLO-based models, especially when coupled with Darknet, emerge as robust choices with high recall and precision. Moreover, YOLO's unique grid-based prediction and streamlined design enable faster inference times and higher FPS, crucial for timely detection. The Stanford Drone, UAV-ROD, and PSU Car datasets show significant class imbalance, impacting

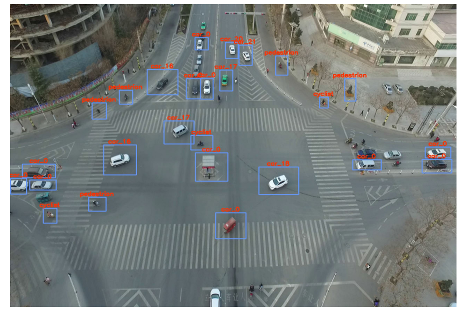
[1] https://github.com/fengkaibit/UAV-ROD
[2] https://hyper.ai/datasets/5419
[3] https://github.com/aniskoubaa/psu-car-dataset
[4] https://github.com/jekhor/aerial-cars-dataset
[5] https://shuochen365.github.io/
[6] https://github.com/tzutalin/labelImg



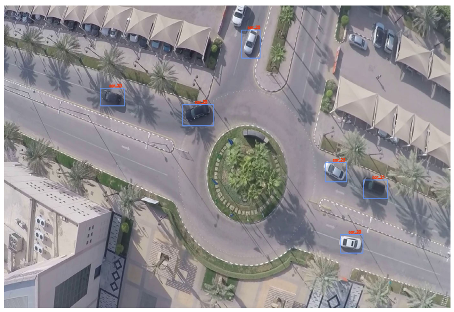
(a) UAV-ROD



(b) UCAS-AOD



(c) FHY-XD-UAV-DATA



(d) PSU Car Dataset

(e) Stanford Dataset

Fig. 1. Examples of aerial images analysed for detection and speed estimation tasks for the datasets used in this study.

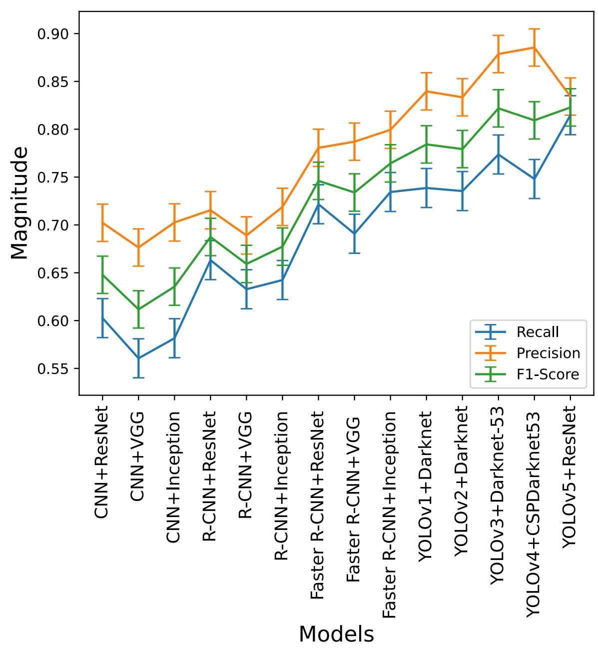


Fig. 2. The average difference in performance across three metrics (i.e., recall, precision, and F1-score) calculated for the DL models applied to datasets, considering both detection and classification tasks.

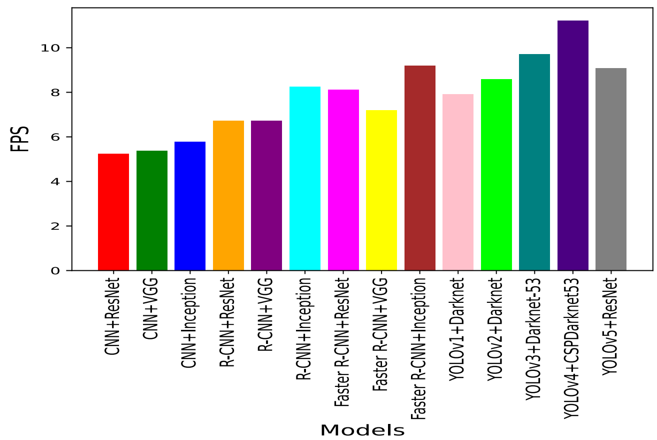
model performance. In the Stanford Drone dataset, pedestrian instances outnumber others, whereas vehicles dominate the UAV-ROD and PSU Car. As a result of this imbalance, it's evident that the models' performance on these datasets is inferior compared to others. This discrepancy is particularly noticeable in metrics like recall and F1-score, which better highlight the models' superiority on imbalanced data. For instance, the Stanford and UAV-ROD datasets exhibited the poorest recall, scoring 0.64 and 0.66, respectively. Moreover, small objects in the Stanford dataset pose a challenge due to their limited size, making them difficult to detect accurately. Their tiny appearance and sparse visual information hinder model performance, leading to lower detection accuracy.

For a more comprehensive visualization of our results in both detecting and classifying objects in aerial images and videos, Fig. 2 presents an error bar plot illustrating the performance of various methods across three criteria: recall, precision, and F1-score. The YOLOv5+ResNet model showcases a commendable average recall of 0.814. In contrast, the YOLOv4+CSPDarknet53 model achieves the highest precision of 0.88. As for the F1-Score, both the YOLOv5+ResNet and YOLOv3+Darknet-53 models exhibit identical performance, with an average score of 0.82. The lower variance in the figure indicates superior performance compared to other models. According to the findings, the YOLOv5+ResNet model exhibited superior performance and demonstrated lower variances compared to other methods.

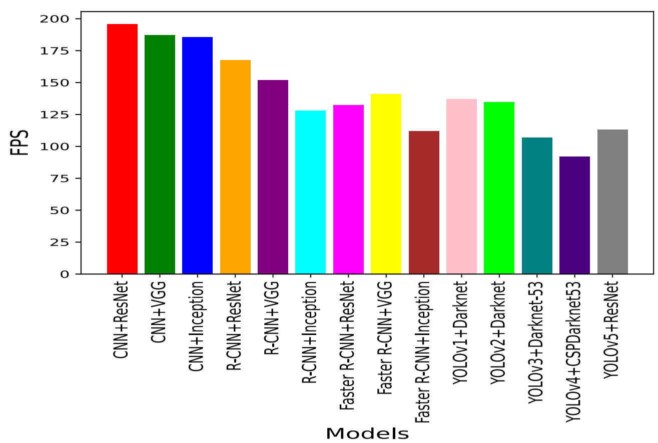
The results in Fig. 3 indicate variations in the average FPS and Inference Time across the datasets. Notably, YOLOv4+CSPDarknet53 exhibits the highest value (11.32), suggesting a faster processing speed. On the other hand, CNN+ResNet needed the most computational resources with an average FPS of 5.24. Overall, YOLO-based models that are a single-shot architecture could provide a faster detection process. This stems from the YOLO architecture and the specific configurations used. YOLO is designed to perform vehicle detection in a single forward pass, making it inherently faster compared to models that involve multiple stages or region proposals, such as Faster R-CNN. Moreover, YOLOv3 and YOLOv4 incorporate feature pyramids, enabling the

TABLE III
AVERAGE PERFORMANCE OF DL MODELS FOR DETECTION TASK (VEHICLE, PEDESTRIAN, AND CYCLIST)
(AR: AVERAGE RECALL, AP: AVERAGE PRECISION, AF: AVERAGE F1-SCORE)

Dataset	Model + Backbone (Feature Extractor)	AR	AP	AF	FAR	FPS	Inference Time (ms)
UAV-ROD	CNN+ResNet	0.522	0.621	0.5672	0.379	5.3	188.6
	CNN+VGG	0.433	0.592	0.5001	0.408	6.1	163.9
	CNN+Inception	0.471	0.603	0.5288	0.397	5.15	194.17
	R-CNN+ResNet	0.658	0.638	0.6478	0.362	6.7	149.2
	R-CNN+VGG	0.587	0.645	0.6146	0.355	7.36	135.86
	R-CNN+Inception	0.627	0.694	0.6588	0.306	8.9	112.35
	Faster R-CNN+ResNet	0.822	0.708	0.7607	0.292	9.2	108.6
	Faster R-CNN+VGG	0.71	0.721	0.7154	0.279	7.52	133
	Faster R-CNN+Inception	0.743	0.731	0.7369	0.269	8.2	122
	YOLOv1 + Darknet	0.805	0.812	0.8084	0.188	11.3	88
	YOLOv2 + Darknet	0.632	0.754	0.6876	0.246	**13.5**	**74**
	YOLOv3 + Darknet-53	0.79	0.902	**0.8422**	0.098	12.6	79.3
	YOLOv4 + CSPDarknet53	0.712	**0.961**	0.8179	**0.039**	12.8	78.12
	YOLOv5 + ResNet	**0.8245**	0.767	0.7947	0.233	10.2	98.03
UCAS-AOD	CNN+ResNet	0.583	0.726	0.646	0.274	6.88	145.34
	CNN+VGG	0.561	0.718	0.629	0.282	5.73	174.52
	CNN+Inception	0.553	0.711	0.622	0.289	8.71	114.81
	R-CNN+ResNet	0.684	0.736	0.709	0.264	11.54	86.655
	R-CNN+VGG	0.631	0.722	0.673	0.278	7.34	136.23
	R-CNN+Inception	0.652	0.753	0.698	0.247	11.2	89.28
	Faster R-CNN+ResNet	0.784	0.782	0.782	0.218	5.17	193.42
	Faster R-CNN+VGG	0.765	0.829	0.795	0.171	5.61	178.25
	Faster R-CNN+Inception	**0.821**	0.855	0.837	0.145	12.6	79.36
	YOLOv1 + Darknet	0.734	0.86	0.792	0.14	9.06	110.37
	YOLOv2 + Darknet	0.752	**0.896**	0.817	**0.104**	9.54	104.82
	YOLOv3 + Darknet-53	0.814	0.883	**0.847**	0.117	10.46	95.6
	YOLOv4 +CSPDarknet53	0.728	0.851	0.784	0.149	**13.2**	**75.75**
	YOLOv5 + ResNet	0.808	0.836	0.821	0.164	6.42	155.76
FHY-XD-UAV-DATA	CNN+ResNet	0.623	0.678	0.649	0.322	4.3	232.5
	CNN+VGG	0.584	0.624	0.603	0.376	5.1	196
	CNN+Inception	0.636	0.712	0.671	0.288	3.8	263.15
	R-CNN+ResNet	0.651	0.738	0.691	0.262	4.68	213.67
	R-CNN+VGG	0.644	0.681	0.661	0.319	5.21	191.93
	R-CNN+Inception	0.597	0.704	0.646	0.296	7.44	134.4
	Faster R-CNN+ResNet	0.674	0.814	0.737	0.186	**10.2**	**98.03**
	Faster R-CNN+VGG	0.622	0.826	0.709	0.174	7.26	137.74
	Faster R-CNN+Inception	0.708	0.834	0.765	0.166	8.55	116.95
	YOLOv1 + Darknet	0.685	0.816	0.744	0.184	4.69	213.21
	YOLOv2 + Darknet	**0.755**	0.831	0.791	0.169	4.38	228.31
	YOLOv3 + Darknet-53	0.738	0.838	0.784	0.162	7.23	138.31
	YOLOv4 +CSPDarknet53	0.743	0.841	0.788	0.159	9.06	110.37
	YOLOv5 + ResNet	0.772	**0.864**	**0.815**	**0.136**	9.78	102.24
PSU Car Dataset	CNN+ResNet	0.674	0.761	0.714	0.239	5.12	195.31
	CNN+VGG	0.653	0.743	0.695	0.257	4.88	204.91
	CNN+Inception	0.697	0.802	0.745	0.198	5.54	180.5
	R-CNN+ResNet	0.715	0.813	0.76	0.187	6.33	157.97
	R-CNN+VGG	0.708	0.763	0.734	0.237	7.56	132.27
	R-CNN+Inception	0.713	0.832	0.767	0.168	8.23	121.5
	Faster R-CNN+ResNet	0.71	0.844	0.771	0.156	**9.8**	**102.04**
	Faster R-CNN+VGG	0.725	0.823	0.77	0.177	7.81	128.02
	Faster R-CNN+Inception	0.743	0.851	0.793	0.149	8.4	119
	YOLOv1 + Darknet	0.786	0.834	0.809	0.166	7.13	140.25
	YOLOv2 + Darknet	0.832	0.795	0.813	0.205	6.34	157.72
	YOLOv3 + Darknet-53	0.811	0.867	0.838	0.133	8.04	124.37
	YOLOv4 +CSPDarknet53	0.824	**0.887**	0.854	**0.113**	8.7	114.94
	YOLOv5 + ResNet	**0.915**	0.869	**0.891**	0.131	9.3	107.52
Stanford Dataset	CNN+ResNet	0.611	0.725	0.6631	0.275	4.6	217.3
	CNN+VGG	0.572	0.705	0.6315	0.295	5.1	196
	CNN+Inception	0.551	0.685	0.6107	0.315	5.7	175.4
	R-CNN+ResNet	0.608	0.652	0.6292	0.348	4.36	229.3
	R-CNN+VGG	0.594	0.634	0.6133	0.366	6.14	162.8
	R-CNN+Inception	0.623	0.611	0.6169	0.389	5.5	181.8
	Faster R-CNN+ResNet	0.618	0.755	0.6796	0.245	6.25	160
	Faster R-CNN+VGG	0.632	0.736	0.68	0.264	7.8	128.2
	Faster R-CNN+Inception	0.657	0.726	0.6897	0.274	8.2	121.9
	YOLOv1 + Darknet	0.683	0.876	0.7675	0.124	7.46	134
	YOLOv2 + Darknet	0.706	0.891	0.7877	0.109	9.25	108.1
	YOLOv3 + Darknet-53	0.715	**0.903**	0.798	**0.097**	10.3	97
	YOLOv4 +CSPDarknet53	0.733	0.887	**0.8026**	0.113	**12.4**	**80.6**
	YOLOv5 + ResNet	**0.754**	0.835	0.7924	0.165	9.78	102.2



(a) Average FPS on the datasets



(b) Average Inference Time on the datasets

Fig. 3. Average efficiency of the DL models for classification- and detection-related tasks on the datasets in terms of average FPS and Inference time.

model to detect objects at different scales efficiently. Among double-shot architectures (i.e., CNN, R-CNN and Faster R-CNN using different backbones), Faster R-CNN+Inception also performs well in terms of FPS and inference time with average values of 9.19 and 111.8 ms, respectively.

Fig. 4 displays the Precision-Recall Curves of the methods for the ten test images. As can be seen from these, the YOLO-based and Faster R-CNN-based methods have much better detection performance than the CNN-based and R-CNN-based detectors. Furthermore, the outcomes indicated that deeper networks, such as YOLO-based models and those incorporating ResNet, exhibit enhanced feature representation capabilities, leading to improved detection performance. This highlights the effectiveness of leveraging features from various layers and anchor boxes, ultimately contributing to superior detection results.

2) Results for Traffic Extraction (Speed and Counting): Fig. 5 shows the estimated count in comparison with the ground truth count for test set frames of the datasets. The accuracy of vehicle count estimation was closely tied to the detector's performance, as the count was determined by the number of detection windows identified in each frame. The calculation accuracies were high for the FHY-XD-UAV-DATA and Stanford Datasets, around 0.72 and 0.55, respectively. Our study reveals that the performance of DL models in object detection and counting on roads is significantly influenced by the traffic composition and the attitude of the images or videos. Notably, the Stanford dataset, recorded using a high-attitude drone, presents specific challenges. The dataset encompasses small-sized objects like pedestrians and cyclists, often accompanied by shadows, adding complexity to the vehicle localization process.

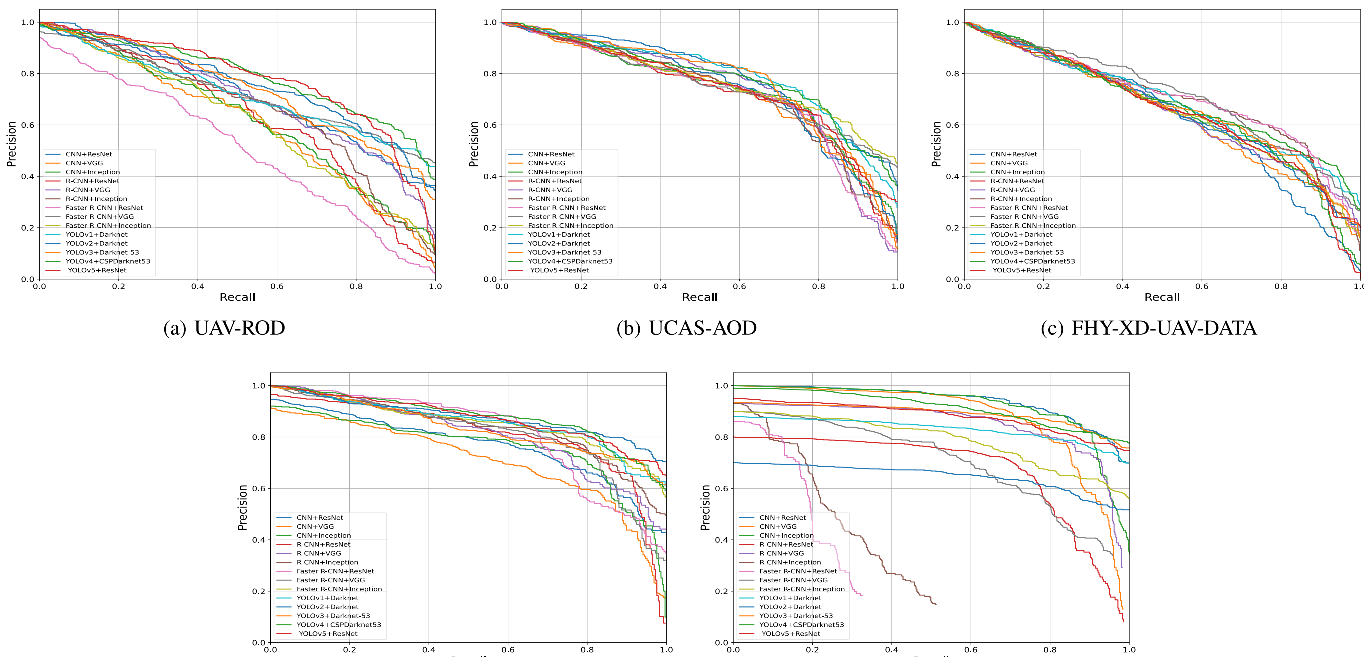
To evaluate the traffic parameter estimation accuracy of the DL methods, ground truth data on the average speed per pixel by frame was measured from the aerial videos. It is crucial to consider an appropriate measurement interval for speed estimation. For dynamic processes or systems with rapid changes, a smaller interval might be necessary to capture fluctuations accurately. On the other hand, for processes with slower changes or when a general trend is sufficient, a larger interval might be more practical, reducing the amount of data to process and store. In our experiments, the vehicle's motion is analyzed over a sequence of five frames, and the vehicle's speed is then determined based on this measurement interval, expressed in miles per hour. A tracking algorithm should be applied to predict the position of the vehicle based on its previous state. For this reason, we employed the KLT tracker [168].

The findings depicted in Fig. 6 indicate that, in both directions, the average speed of the traffic streams remained relatively consistent over time for the FHY-XD-UAV-DATA Dataset while varying more obviously in the UAV-ROD and PSU Car datasets. From visual inspections, the average speed difference between the ground truth and the estimated for the two traffic directions over the three datasets of the UAV-ROD, FHY-XD-UAV-DATA, PSU Car datasets are 4.25, 0.93, and 3, respectively. The outcomes are deemed reasonable since, under free-flow and moderately congested traffic conditions in the FHY-XD-UAV-DATA Dataset, vehicles maintain a relatively steady speed dictated by the prevailing traffic state.

V. Open Issues and Future Directions

There is no doubt that ML and DL techniques have great potential to improve the surveillance and decision-making of UAVs and drones in the ITS context. However, our review in Section III and experimental findings in Section IV showed that several limitations and challenges need to be considered for future work.

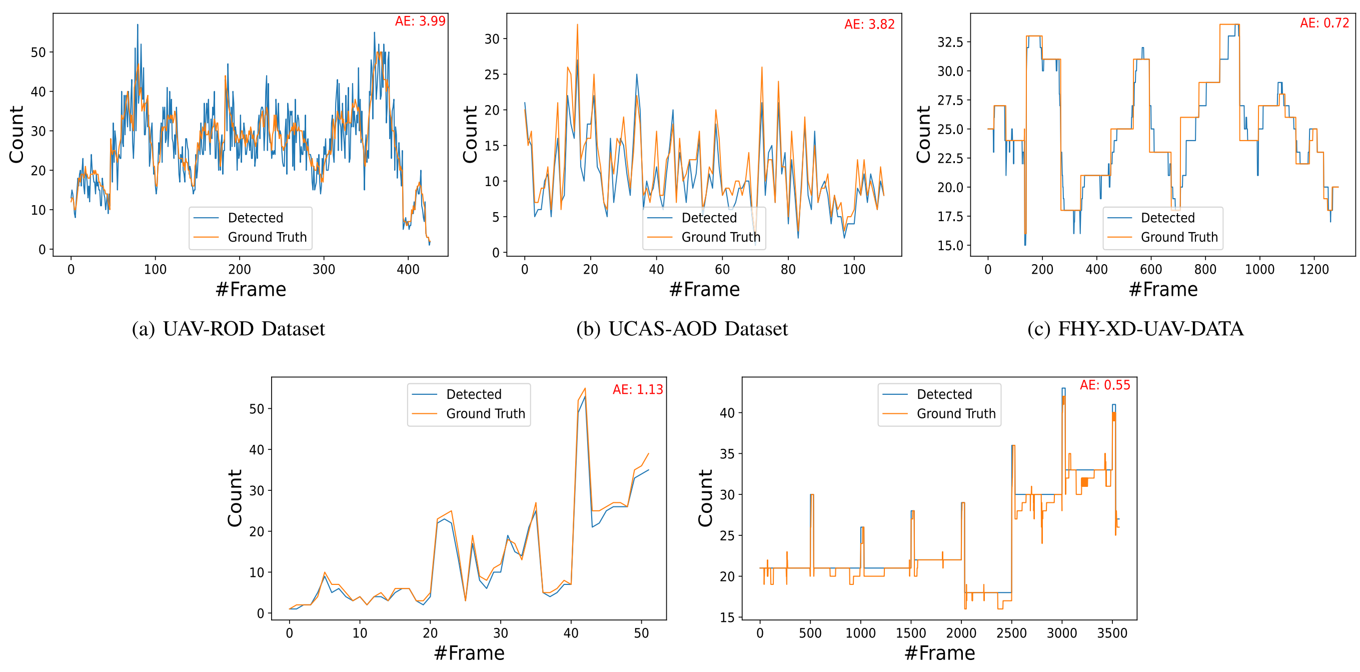
- ***Real-time Processing:*** Real-time processing capabilities ensure timely decision-making during UAV missions. This could involve optimizing models for deployment on UAVs with limited computational resources. Multi-GPU configurations can be incorporated to reduce the computational time.
- ***Federated learning for secure model training:*** Federated learning in UAV-aided ITS ensures decentralized model training, prioritizes privacy, and leverages the collective intelligence of UAVs to improve model adaptability. This collaborative approach aligns with the distributed and dynamic nature of UAV networks, contributing to more



(a) UAV-ROD (b) UCAS-AOD (c) FHY-XD-UAV-DATA

(d) PSU Car Dataset (e) Stanford Dataset

Fig. 4. Precision-recall curves of the DL methods on the datasets used in this study.



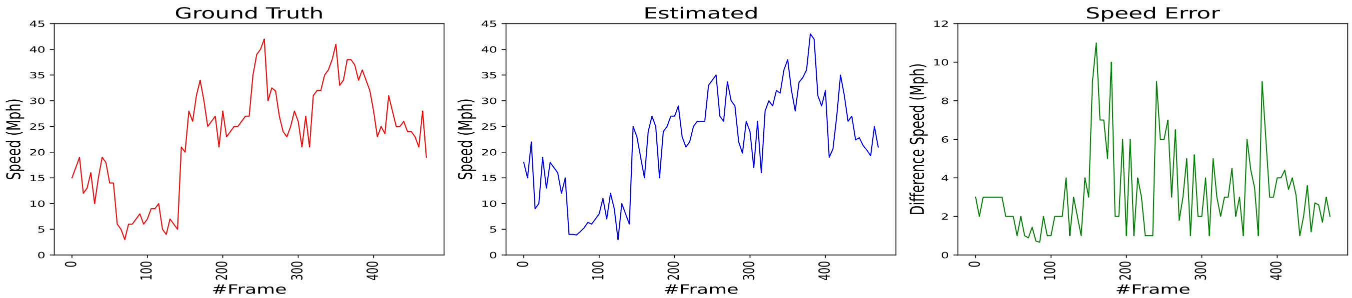
(a) UAV-ROD Dataset (b) UCAS-AOD Dataset (c) FHY-XD-UAV-DATA

(d) PSU Car Dataset (e) Stanford Dataset

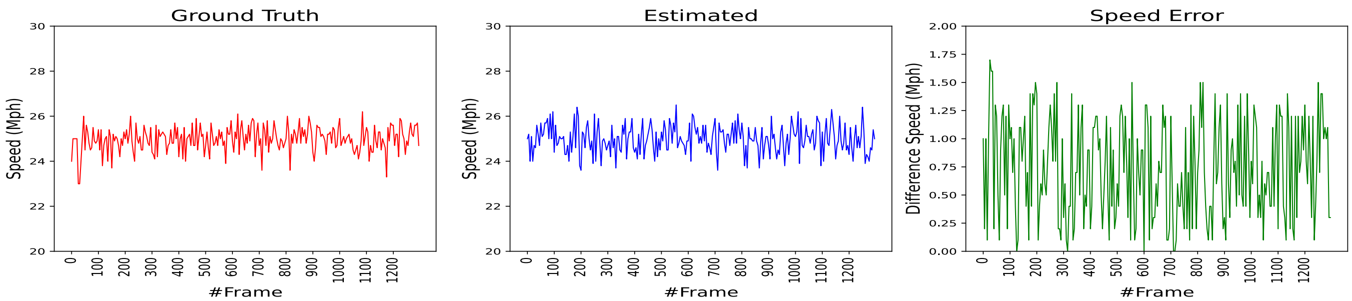
Fig. 5. The number of Vehicles detected against the actual value for the UAV-ROD dataset.

effective and privacy-conscious intelligent transportation systems. For example, Federated learning enables UAVs to update their detection and classification models without sharing raw data. Only model updates, in the form of model parameters, are exchanged between UAVs and a central server.

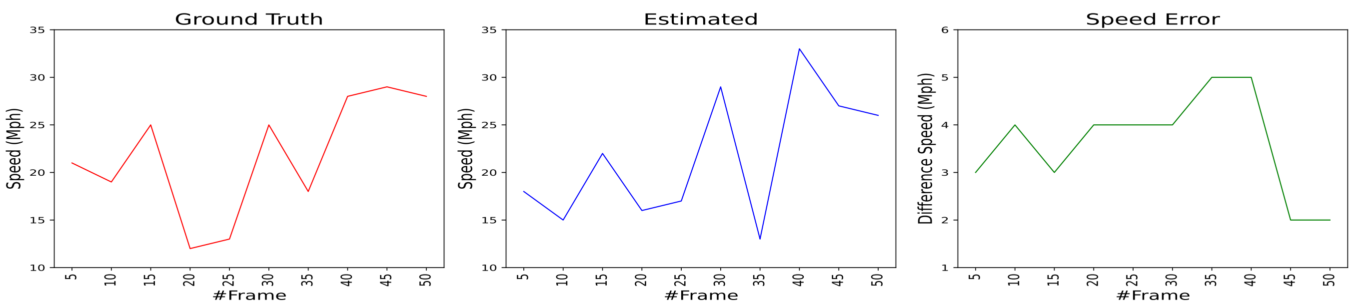
- ***Adversarial training for UAV-aided ITS:*** ML approaches employed in UAV-assisted ITS are susceptible to adversarial attacks. Adversarial attacks target these models by introducing meticulously crafted perturbations or inputs. The intention is to manipulate ML models, causing misclassification or generating incorrect predictions. Strategies, such as adversarial training, are essential to bolster the resilience of ML-based UAV systems against potential adversarial threats. Our team plans to investigate the impact of adversarial attacks on different DL models for vehicle detection and classification tasks from aerial videos/images.



(a) UAV-ROD Dataset



(b) FHY-XD-UAV-DATA Dataset



(c) PSU Car Dataset

Fig. 6. Accuracy of traffic flow speed estimation.

- ***Architectural adaptation for vehicle size variability in DL models:*** The experiments demonstrated that the implementation of the DL models was hindered by the presence of very small objects against cluttered backgrounds, especially at greater altitudes. To address this issue, adding more layers with various output sizes is helpful. In this way, a DL model can process features at different spatial scales. Moreover, this facilitates hierarchical feature extraction, allowing the model to learn complex representations that encompass a wide range of object sizes.
  item ***Anchor-free DL methods:*** YOLO models rely heavily on anchor boxes for predictions. Anchor-free methods like YOLOX [169] offer potential solutions to streamline detection complexity. By eliminating the need for anchor boxes, YOLOX streamlines the detection process, offering a promising alternative for object detection tasks and contributing to the ongoing evolution of DL techniques in computer vision research.
- ***Applying Deconvolution layers to DL models:*** By upsampling the feature maps, deconvolution layers enable a DL model to capture finer details and spatial information, thereby improving its ability to accurately localize and recognize objects in images. This process effectively enhances the model's understanding of complex visual patterns and facilitates more precise object detection and segmentation, particularly in tasks where high-resolution representations.
- ***Advanced ensemble Learning using diverse DL models:*** Exploration of advanced ensemble techniques such as stacking, boosting, or meta-learning can provide insights into optimizing model fusion and decision-making processes. In addition, the ensemble transformer with the attention module encoder can improve the detection of small objects.
- ***Strategies for class imbalance problem:*** Class imbalance, particularly the foreground-background (when the background class vastly outnumbers the foreground class) and foreground classes are much more prevalent than others), which is a se foreground (where certain foreground rious problem in object detection has not been touched in the ITS community. Re-weighting of focal loss and few-shot learning are two potential solutions for this reason.

- ***Hyper-parameters optimization for DL models:*** Selecting suitable values for hyperparameters, such as learning rate, activation functions, and optimizer Parameters (e.g., momentum, decay rates, epsilon) of DL models significantly influence their effectiveness and resource efficiency in tasks such as detection and classification in UAV-assisted ITS. Optimization strategies can be grid search, Bayesian optimization, evolutionary algorithms, and gradient-based optimization.

VI. Conclusion

In this work, we presented a survey and a comparative study on the contribution of ML and DL approaches in the field of UAV-aided ITS. First, we systematically summarized existing methods in ML employed for aerial ITS-related datasets regarding traffic management and optimization. ML models were categorized into 11 groups and investigated for nine ITS-related tasks. Our survey demonstrated that DL techniques significantly contributed to the perception layer of UAV-aided ITS, while RL algorithm was employed in decision layer and optimization tasks. Moreover, in this study, we conducted a thorough experimental comparison of 14 learners with the combination of eight DL models of CNN, R-CNN, Faster R-CNN, and YOLO versions 1-5 with different backbone architectures on five UAV datasets. The findings demonstrated that the characteristics of the datasets, particularly the attitude of captured images/videos, largely affected the performance of DL algorithms. Furthermore, despite their performance for detection and classification tasks, YOLOv5+ResNet outperformed the other DL models in terms of average recall and F1-Score, 81.47% and 82.2%, respectively. On the other hand, YOLOv4+CSPDarknet53 provided a superiority over the others for the average precision, 88.5%. Regarding efficiency, the CNN+ResNet model provided the worst average FPS and inference time.

Despite many challenges ahead, the study showcases the potential of DL models in accurately estimating traffic parameters, emphasizing the importance of considering appropriate measurement intervals for speed estimation. The results presented for the comparative study also demonstrate the effectiveness of the YOLO-based models, particularly YOLOv5+ResNet, across diverse datasets. The YOLO architecture's unique grid-based prediction and streamlined design contribute to its superior balance between accuracy and speed, crucial for real-world applications in dynamic environments like vehicle detection. The study also delves into the challenges associated with the application of these techniques, such as limited computational resources, data quality and diversity, adaptability to dynamic environments, and privacy concerns. These challenges underscore the need for ongoing research and development efforts to optimize algorithms, enhance adaptability, and address privacy considerations for widespread deployment in ITS.

Acknowledgment

The authors extend their appreciation to Sky Shine Innovation and HoverUAV for their valuable contributions and collaboration.

References

[1] R. Jabbar et al., "Blockchain technology for intelligent transportation systems: A systematic literature review," *IEEE Access*, vol. 10, pp. 20995–21031, 2022.

[2] H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer, "UAV-enabled intelligent transportation systems for the smart city: Applications and challenges," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 22–28, Mar. 2017.

[3] M. A. Khan et al., "A provable and privacy-preserving authentication scheme for UAV-enabled intelligent transportation systems," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3416–3425, May 2022.

[4] Y. M. Chen, L. Dong, and J.-S. Oh, "Real-time video relay for UAV traffic surveillance systems through available communication networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Mar. 2007, pp. 2608–2612.

[5] L. Deng, G. Wu, J. Fu, Y. Zhang, and Y. Yang, "Joint resource allocation and trajectory control for UAV-enabled vehicular communications," *IEEE Access*, vol. 7, pp. 132806–132815, 2019.

[6] *Irma: A Hurricane for the History Books*, CNN, Atlanta, GA, USA, 2017.

[7] M. Samir, M. Chraiti, C. Assi, and A. Ghrayeb, "Joint optimization of UAV trajectory and radio resource allocation for drive-thru vehicular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.

[8] A. Raza, S. H. R. Bukhari, F. Aadil, and Z. Iqbal, "An UAV-assisted VANET architecture for intelligent transportation system in smart cities," *Int. J. Distrib. Sensor Netw.*, vol. 17, no. 7, 2021, Art. no. 15501477211031750.

[9] M. Samir, C. Assi, S. Sharafeddine, D. Ebrahimi, and A. Ghrayeb, "Age of information aware trajectory planning of UAVs in intelligent transportation systems: A deep learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12382–12395, Nov. 2020.

[10] T. Tang, S. Zhou, Z. Deng, L. Lei, and H. Zou, "Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks," *Remote Sens.*, vol. 9, no. 11, p. 1170, Nov. 2017.

[11] M. Samir, D. Ebrahimi, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Leveraging UAVs for coverage in cell-free vehicular networks: A deep reinforcement learning approach," *IEEE Trans. Mobile Comput.*, vol. 20, no. 9, pp. 2835–2847, Sep. 2021.

[12] K. Kanistras, G. Martins, M. J. Rutherford, and K. P. Valavanis, "A survey of unmanned aerial vehicles (UAVs) for traffic monitoring," in *Proc. Int. Conf. Unmanned Aircr. Syst. (ICUAS)*, May 2013, pp. 221–234.

[13] F. Outay, H. A. Mengash, and M. Adnan, "Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges," *Transp. Res. A, Policy Pract.*, vol. 141, pp. 116–129, Nov. 2020.

[14] A. Gupta, T. Afrin, E. Scully, and N. Yodo, "Advances of UAVs toward future transportation: The state-of-the-art, challenges, and opportunities," *Future Transp.*, vol. 1, no. 2, pp. 326–350, Sep. 2021.

[15] A. Saboor, S. Coene, E. Vinogradov, E. Tanghe, W. Joseph, and S. Pollin, "Elevating the future of mobility: UAV-enabled intelligent transportation systems," 2021, *arXiv:2110.09934*.

[16] S. Iftikhar et al., "Target detection and recognition for traffic congestion in smart cities using deep learning-enabled UAVs: A review and analysis," *Appl. Sci.*, vol. 13, no. 6, p. 3995, Mar. 2023.

[17] S. Liu et al., "Performance analysis of vehicle detection algorithm in aerial traffic videos," in *Proc. Int. Conf. Virtual Reality Vis. (ICVRV)*, Nov. 2019, pp. 59–64.

[18] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, Jun. 2011.

[19] B. Benjdira, T. Khursheed, A. Koubaa, A. Ammar, and K. Ouni, "Car detection using unmanned aerial vehicles: Comparison between faster R-CNN and YOLOv3," in *Proc. 1st Int. Conf. Unmanned Vehicle Syst.-Oman (UVS)*, Feb. 2019, pp. 1–6.

[20] A. Ammar, A. Koubaa, M. Ahmed, A. Saad, and B. Benjdira, "Vehicle detection from aerial images using deep learning: A comparative study," *Electronics*, vol. 10, no. 7, p. 820, Mar. 2021.

[21] L. Tan, X. Lv, X. Lian, and G. Wang, "YOLOv4_Drone: UAV image target detection based on an improved YOLOv4 algorithm," *Comput. Electr. Eng.*, vol. 93, Jul. 2021, Art. no. 107261.

[22] X. Luo, Y. Wu, and L. Zhao, "YOLOD: A target detection method for UAV aerial imagery," *Remote Sens.*, vol. 14, no. 14, p. 3240, Jul. 2022.

[23] H. Gupta and O. P. Verma, "Monitoring and surveillance of urban road traffic using low altitude drone images: A deep learning approach," *Multimedia Tools Appl.*, vol. 81, no. 14, pp. 19683–19703, Jun. 2022.

[24] A. K. Haghighat, V. Ravichandra-Mouli, P. Chakraborty, Y. Esfandiari, S. Arabi, and A. Sharma, "Applications of deep learning in intelligent transportation systems," *J. Big Data Anal. Transp.*, vol. 2, no. 2, pp. 115–145, 2020.

[25] D. Gangwani and P. Gangwani, "Applications of machine learning and artificial intelligence in intelligent transportation system: A review," in *Proc. Appl. Artif. Intell. Mach. Learn. (ICAAAIML)*, 2020, pp. 203–216.

[26] A. E. Ezugwu et al., "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Eng. Appl. Artif. Intell.*, vol. 110, Apr. 2022, Art. no. 104743.

[27] I. Aljarah, M. Mafarja, A. A. Heidari, H. Faris, and S. Mirjalili, "Clustering analysis using a novel locality-informed grey wolf-inspired clustering approach," *Knowl. Inf. Syst.*, vol. 62, no. 2, pp. 507–539, Feb. 2020.

[28] A. Telikani, A. Tahmassebi, W. Banzhaf, and A. H. Gandomi, "Evolutionary machine learning: A survey," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–35, 2021.

[29] S. Pouyanfar et al., "A survey on deep learning: Algorithms, techniques, and applications," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–36, Sep. 2018.

[30] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022.

[31] J. Gu et al., "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[32] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587.

[33] F. Piccialli, V. D. Somma, F. Giampaolo, S. Cuomo, and G. Fortino, "A survey on deep learning in medicine: Why, how and when?" *Inf. Fusion*, vol. 66, pp. 111–137, Feb. 2021.

[34] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2016.

[36] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[37] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[38] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[39] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[40] S. Pateria, B. Subagdja, A.-H. Tan, and C. Quek, "Hierarchical reinforcement learning: A comprehensive survey," *ACM Comput. Surv.*, vol. 54, no. 5, pp. 1–35, 2021.

[41] T. Zhao and R. Nevatia, "Car detection in low resolution aerial images," *Image Vis. Comput.*, vol. 21, no. 8, pp. 693–703, Aug. 2003.

[42] H. Yalcin, M. Hebert, R. Collins, and M. J. Black, "A flow-based approach to vehicle detection and background mosaicking in airborne video," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, p. 1202.

[43] H.-Y. Cheng, C.-C. Weng, and Y.-Y. Chen, "Vehicle detection in aerial surveillance using dynamic Bayesian networks," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2152–2159, Apr. 2012.

[44] W. Shao, W. Yang, G. Liu, and J. Liu, "Car detection from high-resolution aerial imagery using multiple features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 4379–4382.

[45] Y. Ma, X. Wu, G. Yu, Y. Xu, and Y. Wang, "Pedestrian detection and tracking from low-resolution unmanned aerial vehicle thermal imagery," *Sensors*, vol. 16, no. 4, p. 446, Mar. 2016.

[46] X. Cao, C. Wu, P. Yan, and X. Li, "Linear SVM classification using boosting HOG features for vehicle detection in low-altitude airborne videos," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2421–2424.

[47] K. V. Najiya and M. Archana, "UAV video processing for traffic surveillence with enhanced vehicle detection," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Apr. 2018, pp. 662–668.

[48] H. Zhou, L. Wei, C. P. Lim, D. Creighton, and S. Nahavandi, "Robust vehicle detection in aerial images using bag-of-words and orientation aware scanning," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7074–7085, Dec. 2018.

[49] Y. Li, "Vehicle extraction using histogram and genetic algorithm based fuzzy image segmentation from high resolution UAV aerial imagery," in *Proc. ISPRS*, vol. 529, 2008, p. B3b.

[50] Á. F. García-Fernández and J. Xiao, "Trajectory Poisson multi-Bernoulli mixture filter for traffic monitoring using a drone," *IEEE Trans. Veh. Technol.*, vol. 73, no. 1, pp. 402–413, Jan. 2024.

[51] R. Ke, S. Feng, Z. Cui, and Y. Wang, "Advanced framework for microscopic and lane-level macroscopic traffic parameters estimation from UAV video," *IET Intell. Transp. Syst.*, vol. 14, no. 7, pp. 724–734, Jul. 2020.

[52] R. Ke, S. Kim, Z. Li, and Y. Wang, "Motion-vector clustering for traffic speed detection from UAV video," in *Proc. IEEE 1st Int. Smart Cities Conf. (ISC2)*, Oct. 2015, pp. 1–5.

[53] R. Ke, Z. Li, S. Kim, J. Ash, Z. Cui, and Y. Wang, "Real-time bidirectional traffic flow parameter estimation from aerial videos," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 890–901, Apr. 2017.

[54] A. Gomaa, M. M. Abdelwahab, and M. Abo-Zahhad, "Real-time algorithm for simultaneous vehicle detection and tracking in aerial view videos," in *Proc. IEEE 61st Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2018, pp. 222–225.

[55] A. Gomaa, M. M. Abdelwahab, and M. Abo-Zahhad, "Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis," *Multimedia Tools Appl.*, vol. 79, nos. 35–36, pp. 26023–26043, Sep. 2020.

[56] Z. Chen et al., "Vehicle detection in high-resolution aerial images based on fast sparse representation classification and multiorder feature," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2296–2309, Aug. 2016.

[57] J.-Y. Choi and Y.-K. Yang, "Vehicle detection from aerial images using local shape information," in *Proc. 3rd Pacific Rim Symp. Adv. Image Video Technol. (PSIVT)*, Tokyo, Japan. Springer, Jan. 2009, pp. 227–236.

[58] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Dec. 2001, p. I.

[59] A. Gaszczak, T. P. Breckon, and J. Han, "Real-time people and vehicle detection from UAV imagery," *Proc. SPIE*, vol. 7878, pp. 71–83, Jan. 2011.

[60] T. P. Breckon, S. E. Barnes, M. L. Eichner, and K. Wahren, "Autonomous real-time vehicle detection from a medium-level UAV," in *Proc. 24th Int. Conf. Unmanned Air Vehicle Syst.* Princeton, NJ, USA: Citeseer, 2009, p. 29.

[61] H. Niu, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "A UAV-based traffic monitoring system—Invited paper," in *Proc. IEEE 87th Veh. Technol. Conf. (VTC Spring)*, Jun. 2018, pp. 1–5.

[62] Y. Xu, G. Yu, Y. Wang, and X. Wu, "Vehicle detection and tracking from airborne images," in *Proc. CICTP*, Jul. 2015, pp. 641–649.

[63] X. Cao, C. Wu, J. Lan, P. Yan, and X. Li, "Vehicle detection and motion analysis in low-altitude airborne video under urban environment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 10, pp. 1522–1533, Oct. 2011.

[64] J. Leitloff, D. Rosenbaum, F. Kurz, O. Meynberg, and P. Reinartz, "An operational system for estimating road traffic information from aerial images," *Remote Sens.*, vol. 6, no. 11, pp. 11315–11341, Nov. 2014.

[65] T. Moranduzzo and F. Melgani, "Detecting cars in UAV images with a catalog-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6356–6367, Oct. 2014.

[66] T. Moranduzzo and F. Melgani, "A SIFT-SVM method for detecting cars in UAV images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 6868–6871.

[67] T. Moranduzzo and F. Melgani, "Automatic car counting method for unmanned aerial vehicle images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1635–1647, Mar. 2014.

[68] Y. Xu, G. Yu, X. Wu, Y. Wang, and Y. Ma, "An enhanced Viola–Jones vehicle detection method from unmanned aerial vehicles imagery," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1845–1856, Jul. 2017.

[69] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "A hybrid vehicle detection method based on viola-jones and HOG + SVM from UAV images," *Sensors*, vol. 16, no. 8, p. 1325, Aug. 2016.

[70] G. Hu, C. Yin, M. Wan, Y. Zhang, and Y. Fang, "Recognition of diseased pinus trees in UAV images using deep learning and AdaBoost classifier," *Biosyst. Eng.*, vol. 194, pp. 138–151, Jun. 2020.

[71] H. Grabner, T. T. Nguyen, B. Gruber, and H. Bischof, "On-line boosting-based car detection from aerial images," *ISPRS J. Photogramm. Remote Sens.*, vol. 63, no. 3, pp. 382–396, May 2008.

[72] K. Liu and G. Mattyus, "Fast multiclass vehicle detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1938–1942, Sep. 2015.

[73] T. T. Nguyen, H. Grabner, H. Bischof, and B. Gruber, "On-line boosting for car detection from aerial images," in *Proc. IEEE Int. Conf. Res., Innov. Vis. Future*, Mar. 2007, pp. 87–95.

[74] S. Kluckner, G. Pacher, H. Grabner, H. Bischof, and J. Bauer, "A 3D teacher for car detection in aerial images," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[75] S. Tuermer, F. Kurz, P. Reinartz, and U. Stilla, "Airborne vehicle detection in dense urban areas using HoG features and disparity maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2327–2337, Dec. 2013.

[76] V. Mahajan, E. Barmpounakis, M. R. Alam, N. Geroliminis, and C. Antoniou, "Treating noise and anomalies in vehicle trajectories from an experiment with a swarm of drones," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 9, pp. 9055–9067, May 2023.

[77] R. W. Schafer, "What is a Savitzky–Golay filter? [Lecture notes]," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, Jul. 2011.

[78] X.-F. Liu, L.-M. Gao, Z.-W. Guang, and Y.-Q. Song, "A UAV allocation method for traffic surveillance in sparse road network," *J. Highway Transp. Res. Develop., English Ed.*, vol. 7, no. 2, pp. 81–87, Jun. 2013.

[79] H. Ghazzai, H. Menouar, and A. Kadri, "On the placement of UAV docking stations for future intelligent transportation systems," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–6.

[80] C. Kyrkou, S. Timotheou, P. Kolios, T. Theocharides, and C. G. Panayiotou, "Optimized vision-directed deployment of UAVs for rapid traffic monitoring," in *Proc. IEEE Int. Conf. Consum. Electron. (ICCE)*, Jan. 2018, pp. 1–6.

[81] T. Qu, Q. Zhang, and S. Sun, "Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21651–21663, Oct. 2017.

[82] R. Zhang, S. Newsam, Z. Shao, X. Huang, J. Wang, and D. Li, "Multi-scale adversarial network for vehicle detection in UAV imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 180, pp. 283–295, Oct. 2021.

[83] F. Shi, T. Zhang, and T. Zhang, "Orientation-aware vehicle detection in aerial images via an anchor-free object detection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5221–5233, Jun. 2021.

[84] X. Li, X. Li, and H. Pan, "Multi-scale vehicle detection in high-resolution aerial images with context information," *IEEE Access*, vol. 8, pp. 208643–208657, 2020.

[85] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[86] S. Kim, G. Anagnostopoulos, E. Barmpounakis, and N. Geroliminis, "Visual extensions and anomaly detection in the pNEUMA experiment with a swarm of drones," *Transp. Res. C, Emerg. Technol.*, vol. 147, Feb. 2023, Art. no. 103966.

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

[88] E. I. Vlahogianni, J. D. Ser, K. Kepaptsoglou, and I. La na, "Model free identification of traffic conditions using unmanned aerial vehicles and deep learning," *J. Big Data Anal. Transp.*, vol. 3, no. 1, pp. 1–13, Apr. 2021.

[89] W. Li, H. Li, Q. Wu, X. Chen, and K. N. Ngan, "Simultaneously detecting and counting dense vehicles from drone images," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9651–9662, Dec. 2019.

[90] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, pp. 154–171, Apr. 2013.

[91] Y. Youssef and M. Elshenawy, "Automatic vehicle counting and tracking in aerial video feeds using cascade region-based convolutional neural networks and feature pyramid networks," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2675, no. 8, pp. 304–317, Aug. 2021.

[92] Z. Deng, H. Sun, S. Zhou, J. Zhao, and H. Zou, "Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3652–3664, Aug. 2017.

[93] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "Car detection from low-altitude UAV imagery with the faster R-CNN," *J. Adv. Transp.*, vol. 2017, no. 1, pp. 1–10, 2017.

[94] M. Wang, X. Luo, X. Wang, and X. Tian, "Research on vehicle detection based on faster R-CNN for UAV images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 1177–1180.

[95] T. Tang, S. Zhou, Z. Deng, H. Zou, and L. Lei, "Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining," *Sensors*, vol. 17, no. 2, p. 336, 2017.

[96] E.-J. Kim, H.-C. Park, S.-W. Ham, S.-Y. Kho, and D.-K. Kim, "Extracting vehicle trajectories using unmanned aerial vehicles in congested traffic conditions," *J. Adv. Transp.*, vol. 2019, pp. 1–16, Apr. 2019.

[97] X. Kong, Y. Zhang, S. Tu, C. Xu, and W. Yang, "Vehicle detection in high-resolution aerial images with parallel RPN and density-assigner," *Remote Sens.*, vol. 15, no. 6, p. 1659, Mar. 2023.

[98] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–2969.

[99] H. Zhang, M. Liptrott, N. Bessis, and J. Cheng, "Real-time traffic analysis using deep learning techniques and UAV based video," in *Proc. 16th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, Sep. 2019, pp. 1–5.

[100] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[101] J. Zhu et al., "Urban traffic density estimation based on ultrahigh-resolution UAV video and deep neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4968–4981, Dec. 2018.

[102] L. Mou and X. X. Zhu, "Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6699–6711, Nov. 2018.

[103] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[104] N. Audebert, B. Le Saux, and S. Lefèvre, "Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images," *Remote Sens.*, vol. 9, no. 4, p. 368, Apr. 2017.

[105] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery: A small target detection benchmark," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 187–203, Jan. 2016.

[106] M. Masuduzzaman, A. Islam, K. Sadia, and S. Y. Shin, "UAV-based MEC-assisted automated traffic management scheme using blockchain," *Future Gener. Comput. Syst.*, vol. 134, pp. 256–270, Sep. 2022.

[107] T. Tang, Z. Deng, S. Zhou, L. Lei, and H. Zou, "Fast vehicle detection in UAV images," in *Proc. Int. Workshop Remote Sens. Intell. Process. (RSIP)*, May 2017, pp. 1–5.

[108] J. Liu, J. Wu, and M. Liu, "UAV monitoring and forecasting model in intelligent traffic oriented applications," *Comput. Commun.*, vol. 153, pp. 499–506, Mar. 2020.

[109] R. Feng, C. Fan, Z. Li, and X. Chen, "Mixed road user trajectory extraction from moving aerial videos based on convolution neural network detection," *IEEE Access*, vol. 8, pp. 43508–43519, 2020.

[110] B. Benjdira, A. Koubaa, A. T. Azar, Z. Khan, A. Ammar, and W. Boulila, "TAU: A framework for video-based traffic analytics leveraging artificial intelligence and unmanned aerial systems," *Eng. Appl. Artif. Intell.*, vol. 114, Sep. 2022, Art. no. 105095.

[111] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3645–3649.

[112] J. Li, S. Chen, F. Zhang, E. Li, T. Yang, and Z. Lu, "An adaptive framework for multi-vehicle ground speed estimation in airborne videos," *Remote Sens.*, vol. 11, no. 10, p. 1241, May 2019.

[113] Y. Chen, D. Zhao, M. J. Er, Y. Zhuang, and H. Hu, "A novel vehicle tracking and speed estimation with varying UAV altitude and video resolution," *Int. J. Remote Sens.*, vol. 42, no. 12, pp. 4441–4466, Jun. 2021.

[114] D. Wu, S. Lv, M. Jiang, and H. Song, "Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments," *Comput. Electron. Agricult.*, vol. 178, Nov. 2020, Art. no. 105742.

[115] M. A. Momin, M. H. Junos, A. S. M. Khairuddin, and M. S. A. Talip, "Lightweight CNN model: Automated vehicle detection in aerial images," *Signal, Image Video Process.*, vol. 17, no. 4, pp. 1209–1217, Jun. 2023.

[116] H. V. Koay, J. H. Chuah, C.-O. Chow, Y.-L. Chang, and K. K. Yong, "YOLO-RTUAV: Towards real-time vehicle detection through aerial images with low-cost edge devices," *Remote Sens.*, vol. 13, no. 21, p. 4196, Oct. 2021.

[117] J. Feng and C. Yi, "Lightweight detection network for arbitrary-oriented vehicles in UAV imagery via global attentive relation and multi-path fusion," *Drones*, vol. 6, no. 5, p. 108, Apr. 2022.

[118] M. H. Hamzenejadi and H. Mohseni, "Fine-tuned YOLOv5 for real-time vehicle detection in UAV imagery: Architectural improvements and performance boost," *Expert Syst. Appl.*, vol. 231, Nov. 2023, Art. no. 120845.

[119] J. Feng, J. Wang, and R. Qin, "Lightweight detection network for arbitrary-oriented vehicles in UAV imagery via precise positional information encoding and bidirectional feature fusion," *Int. J. Remote Sens.*, vol. 44, no. 15, pp. 4529–4558, Aug. 2023.

[120] Y. M. Bhavsar, M. S. Zaveri, M. S. Raval, and S. B. Zaveri, "Vision-based investigation of road traffic and violations at urban roundabout in India using UAV video: A case study," *Transp. Eng.*, vol. 14, Dec. 2023, Art. no. 100207.

[121] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[122] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3464–3468.

[123] B. Ma, Z. Liu, F. Jiang, Y. Yan, J. Yuan, and S. Bu, "Vehicle detection in aerial images using rotation-invariant cascaded forest," *IEEE Access*, vol. 7, pp. 59613–59623, 2019.

[124] Z.-H. Zhou and J. Feng, "Deep forest," *Nat. Sci. Rev.*, vol. 6, no. 1, pp. 74–86, Jan. 2019.

[125] K. Zheng, M. Wei, G. Sun, B. Anas, and Y. Li, "Using vehicle synthesis generative adversarial networks to improve vehicle detection in remote sensing images," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 9, p. 390, Sep. 2019.

[126] R. Makrigiorgis, N. Hadjittoouli, C. Kyrkou, and T. Theocharides, "AirCamRTM: Enhancing vehicle detection for efficient aerial camera-based road traffic monitoring," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 2119–2128.

[127] R. Jonker and T. Volgenant, "Improving the Hungarian assignment algorithm," *Oper. Res. Lett.*, vol. 5, no. 4, pp. 171–175, 1986.

[128] G. Bishop and G. Welch, "An introduction to the Kalman filter," in *Proc. SIGGRAPH, Course*, vo. 8, nos. 27599–23175, 2001, p. 41.

[129] W. Y. B. Lim et al., "Towards federated learning in UAV-enabled Internet of Vehicles: A multi-dimensional contract-matching approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 5140–5154, Aug. 2021.

[130] B. Jiang, S. N. Givigi, and J.-A. Delamer, "A MARL approach for optimizing positions of VANET aerial base-stations on a sparse highway," *IEEE Access*, vol. 9, pp. 133989–134004, 2021.

[131] S. Liu and Y. Bai, "Multiple UAVs collaborative traffic monitoring with intention-based communication," *Comput. Commun.*, vol. 210, pp. 116–129, Oct. 2023.

[132] Z. Niu, X. Shen, Q. Zhang, and Y. Tang, "Space-air-ground integrated vehicular network for connected and automated vehicles: Challenges and solutions," *Intell. Converg. Netw.*, vol. 1, no. 2, pp. 142–169, Sep. 2020.

[133] L. Zhao et al., "Vehicular computation offloading for industrial mobile edge computing," *IEEE Trans. Ind. Informat.*, vol. 17, no. 11, pp. 7871–7881, 2021.

[134] H. Peng, H. Wu, and X. S. Shen, "Edge intelligence for multi-dimensional resource management in aerial-assisted vehicular networks," *IEEE Wireless Commun.*, vol. 28, no. 5, pp. 59–65, Oct. 2021.

[135] H. Peng and X. S. Shen, "DDPG-based resource management for MEC/UAV-assisted vehicular networks," in *Proc. IEEE 92nd Veh. Technol. Conf. (VTC-Fall)*, Nov. 2020, pp. 1–6.

[136] H. Peng and X. Shen, "Multi-agent reinforcement learning based resource management in MEC-and UAV-assisted vehicular networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 131–141, Jan. 2020.

[137] Y. Su, M. Liwang, S. Hosseinalipour, L. Huang, and H. Dai, "Cooperative relaying and power control for UAV-assisted vehicular networks with deep Q-network," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Jul. 2021, pp. 318–323.

[138] M. Zhu, X.-Y. Liu, and A. Walid, "Deep reinforcement learning for unmanned aerial vehicle-assisted vehicular networks," 2019, *arXiv:1906.05015*.

[139] C. Yang, B. Liu, H. Li, B. Li, K. Xie, and S. Xie, "Learning based channel allocation and task offloading in temporary UAV-assisted vehicular edge computing networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9884–9895, Sep. 2022.

[140] Y. He, D. Zhai, Y. Jiang, and R. Zhang, "Relay selection for UAV-assisted urban vehicular ad hoc networks," *IEEE Wireless Commun. Lett.*, vol. 9, no. 9, pp. 1379–1383, Sep. 2020.

[141] A. Al-Hilo, M. Samir, C. Assi, S. Sharafeddine, and D. Ebrahimi, "A cooperative approach for content caching and delivery in UAV-assisted vehicular networks," *Veh. Commun.*, vol. 32, Dec. 2021, Art. no. 100391.

[142] L. Spampinato, A. Tarozzi, C. Buratti, and R. Marini, "DRL path planning for UAV-aided V2X networks: Comparing discrete to continuous action spaces," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[143] R. Marini, S. Park, O. Simeone, and C. Buratti, "Continual meta-reinforcement learning for UAV-aided vehicular wireless networks," in *Proc. IEEE Int. Conf. Commun.*, May 2023, pp. 5664–5669.

[144] R. Marini, L. Spampinato, S. Mignardi, R. Verdone, and C. Buratti, "Reinforcement learning-based trajectory planning for UAV-aided vehicular communications," in *Proc. 30th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2022, pp. 967–971.

[145] M. Samir, D. Ebrahimi, C. Assi, S. Sharafeddine, and A. Ghrayeb, "Trajectory planning of multiple dronecells in vehicular networks: A reinforcement learning approach," *IEEE Netw. Lett.*, vol. 2, no. 1, pp. 14–18, Mar. 2020.

[146] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.

[147] A. Al-Hilo, M. Samir, C. Assi, S. Sharafeddine, and D. Ebrahimi, "UAV-assisted content delivery in intelligent transportation systems-joint trajectory planning and cache management," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 8, pp. 5155–5167, Aug. 2021.

[148] O. S. Oubbati, M. Atiquzzaman, A. Baz, H. Alhakami, and J. Ben-Othman, "Dispatch of UAVs for urban vehicular networks: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 13174–13189, Dec. 2021.

[149] Z. Wu, Z. Yang, C. Yang, J. Lin, Y. Liu, and X. Chen, "Joint deployment and trajectory optimization in UAV-assisted vehicular edge computing networks," *J. Commun. Netw.*, vol. 24, no. 1, pp. 47–58, Feb. 2022.

[150] Y. H. Chan and L. C. Lau, "On linear and semidefinite programming relaxations for hypergraph matching," *Math. Program.*, vol. 135, nos. 1–2, pp. 123–148, Oct. 2012.

[151] W. Qi, Q. Song, L. Guo, and A. Jamalipour, "Energy-efficient resource allocation for UAV-assisted vehicular networks with spectrum sharing," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7691–7702, Jul. 2022.

[152] M. Ester et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, vol. 96, 1996, pp. 226–231.

[153] C. K. Toth and D. Grejner-Brzezinska, "Extracting dynamic spatial data from airborne imaging sensors to support traffic flow estimation," *ISPRS J. Photogramm. Remote Sens.*, vol. 61, nos. 3–4, pp. 137–148, Dec. 2006.

[154] J. Zhao, Y. Nie, H. Zhang, and F. R. Yu, "A UAV-aided vehicular integrated platooning network for heterogeneous resource management," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 1, pp. 512–521, Mar. 2023.

[155] H. Sun and G. Xing, "A YOLOv4-based vehicle detection method from UAV videos," in *Proc. 33rd Chin. Control Decis. Conf. (CCDC)*, May 2021, pp. 3082–3087.

[156] X. Luo et al., "Fast automatic vehicle detection in UAV images using convolutional neural networks," *Remote Sens.*, vol. 12, no. 12, p. 1994, Jun. 2020.

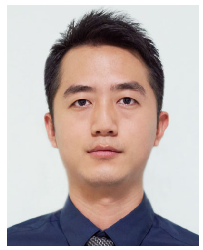
[157] N. Ammour, H. Alhichri, Y. Bazi, B. Benjdira, N. Alajlan, and M. Zuair, "Deep learning approach for car detection in UAV imagery," *Remote Sens.*, vol. 9, no. 4, p. 312, Mar. 2017.

[158] H. Wu, F. Lyu, C. Zhou, J. Chen, L. Wang, and X. Shen, "Optimal UAV caching and trajectory in aerial-assisted vehicular networks: A learning-based approach," *IEEE J. Sel. Areas. Commun.*, vol. 38, no. 12, pp. 2783–2797, Dec. 2020.

[159] M. ElMikaty and T. Stathaki, "Car detection in aerial images of dense urban areas," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 1, pp. 51–63, Feb. 2018.
[160] M. ElMikaty and T. Stathaki, "Detection of cars in high-resolution aerial images of complex urban environments," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5913–5924, Oct. 2017.
[161] Z. Ghasemi Darehnaei, M. Shokouhifar, H. Yazdanjouei, and S. M. J. R. Fatemi, "Si-EDTL: Swarm intelligence ensemble deep transfer learning for multiple vehicle detection in UAV images," *Concurrency Comput., Pract. Exp.*, vol. 34, no. 5, Feb. 2022, Art. no. e6726.
[162] Z. G. Darehnaei, S. M. J. R. Fatemi, S. M. Mirhassani, and M. Fouladian, "Ensemble deep learning using faster R-CNN and genetic algorithm for vehicle detection in UAV images," *IETE J. Res.*, vol. 69, no. 8, pp. 5102–5111, Sep. 2023.
[163] E. Frew et al., "Vision-based road-following using a small autonomous aircraft," in *Proc. IEEE Aerosp. Conf.*, vol. 5, Mar. 2004, pp. 3006–3015.
[164] T. Moranduzzo, F. Melgani, Y. Bazi, and N. Alajlan, "A fast object detector based on high-order gradients and Gaussian process regression for UAV images," *Int. J. Remote Sens.*, vol. 36, no. 10, pp. 2713–2733, May 2015.
[165] J. Gleason, A. V. Nefian, X. Bouyssounousse, T. Fong, and G. Bebis, "Vehicle detection from aerial imagery," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 2065–2070.
[166] N. Bodla, B. Singh, R. Chellappa, and L. S. Davis, "Soft-NMS—Improving object detection with one line of code," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 5562–5570.
[167] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory prediction in crowded scenes," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 2, 2016, p. 5.
[168] C. Tomasi and T. Kanade, "Detection and tracking of point," *Int. J. Comput. Vis.*, vol. 9, nos. 137–154, p. 3, 1991.
[169] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

Akbar Telikani (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in computer engineering and information technology in 2012 and 2015, respectively. He is currently pursuing the Ph.D. degree with the University of Wollongong, Australia. His research interests include intelligent transportation systems, privacy and security, ML, the Internet of Things (IoT), and evolutionary computation. His main focus is to develop UAVs-aided ITS and security protection mechanisms for the IoT.

Arupa Sarkar received the B.Eng. and M.Eng. degrees in electrical engineering and the Master of Information Systems (M.I.S.) degree. She is currently pursuing the Ph.D. degree with the University of Wollongong, Australia. Her research interests include privacy and security, machine learning (ML), unmanned aerial vehicles (UAVs), and fuzzy logic. Her central focus in her research is to identify and mitigate battery depletion attacks on UAVs using artificial intelligence (AI).



Bo Du received the Ph.D. degree from Nanyang Technological University, Singapore. He is currently a Senior Lecturer with Griffith University, Australia. His research interests include spanning transportation and logistics system modeling and optimization, zero-emission and emerging mobility solution design, data analytics, and data-driven modeling. He is an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Jun Shen (Senior Member, IEEE) received the Ph.D. degree from Southeast University. Then, he worked at Australian universities. He was a Visiting Professor with MIT and GaTech. He is currently a Full Professor with the University of Wollongong. He has established expertise in multi-domain computational intelligence applications, including education, transport, manufacturing, and bioinformatics. He has published more than 320 papers in relevant fields, including IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, IEEE TRANSACTIONS ON SERVICES COMPUTING, BIB, ICML, KDD, and BJET. He is a Senior Member of ACM. He was a Panel Member of ACM/AIS MSIS2016. He is the PC chair and a member of more than 400 international conferences. He is an associate editor of four tier 1 international journals and an editorial member of other eight journals.