AI-Powered CPS-Enabled Urban Transportation Digital Twin: Methods and Applications

Yongjie Fu, Mehmet K.Turkcan‡, Mahshid Ghasemi‡, Zhaobin Mo, Chengbo Zang, Abhishek Adhikari, Zoran Kostic, Gil Zussman, Xuan Di*, *Member, IEEE*

***Abstract*—We present methods and applications for the development of digital twins (DT) for urban traffic management. While the majority of studies on the DT focus on its "eyes," which is the emerging sensing and perception like object detection and tracking, what really distinguishes the DT from a traditional simulator lies in its "brain," the prediction and decision making capabilities of extracting patterns and making informed decisions from what has been seen and perceived. In order to add value to urban transportation management, DTs need to be powered by artificial intelligence and complement with low-latency high-bandwidth sensing and networking technologies, in other words, cyberphysical systems (CPS). We will first review the DT pipeline enabled by CPS and propose our DT architecture deployed on a real-world testbed in New York City. This paper can be a pointer to help researchers and practitioners identify challenges and opportunities for the development of DTs; a bridge to initiate conversations across disciplines; and a road map to exploiting potentials of DTs for diverse urban transportation applications.**

***Index Terms*—Digital twin, AI, Urban traffic management**

I. Introduction

URBAN transportation systems are complex to model and simulate, due to heterogeneous road users (such as cars, pedestrians, cyclists, scooters) interacting in multimodal traffic environments consisting of public and private travel modes. With fast-changing traffic evolution in time and space, traffic simulation, if improperly calibrated, might produce traffic management strategies that largely deviate from the reality, potentially leading to suboptimal or even detrimental outcomes. With ubiquitous sensors in smart cities, it is the time to *augment* conventional traffic simulators, many of which were developed in an era when only "small data" became available. Emerging traffic sensors are expected to generate big volumes of data, transmitted via communication networks and processed on edge cloud computing with artificial intelligence (AI) for real-time traffic management. Such a transformation calls for the development of a new paradigm, namely, digital twin (DT), which will push the envelope in urban transportation management.

Literally, DT is the digital replica of a physical object or asset [1], where a digital world mirrors a physical world for real-time diagnosis, prognosis, and decision making. Recent years have seen a growing amount of studies on DT in various domains [1], [2], [3], including a sizable body of articles on vehicular DTs [4], [5], [6], [7], [8].

With recent explosive growth of literature on DTs, we would like to restrict the scope of this paper to applications in the urban setting, especially when vulnerable road users (VRU) (i.e., non-motorists such as pedestrian, bicyclists, other cyclists, or persons on personal conveyance [9]) are an integral part of the system and also potential users of the DT. We will primarily focus on use cases accounting for VRUs with improved traffic safety and efficiency.

The studies of DTs for urban traffic management, especially involving VRUs, are lacking, partly because the development of a DT for a system is non-trivial, particularly when involved with humans.

This paper presents methods and applications for the development of DTs for urban transportation systems. We will depict a DT pipeline prototype, leveraging the architecture of cyberphysical systems and AI methods. We will propose a DT for real-time traffic monitoring and optimization, based on an existing physical testbed deployed in New York City (NYC) leveraging cutting-edge sensing, communication, computing, and AI-based automation. The overall contributions of this paper include: (1) introducing AI methods used in the DT pipeline; (2) exploring the architecture of transportation DTs and propose a prototype for reference; and (3) identifying gaps and directions.

The rest of this paper is organized below. Section II introduces transportation DTs enabled by cyberphysical systems, and position this paper; Section III reviews the literature along the pipeline of a DT. Section IV demonstrates the architecture of our DT, building on a real-world testbed. Section V concludes our work, presents potential research directions and open questions.

II. Preliminaries

The definition of DT has evolved rapidly. Despite presenting its own version, articles share common elements and resemble certain characteristics [1], [10], [11]. In general, there is a physical world (aka. the physical) and a digital world (aka. the digital). The physical world evolves in time and space. To ensure that the physical system is run in a desired direction, it requires close monitoring, operation, and management. Thus, the role a digital world plays is to model and simulate the dynamics of the physical world in a synchronized fashion, so that the digital can also predict the future states of the physical precisely, which offers a ground for optimal decision making. The physical and the digital exchange data and information

Manuscript received month day, year; revised month day, year.

This work was supported by NSF CPS-2038984, ERC-2133516. We thank Mengxuan Liu for her assistance in generating 3D models.

*Corresponding author: Xuan Di (Email: sharon.di@columbia.edu).

‡ Mehmet K.Turkcan and Mahshid Ghasemi contribute equally.

Fu, Turkcan, Mo, and Di are with the Department of Civil Engineering and Engineering Mechanics. Ghasemi, Zang, Adhikari, Kostic, and Zussman are with the Department of Electrical Engineering. Di, Kostic, and Zussman are also with the Data Science Institute. All are with Columbia University.

flows via a two-way communication. Specifically, the physical sends the data of its own state to the digital, and the digital feeds back the actuation signals to the physical. The actuation would trigger a change in the stage of the physical, and the updated state is sensed and sent back to the digital again. This iterative process runs between the physical and the digital as time unfolds. The sequential states of the physical should move towards a more desired state than without a DT. In other words, the ultimate goal of a DT development is to add values to the physical for improved safety and efficiency. Below, we first formalize transportation DT, and then discuss the relation between DT and cyberphyscial systems.

Definition 2.1: **Transportation digital twin (T-DT)** is a digital system integrating the pipeline from object detection and tracking, resource allocation, edge-cloud computing and communication, for online simulation, operation, control and management. It is updated online using continuously fed data collected from the physical and send control policies or issue warnings back to the physical, leveraging big data and AI tools. T-DTs are *closed-loop* with *two-way communication*, where data, information, and control signals are exchanged with the physical sequentially and reiteratively.

Definition 2.2: **Cyberphyscial systems (CPS)** [12] are smart systems that include engineered interacting networks of physical and computational components. CPS holds great potential to enable real-time applications thanks to emerging technologies in sensing, communication, and computing.

A transportation CPS interlinks physical and cyber layers, where the cyber layer consists of sensing, networking, computing, and traffic management application modules (see Fig. 1). The DT encloses the cyber layer, and relies on all the modules for two-way interaction with the physical. To enable the technological development of a DT, a physical testbed is needed as a platform for sensing, computing, experimentation, evaluation, as well as design constraints determination. In Sec. III, we will review the CPS technological enablers needed for the development of a DT, and examine the testbed used for our proposed DT in Sec. IV.

We would like to stress that, this paper aims to discuss how AI algorithms and the architecture of CPS contribute to the urban T-DT (UT-DT) pipeline. In contrast, there are highly cited survey papers on T-DT, which are more focused on general transportation scenarios and applications, while AI might not be the key focus. Tab. I outlines the comparison of a partial set of related work.

TABLE I: Comparison of survey papers on DT

Ref.	Topic	Focus
[4][5][6]	Comprehensive review for vehicle mobility apps	Vehicular technology driven pipeline
[13]	Comprehensive review for traffic safety and mobility	CPS pipeline highlighting communication
[14]	Comprehensive review for operation and maintenance apps	Broad travel modes
Ours	Semi-technical review and position for urban traffic apps	AI-Powered CPS pipeline

We first review various CPS and AI methods needed in the pipeline, and then present our T-DT instance. Accordingly, this paper is semi-survey, semi-technical. We employ such an organization, because we find that the publications on T-DT are normally segmented by different communities and journals, which could prevent researchers from understanding the entire pipeline, from upstream sensing and perception, to downstream transportation applications. For example, [15] on networking DT is primarily focused on the development of communication technology with evaluation on latency, even its use case is transportation. Transportation researchers, who hope to implement this system in real-world, have to seek more details about how traffic dynamics would impact the performance, which is unfortunately missing. This is likely because the authors belong to the society of communication and networking. Another example is that, [3] focused on vehicular DT, heavily rely on the foundational knowledge in CPS, which could be somewhat unfamiliar to many transportation researchers. Realizing such a gap in a large body of literature on T-DT, this paper aims to unify the knowledge by presenting a comprehensive summary upfront, and exemplify the pipeline following the summary. After all, the development of a T-DT calls for the interdisciplinary collaboration across transportation, electrical engineering, mechanical engineering, computer science, and human-machine interaction. To foster the readability of the paper, in the next section, we first offer a comprehensive state-of-the-art review on related work along the CPS pipeline.

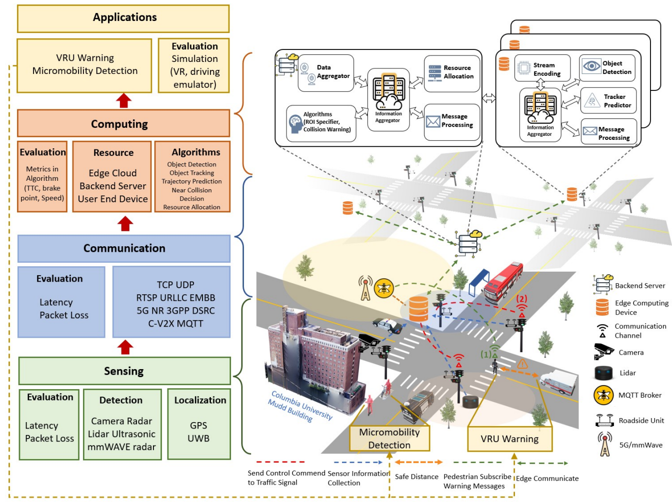


Fig. 1: Schematic diagram of a DT.

III. Related work on CPS-enabled DT

The development of CPS-enabled DT must engage with expertise in sensing, communication, computing, and human-centric perspectives, where AI methods are backbones.

A. Sensing and perception

Sensors are the “eyes” (and “ears”) of a DT. Tab. II summarizes the pros and cons of each sensing technology, namely, mobile devices, on-board vehicles, and roadside infrastructure, for urban traffic applications. The former two are mobile sensors with wider spatial coverage but challenge in precision because of moving references, while the latter at fixed locations could face limited sensing ranges and coverage.

TABLE II: Sensors for urban traffic applications (partially adapted from [16], [17])

Sensor		Purpose	Advantages	Disadvantages
Mobile	GPS	Pedestrian localization	1. Offers global coverage and compatibility with a wide range of devices. 2. Offers reliability and easy access, as it is widely adopted in consumer devices.	1. Accuracy within a few meters, which may be insufficient for safety-critical applications. 2. Poor signals through obstacles. 3. A low update rate for real-time safety applications.
	UWB	Pedestrian localization	1. Achieves high accuracy, often within a few centimeters, making it suitable for precise indoor and short-range outdoor applications. 2. Operates with low latency, providing real-time updates of position data.	1. Requires anchors to be installed at each corner of the designated area before localization. 2. Operates in a frequency range that may overlap with other wireless technologies, such as Wi-Fi, Bluetooth, or cellular networks [18].
On-board Vehicle	Radar	Obstacle Detection	1. Outperforms other sensor types at far distances. 2. Detects vehicle speed and position accurately without the need for calibration. 3. Protects privacy, as this sensor type does not record identifiable images of road users.	1. Performs best only when objects move toward or away from the sensor. 2. A Limited number of classes that can be identified due to the lack of color and resolution. 3. Limited field of view when the range is far [19].
	Camera	Obstacle and lane detection	1. Maintains good resolution when the field of view is wide. 2. Has a long horizon.	1. Difficulties in measuring speed and distance. 2. Performs poorly in bad weather conditions.
	Lidar	Obstacle detection, 3D mapping	1. Measures distance accurately. 2. Constructs 3D models robustly. 3. Shows promising performance in poor weather.	1. Detects nearby objects poorly [19]. 2. Demands high data processing requirements. 3. A shorter effective range than radar.
	Accelero-meter	Acceleration, driving behavior	1. Detects braking, turning, and accelerating accurately. 2. Integrated with vehicles for behavior analysis [20].	1. Performs poorly for slow or subtle movements. 2. Detects poorly in the presence of noise.
Roadside Infrastructure	Camera	Object detection at intersections	1. Provides more details, compared to radar and LiDAR, that can be used to differentiate types of vulnerable road users. 2. Covers a larger area where pedestrians are not confined to a narrow path, such as when people crossing midblock [21].	1. Performs poorly in adverse weather conditions. 2. Difficulty in long-term use due to the cost, power supply, and quantity. 3. No privacy protection.
	Acoustic Ultrasonic	Lane occupancy/vehicle speed	1. Collects data on multiple lanes. 2. Operates during both day and night.	1. Undercounts or overestimates speed. 2. Performs poorly in severe weather.
	MMWave Radar	Vehicle localization, speed measurement	1. Features a compact size and is easy to install. 2. Offers low latency, within 30ms [22]. 3. Penetrates through non-metallic objects.	1. Provides low angular resolution. 2. Measures elevation poorly. 3. Has difficulty with real-time calibration [23], [24].

B. Object detection and tracking

Object detection identifies and classifies objects within the environment using sensors like cameras, Lidar, and Radar. **Object tracking** is the process of monitoring the detected objects over time to determine their position and movement. **Multi-object tracking** is concerned with maintaining the identity of the objects and generating their trajectories. **Trajectory prediction** involves forecasting the future paths of detected and tracked objects. These tasks highly rely on training datasets for urban traffic scenes. Note that there is a much larger size of public datasets collected from on-board vehicles [25], but fewer from other sensor types. Here we thus summarize commonly used and emerging datasets from non-vehicle sensors in Tab. III.

Object detection has been studied extensively for urban applications. A large number of studies focus on low-altitude vehicle and pedestrian detection [51], [52]. Many focus on high-altitude aerial environments, where small object detection becomes an important challenge [31], [35], [33], [36]. Single-stage object detectors, following the original single shot multibox detector [53] and You Look Only Once (YOLO) [54] architectures, have become popular due to their real-time deployment capabilities. In the last few years, transformer-based object detection approaches, competitive with YOLO models, have emerged as the state-of-the-art in object detection when designed to be deployed in the real-time setting [55], [56], [57], [58]. Recent progress in YOLO object detection performance has been enabled through multiple small tricks in architecture and training that all together provide significant improvements in empirical performance [59].

When multiple camera views are available, 3D object detection has been studied heavily for autonomous driving [51], [60], as well as for infrastructure-based 3D object detection [26], [27], [28], [29]. Many approaches to 3D object detection use object queries [61], bird's-eye view transformations [62], or a combination of the two [63].

To build models that make weaker assumptions regarding the sensors, some models have considered the harder task of monocular 3D object detection. MonoCon uses extra regression head branches for learning auxiliary contexts, that are then discarded during inference [64]. DEVIANT is a model architecture equivariant to depth translations [65]. MonoLSS introduces a learnable sample selection module to improve the stability and reliability of the model at test time [66]. Different models have been proposed for infrastructure-based 3D object detection, as many models developed for vehicle-side perception make strong assumptions regarding the position of the cameras. BEVHeight predicts height to the ground to support 3D object detection [67]. CoBEV combines depth and height features to further improve the performance of infrastructure-based 3D object detection [68]. MonoUNI presents the idea of normalized depth, which makes depth prediction independent of camera pitch angle and focal length [69].

To improve the limitations of camera-only perception methods, different sensor combinations are explored. For example, LiDARs or radars, combined with cameras, can detect objects in scenarios where using only the camera is insufficient, such as extreme lighting and weather conditions, or anoma-

TABLE III: Public datasets for urban traffic scenes (On-board vehicle datasets can be found in [25].)

Sensor location	Dataset	Purpose	Sensor Setup	Collection region
Infrastructure	VIRAT [26], Constellation [27]	Urban object detection and tracking, visual event recognition	RGB cameras	Public outdoor spaces in China, city intersection in New York
	Rope3D [28]	2D/3D Road-side object detection, multi-view	RGB cameras, LiDAR	Streets in Beijing
	DeepSense 6G [29]	2D/3D object detection, sensor fusion	RGB cameras, mmWave Phase Arrays, LiDAR, Radar	Various indoor and outdoor spaces
	WILDTRACK [30]	Multi-Object Tracking	RGB cameras	University campus in Zurich
Aerial	VisDrone [31], NGSim [32], highD [33], rounD [34], DOTA [35], CitySim [36]	Aerial object detection and tracking, trajectory forecasting	Drone-based RGB cameras	Urban spaces in China and Aachen, intersections in California and Florida
	MOT Challenge [37], MOTS20 [38], UAVDT [39]	Aerial object detection and tracking, trajectory forecasting	Drone-based RGB cameras	Various Indoor and Outdoor Scenes
Misc	CoCo [40], ADE20K [41], Cityscapes [41]	Object detection, semantic segmentation	RGB cameras	Various indoor and outdoor spaces
Synthetic	CARLA [42], [43], [44]	Autonomous-driving object detection and segmentation	RGB cameras, LiDAR	Urban European/North American environments
	GTAV [45], Synscapes [46], UrbanSyn [47]	Object detection and segmentation	RGB cameras, LiDAR	Urban European/North American environments
	MOTSynth [48]	Multi-object tracking	RGB cameras	Urban European/North American environments
	MatrixCity [49]	Neural-rendering benchmarking (vehicle/pedestrian-free)	RGB cameras	Synthetic city environment
	Boundless [50]	Object detection and segmentation with UE5-synthesised data	RGB cameras	Synthetic urban environments

lous situations where the camera data is significantly out-of-distribution. Sensor fusion for self-driving cars is now being studied including sensor data for these modalities [70], [71]. These methods often involve the projection of camera, radar and LiDAR features independently to a bird's-eye view feature space, wherein an aggregation function could be used to merge the features extracted from these different sensors.

Multi-object tracking involves matching newly detected objects with the existing ones by their inter-frame positional and visual similarity information [72], [73]. ByteTrack improves the traditional Hungarian-algorithm-based matching paradigm to gather more comprehensive information [74]. BoT-SORT further incorporates advanced object re-identification modules and a refined Kalman filter for more accurate performances [75]. BoostTrack explores novel distance and shape similarity measurements to deal with ambiguity caused by unreliable detection results [76], [77].

Predicting future trajectories of detected objects is often a crucial part of safety-critical applications. Numerous deep neural network models have emerged as competitive candidates for trajectory prediction over the past few years. Majority of modern architectures for predicting future trajectories of detected objects adopt Recurrent Neural Networks which is responsible for predicting future object positions based on their historical coordinates, together with generative components which handles the variation and flexibility in social interactions [78], [79]. Neural Social Physics model [80] incorporates learnable parameters into explicit physics models built on top of neural networks. SemanticFormer [81] seeks more structural and humanized environment understanding by constructing a semantic knowledge graph. TrajNet++ [82], TDOR [83], and CASPNet++ [84] predict the distributions of future trajectories based on occupancy grid maps. Models like FRM [85] and PPT [86] decompose the prediction task by taking a multi-stage approach. Larger amounts of data comprised of multiple modalities and more comprehensive frameworks have shown increasing importance as is demonstrated by UniTraj++ [87]. Unlike the tracking algorithms, specific training or fine-tuning is often required before the deployment of trajectory forecasting models to unseen scenarios.

C. Real-time video analytics

Developing end-to-end real-time video analytics systems on a large scale for time-sensitive and safety-critical traffic and crowd management applications presents challenges. Video analytics requires the collection and processing of large volumes of video data, which can be resource-intensive and costly. Optimizing computation and network resource usage while maintaining or enhancing the accuracy of analytical results can be challenging. This challenge is further complicated by the need to adapt to varying network conditions, computational resources, and dynamic scene changes in real-time. Tab. IV provides a comparative analysis of various approaches to address these challenges. The approaches differ in their focus–some prioritize reducing latency and resource consumption, while others emphasize maintaining or enhancing accuracy, especially under constrained conditions. This comparison provided in Tab. IV highlights the trade-offs inherent in real-time video analytics and emphasizes different optimization methods to balance throughput, accuracy, energy consumption, and computational efficiency across diverse deployment scenarios, including edge devices, cloud platforms, and hybrid environments.

D. Communication and networking

A DT for safety-critical applications requires real time communications with aggressively low latency. We explore issues related to low latency communications, and survey

TABLE IV: Comparison of various approaches and optimization objectives in video analytics.

References	Approach	Optimization objective
SPINN [88], Adaptive offloading [89], Shoggoth [90], Sniper [91], JAVP [92], Auction-base [93]	Distributed DNN inference over end devices, edge, and cloud.	Optimize throughput, accuracy, and energy consumption under varying network conditions.
CEVAS [94], SAHI [95], CrossRoI [96], Elf [97]	Adaptive RoI assignment and frame sampling.	Reduce the bandwidth consumption and enhance accuracy.
AdaMask [98], Respire [99], CrossVision [100], VaBUS [101]	Leveraging redundant regions on frames and background understanding.	Minimize network and computation overhead while ensuring high accuracy.
Elf [97], Mobile edge analytics [102], Sniper [91], JAVP [92], Auction-base [93]	Video analytics query scheduling and resource allocation over multiple edge devices.	Reduce latency and increase computation resource utilization.
CEVAS [94], Edge-assisted serverless [103]	Adaptive model selection.	Enhance performance with limited computation resources.
Shoggoth [90], Edge-assisted [104]	Online model fine-tuning and model switching.	Improve the accuracy and efficiency of real-time video inference on edge devices in changing video scenes.
DAO [105], VaBUS [101], AccMPEG [106], AdaMask [98], ILCAS [107]	Adaptive video encoding and compression parameters.	Balance low latency, high accuracy, and low compute overhead on edge devices.
AdaDSR [108], AccDecoder [109]	Camera-side downsampling and server-side super-resolution upsampling.	Balance the trade-offs among accuracy, network cost, and computational cost.
MadEye [110], WiseCam [111]	Dynamical orientation adaption of pan-tilt-zoom (PTZ) cameras.	Boost the overall accuracy while maintaining the resource cost.
EAIS [112], EALI [113], SERAS [114]	Use of an energy-aware scheduler that effectively coordinates batching and dynamic voltage frequency scaling (DVFS) settings.	Minimize energy consumption for CNN inference services on high-performance GPUs while meeting latency of Service-Level Objectives.

component technologies and protocols that can be utilized to achieve very low latency.

1) Real-time requirements and low latency targets

Sensor and control data in a real time system is subject to latency created by the stages, namely, (1) data acquisition from traffic participants (such as camera recordings and encoding, harvesting data from autonomous vehicles, and collect information from fiber); (2) transmission of data across communications links from sensors to inferencing servers using communications protocols such as Transmission Control Protocol (TCP), Unreliable Data Protocol (UDP) and Real-Time Streaming Protocol (RTSP); (3) data preprocessing (video decoding, and cropping); (4) AI inferencing; (5) higher level reasoning about required feedback to traffic participants; and (6) sending feedback to traffic participants across communications links via low-latency broadcast or dedicated channels over wired and wireless.

Smart city applications can be grouped according to their latency requirements. Many, if not all, pedestrian-associated application (facilitated by pedestrian detection/observations and message notifications) are likely to expect the round trip delay in the range of a couple of seconds. Such latency can be supported by contemporary cameras, communication protocols and inferencing engines. Applications which would attempt to close the observation/notification loop for vehicles moving at about 10 km/h may expect latency in tens of millisecond. Using conventional video compression, RTP/RTSP streaming and edge computing is inadequate to support such latency. This presents the opportunity to pursue novel engineering solutions and research problems.

2) Communication techniques and protocols

Ultra-Reliable Low Latency Communications (URLLC), a key component of 5G wireless, can help achieve the low latency targets. Along with enhanced mobile broadband (eMBB) and massive machine-type communication (mMTC), URLLC [115] represents one of the three main capabilities of 5G New Radio (5G NR), as standardized by the 3rd Generation Partnership Project (3GPP). In the context of transportation systems, URLLC aims to deliver up to 99.999% reliability and single-digit millisecond latency [116]. However, meeting these performance metrics is challenging in practice due to complex channel environments, particularly in dense urban areas, which can reduce reliability. For intelligent transportation systems, where URLLC may be used as infrastructure backhaul, the target is an end-to-end latency of 30 ms [117].

An emerging technology in this space is Cellular-Vehicle-to-Everything (C-V2X), which has largely replaced the earlier Wi-Fi-based Dedicated Short-Range Communications (DSRC). Unlike DSRC, C-V2X leverages cellular networks, allowing network providers to offer always-on connectivity, which is a critical feature for time-sensitive applications. Additionally, private 5G networks are being developed to ensure this level of connectivity, overcoming the congestion and range limitations inherent to Wi-Fi [118]. In transportation systems, active collaboration between wireless service providers and vehicle manufacturers is in progress to integrate private 5G networks into vehicular networks [119].

Tab. V summarizes the key characteristics of commonly used IoT communication protocols, Message Queuing Telemetry Transport (MQTT) [120], Constrained Application Protocol (CoAP) [121], and Hypertext Transfer Protocol (HTTP) [122]. Among them, MQTT emerges as a practical choice, due to its combination of low latency, high scalability, and reliable delivery mechanisms. Its lightweight publish/subscribe model, Quality of Service (QoS) guarantees, and session management make it well suited for real-time data exchange in dynamic, safety-critical environments like urban transportation systems [123], [124].

IV. Proposed DT pipeline

In this section, we present a DT architecture for UT-DT, based on the sensing/communication/computing testbed deployed in NYC (Fig. 2). The proposed architecture is enabled

TABLE V: Communication protocol comparison for real-time digital twin systems.

Protocol	Latency	Scalability	Reliability	Best Suited For
MQTT	Persistent TCP connection delivers low latency for ongoing messaging	Highly scalable with a broker-based publish/subscribe model supporting many-to-many communication across thousands of devices	Guarantees reliable delivery with configurable acknowledgment levels, exactly-once delivery via a four-step handshake, persistent sessions	Large-scale IoT, real-time sensor/actuator networks
CoAP	UDP-based, very low latency when network is stable, but performance degrades with packet loss	Limited scalability due to client-server request/response model. Multicast is possible but complex and unreliable at scale	Basic two-way acknowledgment, no exactly-once delivery guarantee, no session continuity	Resource-limited devices that send small, infrequent data
HTTP	Higher latency due to new TCP/TLS handshake and headers per request	Limited scalability due to client-server request/response model and verbose ASCII headers	Relies on TCP for delivery, with no application-level acknowledgment, retries, or session handling	Web APIs, periodic data transfer, backend integration

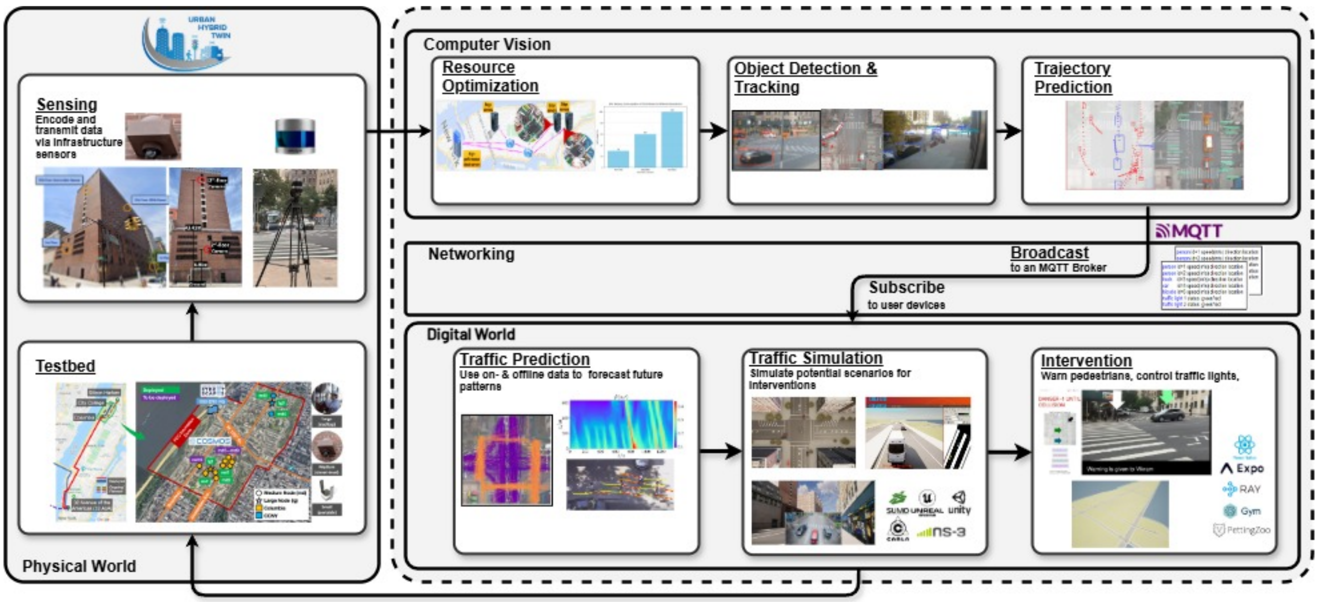


Fig. 2: Architecture of the proposed DT pipeline: Urban Transportation DT (UT-DT).

by cameras and LiDARs, high speed communications, and edge cloud computing.

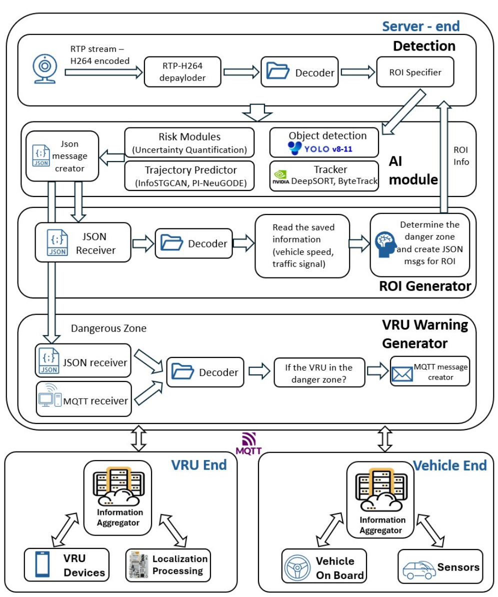
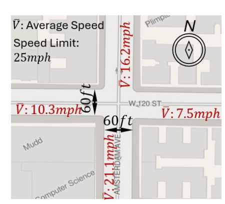


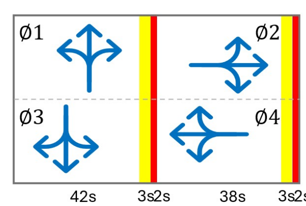
Fig. 3: Diagram of intersection safety warning use case.

A. Physical Infrastructure

Pilot experiments are executed at the signalized intersection of Amsterdam Avenue (major) and 120th Street (minor) near Columbia campus in NYC. The road geometry and traffic statistics are summarized in Fig. 4.



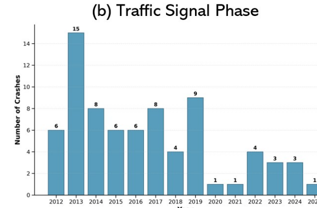
(a) Intersection layout



(b) Traffic Signal Phase

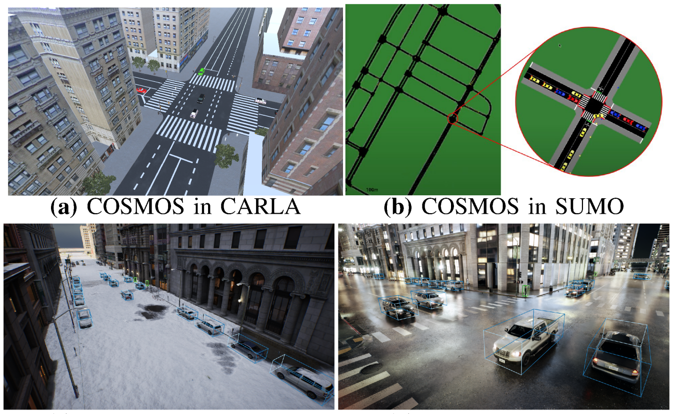
	N	S	W	E
N	/	2985	850	461
S	5356	/	758	224
W	700	512	/	528
E	94	47	307	/

(c) Traffic Volume (Daily)



(d) Number of Collision Events

Fig. 4: Traffic statistics for NYC's intersection of 120th St. & Amsterdam Ave.



(a) COSMOS in CARLA **(b)** COSMOS in SUMO

(c) Boundless with snow **(d)** Boundless with rain

Fig. 5: (Top) Carla-SUMO simulator; (Bottom) Qualitative examples from Boundless. Bounding boxes for vehicles are shown in blue, and pedestrians are shown in green.

1) Technological enabler

The physical functionality of the proposed UT-DT is built on top of the COSMOS testbed ("Cloud enhanced Open Software defined mobile wireless testbed for city-Scale deployment") [125], developed for real-world research, development, and deployment of city-scale advanced wireless communications and innovative applications [125]. It targets the technology "sweet spot" of ultra-high bandwidth and ultra-low latency, a capability that will enable a broad class of safety-time critical applications. Deployed in West Harlem, NYC, next to Columbia University campus, the COSMOS testbed is an enabler for research on the design, development, and deployment of a DT for urban traffic management.

At the intersection of Amsterdam Avenue and 120th Street, cameras, LiDAR, and wireless sensing and communication nodes are deployed. The cameras employ H264 encoding with an I-frame interval set to 10 frames. The live RTSP video streams–with 4K resolution at 30 fps–were processed on a COSMOS edge server equipped with an A100 GPU. The LiDAR is a Velodyne VLP-32C with up to 200m range.

B. Digital twin

Building on data collected from the physical world, we built an Unreal Engine based simulation platform for generating photorealistic street scenes. Moving objects are populated into an integrated SUMO-CARLA simulation platform to validate our safety warning application [126], where vehicle locations are synchronized from the real-world MQTT messages.

Another major use case for DT is data generation, especially given the privacy concerns associated with use and release of data for urban scenarios. We use high quality graphical assets, raytracing and render the scene at 8K resolution (7680×4320) (see Fig. 5) to generate realistic-looking synthetic data for training object detection models for the COSMOS testbed as well as other urban North American environments [50].

C. Use cases

The value of a DT for urban transportation management ranges from traffic state prediction, spatiotemporal traffic flow forecasting, to urban planning and policy making. Here we focus on two that are particularly important in urban settings, leveraging video analytics.

1) Use case 1: Micromobility detection

In the U.S., the e-bike had a market size of almost $2 billion in 2022, with a projected growth of 15.6% from 2023 to 2030. In 2022, 1.1 million e-bikes were sold, four times as many as were sold in 2019. In 2019, 136 million trips were made using micromobility, a 60% increase from 2018. Because their presence and surge have profound implications for road safety and urban traffic management, automatic detection of micromobility for regulation becomes increasingly important. However, there does not exist a standard AI model nor a dataset for the object detection of this emerging transportation mode.

TABLE VI: Class-wise recall (%) on the Micromobility test set. Higher values indicate better detection performance.

Model	**Bicycle ↑**	**e-Bike ↑**	**Motorcycle ↑**
YOLOv9e	52.9	–	54.7
Ours	**68.1**	**61.0**	**91.8**

AI model evaluation

We benchmarked the performance of off-the-shelf object detection models against a real-world dataset of infrastructure cameras placed in NYC looking at the same intersection with 4K resolution (3840×2160). We found that models trained with standard datasets have limited performance in these real-world scenarios, showing a significant gap to be addressed. To help bridge the gap, we trained YOLO models in native 4K resolution. Our dataset contains 14,000 images collected from the COSMOS testbed, of which a subset of 4,000 images collected at a different timeframe that does not overlap with the training test are held out for testing. We present the results in Table VI. We use recall at the default confidence threshold for both our model and YOLOv9e, a top-performing real-time object detection model pretrained on the COCO dataset [40]. We use recall due to the lack of a corresponding "e-bike" class in existing datasets, whose datasets predate the modern e-bikes. Selected samples are shown in Fig. 6. Our findings reveal a significant performance gap that shows the need to collect datasets for object detection in urban metropolises.

2) Use case 2: VRU safety warning

Safety-critical applications, making automated life-and-death decisions such as collision avoidance warning between automobiles and pedestrians, need to activate at the precise time and the right moment with bounded latency. Safety-critical systems are often time-critical and reactive, because they need to react to external signals in precise time [127], and require tasks to be executed in a timely manner. Theses systems hold the substantial potential to enhance road safety and save lives. They are, however, risky to test and run, because rare events like automobile collisions are challenging and unethical to replicate. Thanks to emerging technologies in ubiquitous sensing, low-latency high-bandwidth communication, high-speed computing and AI, safety critical applications

Fig. 6: Four sample frames showing the small scale of micromobility instances from infrastructure-based 4K resolution cameras in the COSMOS testbed.

could potentially be modeled, simulated, processed, and tested in a DT. The applications of DTs on safety-critical scenarios, however, remains understudied, because it necessities short runtime and real-time reaction, posing high requirements for communication and computing technologies, pipeline architecture design, and testing. We summarize the safety-critical applications to address conflict risks between vehicles and VRUs in Tab. VII, and offer an outlook of safety guarantee related methods in Sec. V-A3.

Intersections, where sixty percent of crashes happen [135], [136], are critical bottlenecks of an urban transportation network. To improve urban road safety and increase traffic capacities, safety warning is the key. Leveraging existing sensors at the intersection, we have developed a combined VRU and vehicle warning application. The cameras and the LiDAR sense the presence of VRUs at urban intersections, make predictions of their movements, apply traffic operation and control strategies, and feedback to system controllers and road users, with the primary goal of increasing traffic safety and efficiency (see Fig. 3).

Resource optimization

Due to the computational needs of complex computer vision models and the high volume of video data, ensuring scalability in a video analytics pipeline requires optimization of its configuration to maintain performance. However, the system's performance is impacted by factors such as the video content, network conditions, and available computational resources. As a result, it is crucial to implement a dynamic, real-time optimization mechanism that adjusts key configuration parameters—such as resolution, frame rate, and bitrate—based on these varying conditions. This adaptive approach allows the system to continuously balance performance with resource efficiency, ensuring scalable and reliable video analytics. Accordingly, we have equipped our video analytics pipeline with a Resource Optimization (see Fig. 2) component that continuously adapts the system's configuration parameters to maintain its performance.

In addition to dynamic adaptation of the system's configuration, we can reduce latency and GPU consumption by identifying Regions of Interest (RoIs) where pedestrians might be in danger, using lightweight processing (e.g., low resolution, small models). As shown in Table VIII, smaller models ("YOLOv8s" in Column 4) incur significantly lower latency. Larger models ("YOLOv8x" in Column 6) with higher resolution and frame rate are only triggered when a critical danger area, i.e., RoI, is detected. For example, detecting large objects such as vehicles does not require large models or high resolution. Therefore, we can use a smaller model and lower resolution to detect vehicles and their trajectories, and based on that, determine the danger areas where pedestrians may be at risk or in the blind zones of vehicles. These identified danger areas are then processed using larger models and higher resolution to detect pedestrians at risk and notify them or the vehicles if necessary. In our intersection safety warning system (shown in Fig. 3), we have embedded an RoI Specifier element to facilitate this approach.

AI models

A variety of AI models are developed to predict future trajectories of interacting road users at the intersection, including InfoSTGCAN (An Information-Maximizing Spatial-Temporal Graph Convolutional Attention Network) [137], which encodes road user trajectories into quantized latent codes to account for heterogeneity in road users, PI-NeuGODE (Physics-Informed Graph Neural Ordinary Differential Equations), using physics-informed graph neural ordinary differential equations [138], [139], and uncertainty quantification (UQ) to characterize the predictive confidence [140]. These probabilistic trajectory prediction methods enhance pedestrian safety at intersections by incorporating UQ into risk assessment. The Kalman Filter (KF) [141] performs prediction by recursively estimating the state of a moving object using a linear dynamic model with Gaussian noise. Trajectron++ [142] models multimodal future trajectories by combining recurrent neural networks with conditional variational autoencoders, enabling probabilistic and socially-aware predictions. We present the predicted trajectories and performance comparison in Fig. 7, where the performance is evaluated using Average Displacement Error (ADE, i.e., the mean Euclidean distance between predicted and ground-truth positions over all future time steps) and Final Displacement Error (FDE, i.e., the Euclidean distance at the final prediction step).

Evaluation

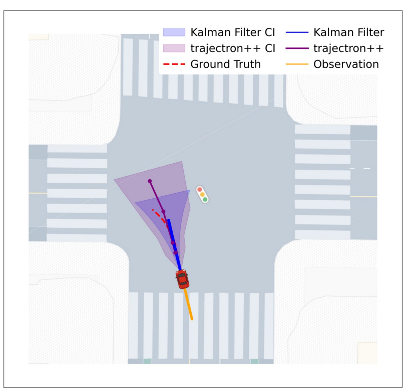
In safety-critical applications, it is infeasible to perform field experiments for such assessment. DT simulation is thus used to assess the effectiveness of the system.

a. End-to-end latency

Measuring latency could be particularly challenging due to the need for cross-network timing synchronization of devices and compute server. Table VIII presents the latency incurred by the main components of a typical video analytics pipeline for object detection and tracking, as shown in Fig. 3. We

TABLE VII: Literature review on safety warning to VRUs.

Ref	Technology	Objective	Risk Assessment Metrics	Evaluation Method/Metric
[128]	**Sensing**: roadside sensors and VRU smartphones **Communication&Networking**: 5G, MNO infrastructure, and ITS-G5 **Computing**: Edge-cloud hybrid computing	Reduces latency and optimizes resource utilization through dynamic service placement.	The distance between the VRU and vehicles	End-to-end delay: 200 ms
[129]	**Sensing**: Real-time camera detection, YOLOv7 **Communication&Networking**: 5G, 802.11p, C-V2X, Vehicular Basic Safety Message (BSM) **Computing**: Local server equipped with GPU	Develops a video-based vehicular BSMs method with lower error and latency that outperforms the cellular vehicle-to-everything (C-V2X) method.	N/A	1. End-to-end delay: < 100 ms 2. Localization/Speed accuracy
[130]	**Sensing**: GPS on smartphones **Communication&Networking**: LTE, Node B, Wi-Fi, Cooperative Awareness Message (CAM) **Computing**: CAM server deployed at edge/cloud	Proposes a system using commercial devices and standard messages for road user communication.	The distance between VRUs and nearby entities	Latency from VRU to CAM < 50 ms
[131]	**Sensing**: Camera, Android phone's GPS module **Communication&Networking**: 3G, 4G, 5G, and MQTT protocol **Computing**: Coral Edge TPU with TensorFlow lite	Develops a traffic safety system using edge computing and 5G to deliver low-latency warnings.	The coordinates of pedestrians and cyclists in a driver's blind spots	Latency: 1. 4G - 109.35 ms 2. 5G - 90.95 ms
[132]	**Sensing**: N/A **Communication&Networking**: Wi-Fi, C-V2X, 802.11p, ITS-G5, Cooperative Awareness Message **Computing**: Edge computing server, smartphones	Develops a system to deliver CAM to VRUs on smartphones using Beacon stuffing without the need for root access to utilize 802.11p.	N/A	1. End-to-end latency ~ 2500 ms 2. MAC channel utilization
[133]	**Sensing**: Cameras, Radar, YOLOv3 on NVIDIA JETSON, On Board Units (OBUs) **Communication&Networking**: 5G, Fiber, LTE, ITS-G5, and MQTT **Computing**: Road side units	Develops a system with sensing and communication, along with fusion and collision detection algorithms, to predict potential collisions and warn VRUs.	The distance between VRUs and nearby entities	1. End-to-end latency < 300 ms 2. Distance error of vehicles and VRUs
[126]	**Sensing**: Camera, real time object detection **Communication**: LTE, Wi-Fi, and MQTT protocol **Computing**: GPU on server end	Develops a real-time system with a mobile application to warn pedestrians to avoid vehicle and walker collisions.	Time to collision (TTC)	1. End-to-end latency 400 ms. 2. Simulation.
[134]	**Sensing**: CAN, GNSS, roadside, cameras, Google MediaPipe Posekeypoint **Communication&Networking**: Fiber, and Simulated V2X **Computing**: Server end GPU and portable GPU	Develops a DT framework for connected vehicles and pedestrian in-the-loop simulation. Test it with a V2P collision warning use case.	TTC	1. Speed 2. Brake point 3. Distance to conflict point
Ours	**Sensing**: Camera, YOLOv8 object detection. **Communication&Networking**: Fiber, LTE, and MQTT protocol **Computing**: Server end GPU	Develops a DT pipeline to demonstrate use cases (including intersection safety warning and ATSC) in urban settings.	TTC	1. Accuracy for the warning issued 2. Granular latency per frame



Method	ADE ↓	FDE ↓
Kalman Filter	0.91	1.92
Trajectron++	**0.77**	**1.34**

ADE and FDE comparison. Lower values are better (↓).

Fig. 7: Performance comparison for trajectory prediction

measured the latency for three sizes of YOLOv8 object detection model: YOLOv8s (small, ∼11.2 million parameters), YOLOv8m (medium, ∼25.9 million parameters), and YOLOv8x (large, ∼68.2 million parameters). Larger models offer improved detection accuracy but come at the cost of increased latency and higher GPU and memory usage [143]. The pipeline has been systematically optimized in terms of memory and resource usage. These elements usually run on an edge device or server. The results identify potential bottlenecks within the pipeline and indicate which components could benefit from further optimization.

b. Safety message generation accuracy

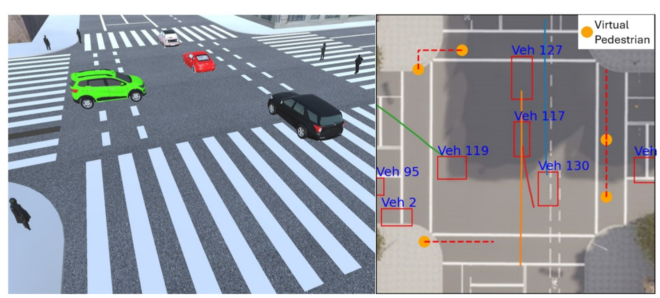


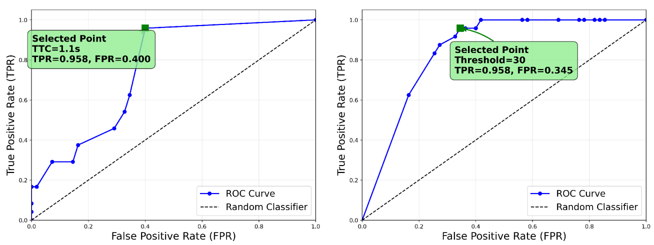
Fig. 8: A simulated pedestrian in CARLA (left) surrounded by cars with real-world positions (right).

To validate the accuracy of our trajectory prediction and risk assessment algorithms for warning generation, we con-

TABLE VIII: Average latency and Standard Deviation for different pipeline elements.

Metric	Inference Elements						Downstream Elements				
	Reception	Pre-processing	Object Detection Model			Object Tracking	MQTT Msg Creator	MQTT Msg Retrieval			
			Small (YOLOv8s)	Medium (YOLOv8m)	Large (YOLOv8x)			Ethernet	Wi-Fi	LTE	5G
Avg Latency (Std Dev) [ms]	1.94 (1.69)	0.108 (0.024)	4.034 (0.084)	7.216 (0.086)	11.140 (1.800)	0.973 (0.173)	0.081 (0.021)	3.21 (0.315)	6.86 (1.19)	45.72 (15.30)	39.21 (7.12)

ducted three rounds of simulation, each lasting 10 minutes and generating a total of 232 virtual pedestrians in CARLA. We then compared the number of collision warning messages with the actual number of simulated collisions. The process is illustrated in Fig. 8.



(a) TTC (in seconds) (b) Distance (in pixels)

Fig. 9: ROC curves for trajectory prediction to determine optimal thresholds

Note that metrics for risk assessment include time to collision (TTC) (i.e., the time remaining before a collision occurs) and post-encroachment time (PET) (i.e., the time interval between when the encroaching vehicle leaves the conflict point and when the vehicle of the right-of-way arrives at the conflict point). TTC is commonly used, while PET is typically for post-event analysis. To select the optimal thresholds of TTC, we first generate the ROC (Receiver Operating Characteristic) curve in Fig. 9. Fig. 9a illustrates the relationship between the true positive rate (TPR) and the false positive rate (FPR) across various TTC threshold values, ranging from 0.1 to 1.2 seconds. The optimal TTC threshold is 1.1 seconds (indicated by a blue square), at which the TPR reaches 0.958 and the FPR is 0.4. To compute TTC, we need to compare each pair of predicted trajectory points in the next t time steps, determined by a danger distance threshold that is defined as when the proximity between a vehicle and a pedestrian constitutes a dangerous interaction. Fig. 9b presents the TPR and FPR for the threshold of danger distances ranging from 5 to 100 pixels. The optimal threshold is valued at 30 pixels, which yields a TPR of 0.958 and an FPR of 0.345. Based on the above two thresholds, we run the collision prediction model. The resultant confusion matrix for predicted collisions under the selected thresholds is given by $[[\mathrm{TP}, \mathrm{FP}], [\mathrm{FN}, \mathrm{TN}]] = [[66, 45], [2, 119]]$.

c. Human response time assessment

Would issuing warnings to pedestrians help reduce users' response time and increase their safety awareness? To test this hypothesis, we designed virtual reality (VR) experiments in Unity3D [144], where warnings are provided via voice and text displayed on the VR headset Meta Quest 3. There are two traffic scenarios, one involving an interaction between the participant (i.e., a pedestrian) and an oncoming scooter, and the other between the participant and an oncoming vehicle.

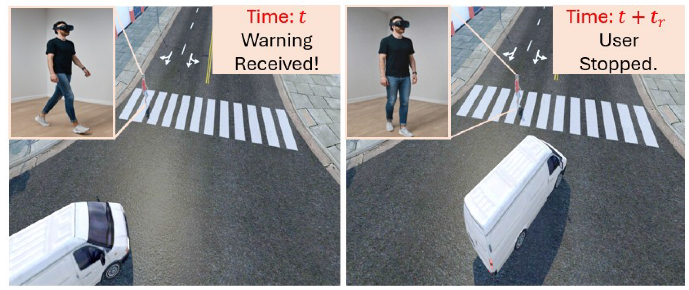


Fig. 10: Response time determination in the VR experiment

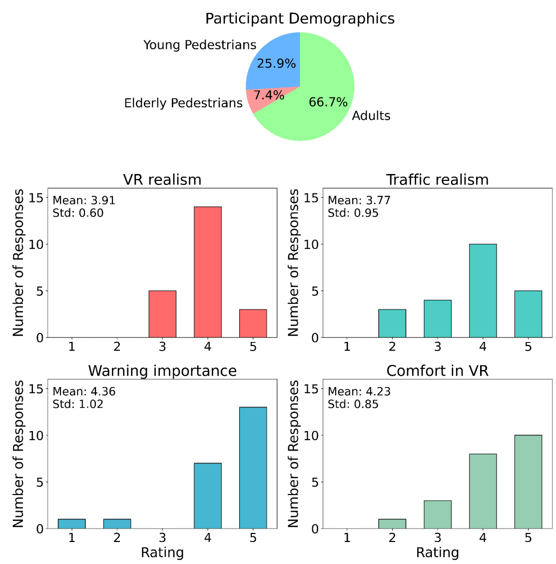


Fig. 11: Survey results from the VR experiment

In each scenario, participants first receive no warning and then a warning, and their response times are recorded. As illustrated in Fig. 10, when traffic approaches at time t, the user either receives a warning or does not. The user stops after a response time t_r, which is then recorded. The response time distributions are presented in Fig. 12, with the 'no warning' condition shown in yellow and the 'warning' condition in blue. The issuance of warning has reduced people's average response time by 0.62s (in Scenario 1) and 1.11s (in Scenario 2), respectively. The demographics of the participants are shown in the upper section of Fig. 11. The bar chart summarizes the participants' survey responses regarding VR and traffic realism, the perceived importance of safety warnings, and comfort within the VR environment, with mean and standard deviation indicated on the top left corners of each subfigure.

3) *Discussions*

We would like to highlight that our NYC's testbed deployment demonstrates generalizability. As the biggest metropolis of the U.S., NYC's dense population, limited space, and

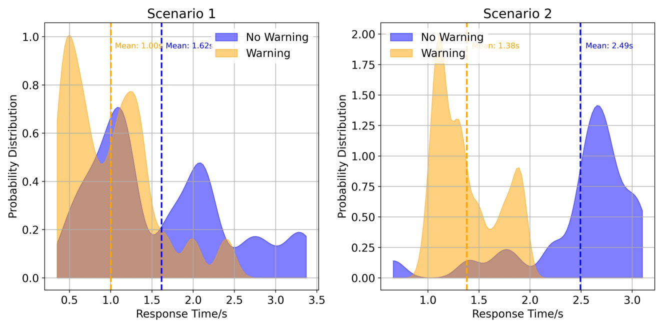


Fig. 12: Comparison of response times with and without a warning. A scooter (in Scenario 1) and a vehicle (in Scenario 2) poses the danger, respectively.

multimodal transportation systems (consisting of public transit with buses and subway, driving, ferry) motivate us to study the interactions between automobiles and VRUs (including pedestrians, cyclists, and surging micromoblity and e-mobility). Located in one of the world's busiest regions - Manhattan of the NYC, the COSMOS testbed offers a natural laboratory for us to collect abundant data and conduct live experiments, which would otherwise not be possible.

We leverage the COSMOS testbed to design and implement sensor (e.g., camera, LiDAR) data analytics systems and evaluate them in urban settings to ensure they can scale to thousands of connected intersections. To achieve this, we ensure that the designed analytics systems are efficient in resource consumption, particularly network bandwidth and GPU usage. We also ensure that these systems are modular and deployable in a distributed manner across multiple edge servers and cloud infrastructure. Fig. 3 demonstrates such a modular and distributable system. This design enables flexibility, allowing the system to be deployed in settings where a single edge server cannot handle the full workload, requiring distribution across several edge servers. Therefore, the system must be designed such that each module can be easily deployed on a different server and can seamlessly communicate with other modules, regardless of whether they reside on the same server. Additionally, the AI models used must be replaceable or generalizable to accommodate different environments, camera angles, heights, and other variations. With these principles in mind, we use the COSMOS testbed to develop scalable systems.

The NYC's dense high-rise buildings enable us to leverage legacy infrastructure for traffic sensing and monitoring, and facilitates the instrumentation of cameras and communication nodes. Note that there is a growing trend of studies relying on emerging communication protocols like C-V2X for vehicle DTs [145], [146], which requires customized infrastructure and user-end devices. For rapid and scalable deployment, our pilot experiments utilize contemporary protocols, which hold the potential to further leverage the city's legacy traffic cameras and wireless communication infrastructure for scalability and cost-effectiveness. Our technologies could be replaced by emerging methods that can achieve lower latency and higher bandwidths. For instance, software defined features of the COSMOS testbed and the flexibility of radios make it possible to deploy emerging technologies such as C-V2X and experiment with novel low-latency applications. Since major intersections in the neighborhood has similar sensor suites, it would not need additional infrastructure deployment to generalize our system to network wide multi-intersection settings.

We will extend our use cases to multi-pedestrian warning scenarios. Smartphone-to-cloud beaconing with an Adaptive Multi-Mode (AMM) scheme enables real-time location sharing from multiple devices, allowing the server to issue timely risk alerts [147]. A Vehicle-to-Pedestrian (V2P) system [148] leveraging Bluetooth Low Energy (BLE) advertising on smartphones broadcasts standardized Personal Safety Messages (PSM) to address the multi-device communication challenge in road safety. Moreover, efficient multi-device communication is essential for real-time data sharing and coordination among mobile users in smart urban environments. Key challenges include ensuring reliable connectivity in dense settings, minimizing communication latency, and maintaining user privacy. Advanced techniques such as device-to-device (D2D) communication [149], opportunistic networking [150], and privacy-preserving data aggregation [151] have been developed to address these issues.

V. Conclusions and Open Questions

In this paper, we first review the AI methods applied to every stage of the DT pipeline, from object detection, tracking, prediction, simulation, to traffic operation and management. Then, leveraging the unique characteristics of NYC, we propose a DT architecture and present a safety-critical use case, namely, intersection safety warning. Three evaluation methods are performed, including measuring latency, simulating collision reduction, and VR experiments for human responses.

A growing number of sensors, explosive amounts of data, and increasing computational powers have opened up tremendous opportunities for researchers to apply AI to create, train, evaluate, and streamline DT pipelines for uban traffic management. Subsequently, we will present emerging trends, challenges, and open questions for the development of DT.

A. Emerging trends

While literature on DTs for individual components has been surging, how each element of the DT works collectively and function organically is key to the development of the next-generation DT, which could empower the intelligence and automation of transportation applications.

1) Engineering the pipeline

Engineering a DT via the integration of multiple subsystems poses technical challenges, and we will name a few.

I. Sensor fusion with time synchronization

Various methods have been proposed to tackle the time-synchronization problem in multi-camera settings for a single intersection or area, either utilizing visual cues or through explicit clock synchronization between the cameras. [152] estimates the spatial transformation between the views. [153]

proposes an approach at time synchronization based on the temporal alignment of matching trajectories of entities present in the overlapping scenes. [154] proposes solving a global alignment problem based on video feature descriptors. [155] uses image features to train neural networks for solving the alignment problem. [156] proposes a neural network that uses pose cues to align videos temporally. [157] proposes the use of abrupt lighting changes as temporal cues for facilitating alignment for rolling shutter cameras. Other approaches include estimations of camera capture and transmission latency [158], or clock synchronization [159], [160], [161]. These approaches address the problem of synchronizing sensors across different road intersections. Efficient, low-latency implementation of these methods are vital for synchronization and fusion of camera predictions.

II. Designing edge-cloud architecture and networking

To facilitate the development of a DT in an urban setting with numerous intersections capable of communicating with vehicles and VRUs, an extended network of cloud-connected edge devices and sensors, such as cameras, is required. These sensors generate substantial amounts of real-time data that must be transmitted to edge or cloud systems and processed with bounded latency. The data from all sensors should be integrated into a centralized platform. Given the extensive distribution of these devices, privacy and data security become critical considerations. To safeguard privacy, encoded sensor data is transmitted only to edge devices, where it is processed to extract relevant metadata. Only this metadata is then forwarded to the cloud or central platform for further analysis and integration, ensuring that no raw data or personal information is shared. To this end, data and device federation using federated (reinforcement) learning has gained growing traction [162], [163], [164].

III. Integrated sensing and communication

Integrated sensing and communication (ISAC) is an emerging direction in the design of Beyond-5G wireless networks, enabling transmitted communication waveforms to be opportunistically used as radar-like sensors [165], [166]. In urban mmWave and sub-THz networks, ISAC can enable real-time tracking of vehicles and pedestrians, enhancing sensor fusion algorithms without compromising the network's primary communication responsibilities. Future research directions may focus on leveraging high angular resolution from densely packed phased arrays to achieve precise beamforming, with the potential to introduce imaging-like capabilities within communication networks [167], [168].

2) Emerging DT applications

With the increasing demand for urban passenger and goods delivery, emerging technology like urban sensing, electrification, connectivity and autonomy, and robotics have been gradually transforming urban streetscapes, which pose new challenges to the operation and management of urban infrastructure and public space. DTs are crucial to improve urban safety and mobility in infrastructure planning, service operation and management, and ultimately, policy making.

Here we do not aim to enumerate comprehensive urban transportation applications with DTs, since there exist survey papers [13], [14], [4], [5], [6], which summarize those in operation (e.g., anomaly detection and warning, emergency response), maintenance, and mobility (e.g., transit operation, or driving). Instead, we focus on emerging applications in urban settings, and the potentials the DT holds for them. In Fig. 13, emerging use cases are categorized based on required communication latency (x-axis), spatial resolution (y-axis), and data bandwidth (z-axis), respectively.

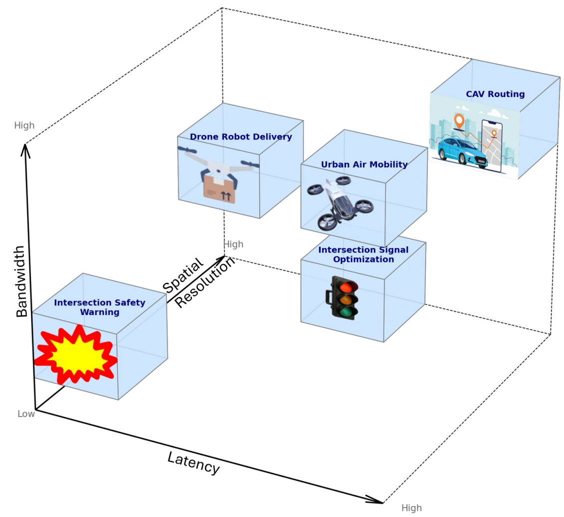


Fig. 13: Urban T-DT use cases.

Stochasticity arising from travel demands, un-predicted VRU movement, and traffic gridlocks requires traffic operators to design adaptive traffic signal control (ATSC). With a large amount of agents (including VRUs, traffic lights) continuously interacting in a stochastic environment, classical optimization tools in deterministic environments could fail to capture such complex decision-making processes. There is a surge in employing learning methods, including reinforcement learning and federated learning, to optimize traffic signals [169], [170], [171], [172], [173], [174], [175], [176], [177]. With an increasing concern in data governance and privacy, federated learning witnesses a growing trend that allows optimizing centralized control while preserving distributed data privacy [178], [179], [180], [164], [181], [182].

Driven by growing urban populations, low-altitude economy aims to exploit vertical space in urban environments, advanced by drones and eVTOL (electric vertical take-off and landing). It will foster wide applications, ranging from urban air mobility, urban logistics, agriculture, emergency service, to infrastructure inspection, surveillance and security. Critical challenges arise in sensing, modeling, predicting urban aerodynamics, and collision avoidance with high-rise buildings and flying objects. Open questions include market demand estimation, infrastructure planning, operation and design. DTs hold the potential to the adoption of these emerging technologies, for its power in simulating emerging demand patterns, modeling their kinetics and dynamics, routing and charging behaviors. To build such DTs, CPS technologies and AI methodologies would facilitate precise sensing and perception [146], reliable and fast communication and networking [183], seamless multimodal coordination (such as drone-truck-mobile robot delivery) [184], as well as safe and optimal decisions [185].

3) AI models

AI methods have been applied to every stage of the DT pipeline. We will point out key challenges of AI methods in transportation.

Physics-informed AI

When it comes to the development of a T-DT, the mutual interaction between domain knowledge and AI is necessary. Domain knowledge in traffic flow and road safety that has been developed for decades, provides valuable insights into every stage of the DT pipeline. In particular, it helps inform technological advancement and testbed deployment, data collection, model selection and training, resource allocation and optimization, as well as performance metrics selection. For instance, the understanding of how traffic evolves across time and space and how entities interact with one another at a road junction, could guide to what degree of granularity and precision the semantic segmentation should be done in object detection and tracking, how communication and networking resources should be prioritized, and what models and metrics should be used for risk prediction.

Safety guarantee of AI

DTs, heavily relying on hardware, software, and algorithms, are vulnerable to risks posed by probabilistic events. Failures in sensors, communication channels, or algorithms can result in erroneous or missing inputs to downstream prediction or decision-making, ultimately compromising the feedback loop from the DT to the physical. The rise of AI-empowered DTs introduces additional risks, including adversarial attacks [186], [187], [188], ethical concerns [189], [190], [191], and trustworthiness in real-world applications [192], [193], [194]. *How can DTs be designed with provable safety guarantees, accounting for uncertainties, failures, and attacks?* To achieve this, several methods have been proposed, including reachability analysis [195], robust optimization [196], and control barrier/Lyapunov functions [197], [198]. These methods are mathematically rigorous to ensure that DTs generate outputs within safe regimes. With the emergence of AI, adapting these methods to ensure the safety guarantees of AI models is becoming increasingly important. For instance, reachability analysis has been employed in safe reinforcement learning [199], [200]. Robust optimization techniques have also been applied to enhance model robustness towards adversarial attacks [201], [202]. Additionally, control barrier functions and Lyapunov functions have been integrated into neural networks to enforce safety constraints and maintain system stability [203], [204].

Evaluation and validation

As opposed to traditional traffic management systems heavily involved with humans, an AI-powered CPS-enabled traffic management system demands high degree of autonomy, with increased complexity and scale. Since testing traffic management strategies could be unethical and unsafe, DT thus becomes a crucial tool for the test and verification. There does not exist a unified scheme about what to validate, verify, and test in a DT. Here, we would like to decompose this problem into several layers. First, we need to evaluate whether an DT represents its corresponding physical world correctly. Ideally, the digital is expected to be the twin of the physical, accordingly, "Grieves performance test" [1] is a high-level abstract way to compare the difference in the outputs of both the physical and the DT. This is associated with real2sim gap to be defined and elaborated more in the next section.

Second, we need to verify the system behavior of an DT-enabled CPS, and ensure that it performs as desired [205], such as safety, efficiency, accuracy, and timeliness. Such testing is non-trivial, due to the integration of cyber and physical components, as well as their two-way coupling via communication and networking. There are normally four types of tests [205], namely, conformance testing (i.e., whether a system conforms to an expected behavior), robustness testing (i.e., whether a system is robust against stochasticity in environments), fragility testing (i.e., whether the output of a system is robust against perturbation in inputs), and security testing (i.e., whether a system is not affected by cyber-attacks). DTs play important roles in the above tests, while experimental design is crucial to cover comprehensive scenarios leveraging game theory [206] and in recent years, generative models [207].

The presence of human factors along the DT pipeline could complicates the evaluation of cyber-physical-human systems, because of randomness and unpredictability in human behaviors [208]. Humans are not only participants in traffic (as drivers, pedestrians, cyclists), they are also creators and users of DTs. For DTs that involve the feedback to humans, human-in-the-loop test is widely used [209], [210], [208]. Recent years have witnessed a growing trend of using augmented reality and virtual reality to engage humans as pedestrians, cyclists, or scooters in virtual environments without inducing real-world risks [134], [211]. Hardware-in-the-loop testing, such as vehicle-in-the-loop [212], could help test the performance of some physical components "live," while allowing the rest to be simulated within a DT. Since there does not exist a unified approach for the DT testing, evaluation methods, metrics, testing platforms should be developed to facilitate standardized assessment and validation of AI models.

Benchmark datasets and methods

To advance the application of AI in DT, we must "stand on the shoulders of giants." In other words, each team does not simply develop one DT for its own internal use. Instead, we hope that an AI-powered DT could be generalized to diverse tasks, transferred to diverse spatiotemporal settings, and shared for co-development among global researchers. Accordingly, we need to standardize application scenarios, benchmark datasets and methods, and unified test environments, for repetitive training and test [213]. Benchmark datasets and methods necessitate performance comparison of any newly proposed AI methods against the state-of-the-art (SOTA) methods. Thus, the transportation community must push to open source data, codes, simulation, algorithms, and results for replicability. For example, standard test platforms, such as Gym [214], Flower [215], PettingZoo [216], [217], and Ray RLlib [218], are important to benchmark various AI models and methods, which is mostly missing in the transportation community.

Moreover, open mixed-perspective datasets that adapt to diverse deployment conditions remain scarce. Advances in open-vocabulary object detection offer a promising avenue to address this gap [219], [220], [221].

B. Challenges and open questions

1) Closing real-to-sim-to-real gap

A key challenge in applying models trained in simulated environments to the real-world is *domain shift* between the training and test environments. Models trained in simulators might likely experience performance degradation when tested in the real world, where the environment includes unpredictable variations that simulators cannot fully replicate. Such a shift – caused by discrepancies between simulated and real-world conditions – can undermine the generalization of models, resulting in reduced performance in scenarios not represented in training. *What distinguishes an DT from a conventional computer simulator?* We believe the key lies in its capability of generating simulations and predictions consistent with the real world, as well as preserving intervention performance. Accordingly, two gaps exist while establishing a DT, namely, real-to-sim (real2sim) gap (i.e., the deviation of the simulated digital world from the real-world), and sim-to-real (sim2real) gap (i.e., the performance deviation of interventions implemented in the real-world from those simulated in the digital world). The smaller these two gaps are, the closer an DT is to reality.

Although these gaps penetrate through each stage along the DT pipeline, from computer vision, tracking, to prediction, and intervention, sim2real transfer is more studied in object detection [222] and policy learning [223], [224], [225], [226]. The major application area of sim2real transfer in DT is robotic manipulation [227], while that in urban navigation and autonomous driving has witnessed a gradual surge, participially in computer vision [228] and reinforcement learning [229], [230]. Key methods include *domain randomization* [231], [207], which introduces variations to augment training data and expose the model to a wide range of scenarios; and *domain adaptation* [232], [233], a technique of finding mappings to transfer data points observed empirically from two different data distributions.

How do we characterize real2sim gap and control such a gap? This boils down to quantifying errors of the digital representation of a physical world. Minimizing the real2sim gap is key to system identification and representation learning. Depending on observability and internal workings, a system could be modeled as white-box, grey-box, or black-box [234]. Domain knowledge could help represent the physical world with higher accuracy. For example, hybrid twin [235], [236] relies on both data-driven and physics-informed AI [237]. Since calibration of a full model could be time-consuming, expensive, and potentially infeasible, the model reduction philosophy has become popular. Digital shadow [11], a digital model with one-way data exchange from the physical to the digital, takes less effort to build, but could fail to update the state of the physical world once feedback is executed. Digital cousin [238] aims not to build a simulation model that replicates the reality exactly, instead, mainly focus on end-to-end gap, namely, from real-world sensing to intervention.

2) Prototyping DTs for human intelligence

Many studies define a DT as emerging technologies, namely, object detection and tracking with edge-cloud computing and communication. These technological enablers, however, are essentially “eyes” of a DT, while what really distinguishes a DT from traditional simulators lies in its “brain,” the prediction and decision making modules that are capable of extracting patterns, modeling semantics, and making informed decisions drawing upon what has been seen and perceived. *How do we establish a foundational DT that develops human-like intelligence with machine automation?* A DT with comparable human-like intelligence or artificial general intelligence (AGI) should consist of a hierarchical cognitive structure analogous to a human’s neural system, backed up by emerging hardware, software, AI algorithms, and API interfaces. The architecture of a proposed DT could consist of:

1) **Eyes**: object detection and tracking, and perception, powered by convolutional neural networks;
2) **Neural systems**: edge-cloud networking and computing backbone, powered by resource allocators [239] and cognitive DT [240];
3) **Brain**: data storage and processing, reasoning and planning, inference and generalization, powered by causal inference and counterfactual analysis [241];
4) **Communication and reasoning**: natural language processing and vision reasoning, powered by generative AI (GenAI) [242].

A T-DT could be deemed as the world model of a transportation manager. A world model is a mental model learned by an AI agent to simulate the evolution of its environment for action planning and reasoning. It is a special type of DT that relies on the agent’s own sensor information. Moreover, it can be embedded into a DT as the AI agent’s internal, abstract representation of the physical world. A world model emerges from the field of robotic learning, and there is an emerging trend to augment a world model with cognitive and reasoning capabilities for more accurate representation and prediction [243]. Such a trend, we strongly believe, must be the pathway for the next-generation of DTs, despite that DTs could be an external representation of a system that would facilitate engineers to monitor, diagnose, control, and manage the system.

In particular, foundational and GenAI models, which are shown to empower cognitive and reasoning architectures of the world model, have demonstrated great potential in DTs [244], [245], [246]. Large language models (LLM) have been used to generate new data for training [247], [248], [249], enhance interactions between human users and the DT system [250], making personalized recommendations [251], [252] and even automate code generation [253] and creation process of DTs [254]. On the other hand, vision language models (VLM) help augment training datasets for robustness, including generation of critical events [255], videos [256], [257] and simulations [258], as well as visual question answering [259], [260]. Thus, we believe that GenAI-powered DT will be the next-generation of DTs for efficient and safe traffic management.

3) Limits in AI

Despite the promising future of AI-powered DTs, application of AI to DTs could face challenges. As opposed to classical statistical methods, AI algorithms are generally black boxes where their inner workings are not transparent to developers nor users. Thus, interpretability or explainability using Shapley value [261], PIDL [262], symbolic regression [263], or Kolmogorov-Arnold networks [264] can potentially reveal to some degree the rationale underlying the predictions. Unlike humans, AI models do not understand causes and heavily rely on correlations to make predictions. Without knowing causality could render AI methods incapable of generalizing to unseen data. Thus, augmenting AIs with causal reasoning and inference could increase its deductive capabilities [265], [266], [267]. In addition, generative AIs could produce hallucination, which might lead to nonphysical predictions or unrealistic decision making. Introducing physics based domain knowledge could help fine tune these models, enhance inductive biases, and generate more meaningful outputs. On the other hand, the emergence of LLM could facilitate the alignment of AI models with human preference. For example, reinforcement learning from human feedback [268] has seen a rapid growth for preferential learning in autonomous driving [269], as well as the generation of more realistic traffic DTs [270].

Last but not the least, there has been a growing trend in developing safe AI systems aligned with human values and objectives. For example, recent studies have focused on evaluating LLMs in terms of toxicity [271], privacy [272], ethics [273], and fairness [274], indicating that LLMs are not sufficiently safe. It is thus crucial to continually achieve safety guarantees of AI, especially for the implementation of DTs.

In a nutshell, transportation applications in urban settings are generally challenging to design, develop, deploy, and test for their potential unsafe and unethical consequences. Thus, AI-empowered DT plays a critical role in effective implementation of these applications. Despite relatively sparse literature in this domain, we review an ensemble body of literature on how to leverage emerging technologies in sensing, communication, edge and cloud computing, for urban traffic management. We hope this paper can serve as a pointer to help researchers and practitioners understand SOTA methods and gaps on the development of DTs; a bridge to initiate conversations across interdisciplinary researchers; and a road map to exploiting potentials of DTs for urban transportation applications.

REFERENCES

[1] M. W. Grieves, "Product lifecycle management: the new paradigm for enterprises," *International Journal of Product Development*, vol. 2, no. 1-2, pp. 71–84, 2005.

[2] L. Li, S. Aslam, A. Wileman, and S. Perinpanayagam, "Digital twin in aerospace industry: A gentle introduction," *IEEE Access*, vol. 10, pp. 9543–9562, 2021.

[3] X. Liao, Z. Wang, X. Zhao, K. Han, P. Tiwari, M. J. Barth, and G. Wu, "Cooperative ramp merging design and field implementation: A digital twin approach based on vehicle-to-cloud communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 5, pp. 4490–4500, 2021.

[4] C. Schwarz and Z. Wang, "The role of digital twins in connected and automated vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 14, no. 6, pp. 41–51, 2022.

[5] Z. Wang, R. Gupta, K. Han, H. Wang, A. Ganlath, N. Ammar, and P. Tiwari, "Mobility digital twin: Concept, architecture, case study, and future challenges," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 452–17 467, 2022.

[6] S. M. Hossain, S. K. Saha, S. Banik, and T. Banik, "A new era of mobility: Exploring digital twin applications in autonomous vehicular systems," in *IEEE AIIoT*. IEEE, 2023, pp. 0493–0499.

[7] Z. Zheng, X. Han, X. Xia, L. Gao, H. Xiang, and J. Ma, "Opencda-ros: Enabling seamless integration of simulation and real-world cooperative driving automation," *IEEE Transactions on Intelligent Vehicles*, 2023.

[8] Y. Li and W. Zhang, "Traffic flow digital twin generation for highway scenario based on radar-camera paired fusion," *Scientific reports*, vol. 13, no. 1, p. 642, 2023.

[9] USDOT FHWA, "Vulnerable road user safety assessment guidance," *Memorandum*, 2022.

[10] M. National Academies of Sciences, Engineering *et al.*, "Foundational research gaps and future directions for digital twins," 2023.

[11] M. Grieves, "Digital model, digital shadow, digital twin," *Preprint*, 2023.

[12] C. P. S. P. W. Group, "Framework for cyber-physical systems release 1.0," *NIST Special Publication 1500-201*, 2016.

[13] M. S. Irfan, S. Dasgupta, and M. Rahman, "Towards transportation digital twin systems for traffic safety and mobility: A review," *IEEE Internet of Things Journal*, 2024.

[14] S. Werbińska-Wojciechowska, R. Giel, and K. Winiarska, "Digital twin approach for operation and maintenance of transportation system—systematic review," *Sensors*, vol. 24, no. 18, p. 6069, 2024.

[15] Z. Liu, H. Sun, G. Marine, and H. Wu, "6g iov networks driven by rf digital twin modeling," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 3, pp. 2976–2986, 2023.

[16] J. Van Brummelen, M. O'brien, D. Gruyer, and H. Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transportation research part C*, vol. 89, pp. 384–406, 2018.

[17] C. Zhao, D. Ding, Z. Du, Y. Shi, G. Su, and S. Yu, "Analysis of perception accuracy of roadside millimeter-wave radar for traffic risk assessment and early warning systems," *International journal of environmental research and public health*, vol. 20, no. 1, p. 879, 2023.

[18] M. Chiani and A. Giorgetti, "Coexistence between uwb and narrow-band wireless communication systems," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 231–254, 2009.

[19] R. H. Rasshofer and K. Gresser, "Automotive radar and lidar systems for next generation driver assistance functions," *Advances in Radio Science*, vol. 3, pp. 205–209, 2005.

[20] Z. Liu, M. Wu, K. Zhu, and L. Zhang, "Sensafe: A smartphone-based traffic safety framework by sensing vehicle and pedestrian behaviors," *Mobile Information Systems*, vol. 2016, no. 1, p. 7967249, 2016.

[21] H. Townsend, A. Gatiba, K. Thompson, P. Wang, K. Wunderlich *et al.*, "Summary report on request for information (rfi): Enhancing the safety of vulnerable road users at intersections," United States. Department of Transportation. Intelligent Transportation ..., Tech. Rep., 2023.

[22] Y. Wang, H. Liu, K. Cui, A. Zhou, W. Li, and H. Ma, "m-activity: Accurate and real-time human activity recognition via millimeter wave radar," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8298–8302.

[23] C. Zhang, J. Wei, J. Dai, S. Qu, X. She, and Z. Wang, "A roadside millimeter-wave radar calibration method based on connected vehicle technology," *IEEE Intelligent Transportation Systems Magazine*, vol. 15, no. 3, pp. 117–131, 2022.

[24] C. Zhang, J. Wei, A. S. Hu, and P. Fu, "A novel method for calibration and verification of roadside millimetre-wave radar," *IET Intelligent Transport Systems*, vol. 16, no. 3, pp. 408–419, 2022.

[25] X. Di and R. Shi, "A survey on autonomous vehicle control in the era of mixed-autonomy: From physics-based to ai-guided driving policy learning," *Transportation research part C*, vol. 125, p. 103008, 2021.

[26] S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C.-C. Chen, J. T. Lee, S. Mukherjee, J. Aggarwal, H. Lee, L. Davis *et al.*, "A large-scale benchmark dataset for event recognition in surveillance video," in *CVPR 2011*. IEEE, 2011, pp. 3153–3160.

[27] M. K. Turkcan, S. Narasimhan, C. Zang, G. H. Je, B. Yu, M. Ghasemi, J. Ghaderi, G. Zussman, and Z. Kostic, "Constellation dataset: Benchmarking high-altitude object detection for an urban intersection," *arXiv preprint arXiv:2404.16944*, 2024.

[28] X. Ye, M. Shu, H. Li, Y. Shi, Y. Li, G. Wang, X. Tan, and E. Ding, "Rope3d: The roadside perception dataset for autonomous driving and monocular 3d object detection task," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 21 341–21 350.

[29] A. Alkhateeb, G. Charan, T. Osman, A. Hredzak, J. Morais, U. Demirhan, and N. Srinivas, "Deepsense 6g: A large-scale real-world multi-modal sensing and communication dataset," *IEEE Communications Magazine*, 2023.

[30] T. Chavdarova, P. Baqué, S. Bouquet, A. Maksai, C. Jose, T. Bagautdinov, L. Lettry, P. Fua, L. Van Gool, and F. Fleuret, "Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5030–5039.

[31] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7380–7399, 2021.

[32] U. D. of Transportation Intelligent Transportation Systems Joint Program Office (JPO), "U.s. department of transportation federal highway administration. (2016). next generation simulation (ngsim) vehicle trajectories and supporting data." Aug 2018. [Online]. Available: http://doi.org/10.21949/1504477

[33] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2118–2125.

[34] R. Krajewski, T. Moers, J. Bock, L. Vater, and L. Eckstein, "The round dataset: A drone dataset of road user trajectories at roundabouts in germany," in *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, 2020, pp. 1–6.

[35] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Dota: A large-scale dataset for object detection in aerial images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3974–3983.

[36] O. Zheng, M. Abdel-Aty, L. Yue, A. Abdelraouf, Z. Wang, and N. Mahmoud, "Citysim: a drone-based vehicle trajectory dataset for safety-oriented research and digital twins," *Transportation research record*, vol. 2678, no. 4, pp. 606–621, 2024.

[37] P. Dendorfer, H. Rezatofighi, A. Milan, J. Shi, D. Cremers, I. Reid, S. Roth, K. Schindler, and L. Leal-Taixé, "Mot20: A benchmark for multi object tracking in crowded scenes," *arXiv:2003.09003[cs]*, Mar. 2020, arXiv: 2003.09003. [Online]. Available: http://arxiv.org/abs/1906.04567

[38] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "Mots: Multi-object tracking and segmentation," *arXiv:1902.03604[cs]*, 2019, arXiv: 1902.03604. [Online]. Available: http://arxiv.org/abs/1902.03604

[39] D. Du, Y. Qi, H. Yu, Y. Yang, K. Duan, G. Li, W. Zhang, Q. Huang, and Q. Tian, "The unmanned aerial vehicle benchmark: Object detection and tracking," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 370–386.

[40] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.

[41] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ade20k dataset," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 633–641.

[42] D. Niranjan, B. C. VinayKarthik, and Mohana, "Deep learning based object detection model for autonomous driving research using carla simulator," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, 2021, pp. 1251–1258.

[43] J. Jang, H. Lee, and J.-C. Kim, "Carfree: Hassle-free object detection dataset generation using carla autonomous driving simulator," *Applied Sciences*, vol. 12, no. 1, p. 281, 2021.

[44] M. Lyssenko, C. Gladisch, C. Heinzemann, M. Woehrle, and R. Triebel, "Instance segmentation in carla: Methodology and analysis for pedestrian-oriented synthetic data generation in crowded scenes," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 988–996.

[45] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 102–118.

[46] C. Meng, S. Zhang, H. Wang, K. Gu, T. Wang, and J. Mei, "Synthesizing data for autonomous driving: Multi-agent reinforcement learning meets augmented reality," SAE Technical Paper, Tech. Rep., 2023.

[47] J. L. Gómez, M. Silva, A. Seoane, A. Borrás, M. Noriega, G. Ros, J. A. Iglesias-Guitian, and A. M. López, "All for one, and one for all: Urbansyn dataset, the third musketeer of synthetic driving scenes," *Neurocomputing*, vol. 637, p. 130038, 2025.

[48] M. Fabbri, G. Brasó, G. Maugeri, O. Cetintas, R. Gasparini, A. Ošep, S. Calderara, L. Leal-Taixé, and R. Cucchiara, "Motsynth: How can synthetic data help pedestrian detection and tracking?" in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 849–10 859.

[49] Y. Li, L. Jiang, L. Xu, Y. Xiangli, Z. Wang, D. Lin, and B. Dai, "Matrixcity: A large-scale city dataset for city-scale neural rendering and beyond," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3205–3215.

[50] M. K. Turkcan, I. Li, C. Zang, J. Ghaderi, G. Zussman, and Z. Kostic, "Boundless: Generating photorealistic synthetic data for object detection in urban streetscapes," 2024. [Online]. Available: https://arxiv.org/abs/2409.03022

[51] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.

[52] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," in *CVPR*, 2020.

[53] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[54] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[55] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European conference on computer vision*. Springer, 2020, pp. 213–229.

[56] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," *arXiv preprint arXiv:2010.04159*, 2020.

[57] Z. Zong, G. Song, and Y. Liu, "Detrs with collaborative hybrid assignments training," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 6748–6758.

[58] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "Detrs beat yolos on real-time object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16 965–16 974.

[59] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "Yolov10: Real-time end-to-end object detection," *arXiv preprint arXiv:2405.14458*, 2024.

[60] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 2446–2454.

[61] Y. Wang, V. C. Guizilini, T. Zhang, Y. Wang, H. Zhao, and J. Solomon, "Detr3d: 3d object detection from multi-view images via 3d-to-2d queries," in *Conference on Robot Learning*. PMLR, 2022, pp. 180–191.

[62] C. Yang, Y. Chen, H. Tian, C. Tao, X. Zhu, Z. Zhang, G. Huang, H. Li, Y. Qiao, L. Lu *et al.*, "Bevformer v2: Adapting modern image backbones to bird's-eye-view recognition via perspective supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 17 830–17 839.

[63] H. Liu, Y. Teng, T. Lu, H. Wang, and L. Wang, "Sparsebev: High-performance sparse 3d object detection from multi-camera videos," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 18 580–18 590.

[64] X. Liu, N. Xue, and T. Wu, "Learning auxiliary monocular contexts helps monocular 3d object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1810–1818.

[65] A. Kumar, G. Brazil, E. Corona, A. Parchami, and X. Liu, "Deviant: Depth equivariant network for monocular 3d object detection," in *European Conference on Computer Vision*. Springer, 2022, pp. 664–683.

[66] Z. Li, J. Jia, and Y. Shi, "Monolss: Learnable sample selection for monocular 3d detection," in *2024 International Conference on 3D Vision (3DV)*. IEEE, 2024, pp. 1125–1135.

[67] L. Yang, K. Yu, T. Tang, J. Li, K. Yuan, L. Wang, X. Zhang, and P. Chen, "Bevheight: A robust framework for vision-based roadside 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 21 611–21 620.

[68] H. Shi, C. Pang, J. Zhang, K. Yang, Y. Wu, H. Ni, Y. Lin, R. Stiefelhagen, and K. Wang, "Cobev: Elevating roadside 3d object detection with depth and height complementarity," *IEEE Transactions on Image Processing*, 2024.

[69] J. Jinrang, Z. Li, and Y. Shi, "Monouni: A unified vehicle and infrastructure-side monocular 3d object detection network with sufficient depth clues," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[70] X. Chen, T. Zhang, Y. Wang, Y. Wang, and H. Zhao, "Futr3d: A unified sensor fusion framework for 3d detection," in *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 172–181.

[71] Z. Lin, Z. Liu, Z. Xia, X. Wang, Y. Wang, S. Qi, Y. Dong, N. Dong, L. Zhang, and C. Zhu, "Rcbevdet: Radar-camera fusion in bird's eye view for 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 14 928–14 937.

[72] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple online and realtime tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, Sep. 2016. [Online]. Available: http://dx.doi.org/10.1109/ICIP.2016.7533003

[73] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," 2017. [Online]. Available: https://arxiv.org/abs/1703.07402

[74] Y. Zhang, P. Sun, Y. Jiang, D. Yu, F. Weng, Z. Yuan, P. Luo, W. Liu, and X. Wang, "Bytetrack: Multi-object tracking by associating every detection box," 2022. [Online]. Available: https://arxiv.org/abs/2110.06864

[75] N. Aharon, R. Orfaig, and B.-Z. Bobrovsky, "Bot-sort: Robust associations multi-pedestrian tracking," 2022. [Online]. Available: https://arxiv.org/abs/2206.14651

[76] V. D. Stanojevic and B. T. Todorovic, "Boosttrack: Boosting the similarity measure and detection confidence for improved multiple object tracking - machine vision and applications," Aug 2024. [Online]. Available: https://link.springer.com/article/10.1007/s00138-024-01531-5

[77] V. Stanojević and B. Todorović, "Boosttrack++: using tracklet information to detect more objects in multiple object tracking," 2024. [Online]. Available: https://arxiv.org/abs/2408.13003

[78] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 961–971.

[79] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," 2018. [Online]. Available: https://arxiv.org/abs/1803.10892

[80] J. Yue, D. Manocha, and H. Wang, "Human trajectory prediction via neural social physics," 2023. [Online]. Available: https://arxiv.org/abs/2207.10435

[81] Z. Sun, Z. Wang, L. Halilaj, and J. Luettin, "Semanticformer: Holistic and semantic traffic scene representation for trajectory prediction using knowledge graphs," 2024. [Online]. Available: https://arxiv.org/abs/2404.19379

[82] P. Kothari, S. Kreiss, and A. Alahi, "Human trajectory forecasting in crowds: A deep learning perspective," 2021. [Online]. Available: https://arxiv.org/abs/2007.03639

[83] K. Guo, W. Liu, and J. Pan, "End-to-end trajectory distribution prediction based on occupancy grid maps," 2022. [Online]. Available: https://arxiv.org/abs/2203.16910

[84] M. Schäfer, K. Zhao, and A. Kummert, "Caspnet++: Joint multi-agent motion prediction," 2023. [Online]. Available: https://arxiv.org/abs/2308.07751

[85] D. Park, H. Ryu, Y. Yang, J. Cho, J. Kim, and K.-J. Yoon, "Leveraging future relationship reasoning for vehicle trajectory prediction," 2023. [Online]. Available: https://arxiv.org/abs/2305.14715

[86] X. Lin, T. Liang, J. Lai, and J.-F. Hu, "Progressive pretext task learning for human trajectory prediction," 2024. [Online]. Available: https://arxiv.org/abs/2407.11588

[87] L. Feng, M. Bahari, K. M. B. Amor, Éloi Zablocki, M. Cord, and A. Alahi, "Unitraj: A unified framework for scalable vehicle trajectory prediction," 2024. [Online]. Available: https://arxiv.org/abs/2403.15098

[88] S. Laskaridis, S. I. Venieris, M. Almeida, I. Leontiadis, and N. D. Lane, "SPINN: synergistic progressive inference of neural networks over device and cloud," in *Proc. ACM MobiCom*, 2020.

[89] L. Zhang, Y. Zhong, J. Liu, and L. Cui, "Resource and bandwidth-aware video analytics with adaptive offloading," in *Proc. IEEE MASS*, 2023.

[90] L. Wang, K. Lu, N. Zhang, X. Qu, J. Wang, J. Wan, G. Li, and J. Xiao, "Shoggoth: towards efficient edge-cloud collaborative real-time video inference via adaptive online learning," in *Proc. ACM/IEEE DAC*, 2023.

[91] W. Liu, J. Geng, Z. Zhu, J. Cao, and Z. Lian, "Sniper: Cloud-edge collaborative inference scheduling with neural network similarity modeling," in *Proc. ACM/IEEE DAC*, 2022.

[92] Z. Yang, W. Ji, Q. Guo, and Z. Wang, "JAVP: Joint-aware video processing with edge-cloud collaboration for DNN inference," in *Proc.ACM MM*, 2023.

[93] K.-J. Fu, Y.-T. Yang, and H.-Y. Wei, "Split computing video analytics performance enhancement with auction-based resource management," *IEEE Access*, vol. 10, pp. 106 495–106 505, 2022.

[94] K. Yang, J. Liu, D. Yang, H. Wang, P. Sun, Y. Zhang, Y. Liu, and L. Song, "A novel efficient multi-view traffic-related object detection framework," in *Proc. IEEE ICASSP*, 2023.

[95] F. C. Akyon, S. O. Altinuc, and A. Temizel, "Slicing aided hyper inference and fine-tuning for small object detection," in *Proc. IEEE ICIP*, 2022.

[96] H. Guo, S. Yao, Z. Yang, Q. Zhou, and K. Nahrstedt, "CrossRoI: Cross-camera region of interest optimization for efficient real-time video analytics at scale," in *Proc. ACM MMSys*, 2021.

[97] W. Zhang, Z. He, L. Liu, Z. Jia, Y. Liu, M. Gruteser, D. Raychaudhuri, and Y. Zhang, "Elf: accelerate high-resolution mobile deep vision with content-aware parallel offloading," in *Proc. ACM MobiCom*, 2021, pp. 201–214.

[98] S. Liu, T. Wang, J. Li, D. Sun, M. Srivastava, and T. Abdelzaher, "AdaMask: Enabling machine-centric video streaming with adaptive frame masking for DNN inference offloading," in *Proc. ACM MM*, 2022.

[99] X. Dai, P. Yang, X. Zhang, Z. Dai, and L. Yu, "Respire: Reducing spatial-temporal redundancy for efficient edge-based industrial video analytics," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9324–9334, 2022.

[100] L. Zhang, J. Xu, Z. Lu, and L. Song, "CrossVsion: Real-time on-camera video analysis via common roi load balancing," *IEEE Trans. Mobile Comput.*, 2023.

[101] H. Wang, Q. Li, H. Sun, Z. Chen, Y. Hao, J. Peng, Z. Yuan, J. Fu, and Y. Jiang, "VaBUS: Edge-cloud real-time video analytics via background understanding and subtraction," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 90–106, 2022.

[102] J. Lin, P. Yang, W. Wu, N. Zhang, T. Han, and L. Yu, "Learning-based query scheduling and resource allocation for low-latency mobile edge video analytics," *IEEE Internet Things J.*, 2023.

[103] Z. Wang, S. Zhang, J. Cheng, Z. Wu, Z. Cao, and Y. Cui, "Edge-assisted adaptive configuration for serverless-based video analytics," in *Proc. IEEE ICDCS*, 2023.

[104] Y. Kong, P. Yang, and Y. Cheng, "Edge-assisted on-device model update for video analytics in adverse environments," in *Proc. ACM MM*, 2023.

[105] T. Murad, A. Nguyen, and Z. Yan, "DAO: Dynamic adaptive offloading for video analytics," in *Proc. ACM MM*, 2022.

[106] K. Du, Q. Zhang, A. Arapin, H. Wang, Z. Xia, and J. Jiang, "Accmpeg: Optimizing video encoding for accurate video analytics," in *Proc. MLSys*, 2022.

[107] D. Wu, D. Zhang, M. Zhang, R. Zhang, F. Wang, and S. Cui, "ILCAS: Imitation learning-based configuration-adaptive streaming for live video analytics with cross-camera collaboration," *IEEE Trans. Mobile Comput.*, 2023.

[108] S. Cen, M. Zhang, Y. Zhu, and J. Liu, "AdaDSR: Adaptive configuration optimization for neural enhanced video analytics streaming," *IEEE Trans.Internet of Things Journal*, 2023.

[109] T. Yuan, L. Mi, W. Wang, H. Dai, and X. Fu, "AccDecoder: Accelerated decoding for neural-enhanced video analytics," in *IEEE INFOCOM*, 2023.

[110] M. Wong, M. Ramanujam, G. Balakrishnan, and R. Netravali, "MadEye: Boosting live video analytics accuracy with adaptive camera configurations," *arXiv preprint arXiv:2304.02101*, 2023.

[111] E. Jinlong, L. He, Z. Li, and Y. Liu, "WiseCam: wisely tuning wireless pan-tilt cameras for cost-effective moving object tracking," in *IEEE INFOCOM*, 2023.

[112] C. Yao, W. Liu, W. Tang, and S. Hu, "EAIS: Energy-aware adaptive scheduling for CNN inference on high-performance GPUs," *Elsevier Future Gen. Comput. Sys.*, vol. 130, pp. 253–268, 2022.

[113] C. Yao, W. Liu, Z. Liu, L. Yan, S. Hu, and W. Tang, "EALI: Energy-aware layer-level scheduling for convolutional neural network inference services on GPUs," *Elsevie Neurocomput.*, vol. 507, pp. 265–281, 2022.

[114] H. A. Hassan, S. A. Salem, and E. M. Saad, "A smart energy and reliability aware scheduling algorithm for workflow execution in DVFS-enabled cloud environment," *Elsevier Future Gener. Comput. Syst.*, vol. 112, pp. 431–448, 2020.

[115] P. Popovski, v. Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, and K. F. Trillingsgaard, "Wireless Access in Ultra-Reliable Low-Latency Communication (URLLC)," *IEEE Transactions on Communications*, vol. 67, no. 8, pp. 5783–5801, 2019.

[116] "Understanding important 5G concepts: What are eMBB, URLLC and mMTC?" https://www.verizon.com/about/news/5g-understanding-embb-urllc-mmtc, 2023.

[117] "New Services & Applications with 5G Ultra-Reliable Low Latency Communications," https://www.5gamericas.org/wp-content/uploads/2019/07/5G_Americas_URLLLC_White_Paper_Final__updateJW.pdf, 2019.

[118] S. Eswaran and P. Honnavalli, "Private 5G networks: a survey on enabling technologies, deployment models, use cases and research directions," *Springer Telecommunications Systems*, vol. 82, pp. 3–26, 2023.

[119] "Verizon to fit Audi's test track with 5G for smart vehicle testing," https://www.reuters.com/technology/verizon-fit-audis-test-track-with-5g-smart-vehicle-testing-2024-02-22/, 2024.

[120] G. Patti, L. Leonardi, G. Testa, and L. L. Bello, "PrioMQTT: A prioritized version of the MQTT protocol," *Elsevier Comput. Commun.*, vol. 220, pp. 43–51, 2024.

[121] Z. Shelby, K. Hartke, and C. Bormann, "The constrained application protocol (coap)," Internet Engineering Task Force, Request for Comments 7252, 2014. [Online]. Available: https://datatracker.ietf.org/doc/html/rfc7252

[122] R. Fielding, J. Gettys, J. C. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee, "Hypertext transfer protocol – http/1.1," Internet Engineering Task Force, Request for Comments 2616, 1999. [Online]. Available: https://www.rfc-editor.org/rfc/rfc2616

[123] Anonymous, "Reducing Communication Overhead in the IoT–Edge–Cloud," *arXiv preprint arXiv:2404.19492*, 2024, provides comparative header and communication model analysis.

[124] C. Caiazza, V. Luconi, and A. Vecchio, "Energy consumption of smartphones and IoT devices when using different versions of the HTTP protocol," *arXiv preprint arXiv:2502.19997*, 2025, examines energy tradeoffs for HTTP/3 in constrained devices.

[125] D. Raychaudhuri, I. Seskar, G. Zussman, T. Korakis, D. Kilper, T. Chen, J. Kolodziejski, M. Sherman, Z. Kostic, X. Gu *et al.*, "Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless," in *Proc. ACM MobiCom*, 2020.

[126] Y. Fu, M. K. Turkcan, V. Anantha, Z. Kostic, G. Zussman, and X. Di, "Digital twin for pedestrian safety warning at a single urban traffic intersection," in *2024 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2024, pp. 2640–2645.

[127] B. Kramer and N. VOlker, *Real-time systems: the international journal of time-critical computing systems*. Springer, 1989.

[128] J. Oliveira, P. Teixeira, P. Rito, M. Luís, S. Sargento, and B. Parreira, "Microservices in edge and cloud computing for safety in intelligent transportation systems," in *NOMS 2024-2024 IEEE Network Operations and Management Symposium*. IEEE, 2024, pp. 1–7.

[129] A. Enan, A. A. Mamun, J. M. Tine, J. Mwakalonge, D. A. Indah, G. Comert, and M. Chowdhury, "Basic safety message generation through a video-based analytics for potential safety applications," *Journal on Autonomous Transportation Systems*, 2024.

[130] A. Napolitano, G. Cecchetti, F. Giannone, A. Ruscelli, F. Civerchia, K. Kondepu, L. Valcarenghi, and P. Castoldi, "Implementation of a mec-based vulnerable road user warning system," in *AEIT AUTOMOTIVE*. IEEE, 2019, pp. 1–6.

[131] I. Lujic, V. D. Maio, K. Pollhammer, I. Brandic, J. Lasic, and I. Brandic, "Increasing traffic safety with real-time edge analytics and 5g," in *Proceedings of the 4th International Workshop on Edge Systems, Analytics and Networking*, 2021, pp. 19–24.

[132] X. Limani, H. C. C. De Resende, V. Charpentier, J. Marquez-Barja, and R. Riggio, "Enabling cross-technology communication to protect vulnerable road users," in *2022 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*. IEEE, 2022, pp. 39–44.

[133] P. Teixeira, S. Sargento, P. Rito, M. Luís, and F. Castro, "A sensing, communication and computing approach for vulnerable road users safety," *IEEE Access*, vol. 11, pp. 4914–4930, 2023.

[134] Z. Wang, O. Zheng, L. Li, M. Abdel-Aty, C. Cruz-Neira, and Z. Islam, "Towards next generation of pedestrian and connected vehicle in-the-loop research: A digital twin co-simulation framework," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 4, pp. 2674–2683, 2023.

[135] IIHS-HLDI, "Fatality facts 2022: Urban/rural comparison," https://www.iihs.org/research-areas/fatality-statistics/detail/urban-rural-comparison, 2024.

[136] FHWA, https://highways.dot.gov/safety/intersection-safety/about#:~:text=Intersecting%20roadways%20are%20necessary%20to,program%20focus%20area%20for%20FHWA., 2024.

[137] K. Ruan and X. Di, "Infostgcan: An information-maximizing spatial-temporal graph convolutional attention network for heterogeneous human trajectory prediction," *Computers*, vol. 13, no. 6, p. 151, 2024.

[138] Z. Mo, Y. Fu, and X. Di, "Pi-neugode: Physics-informed graph neural ordinary differential equations for spatiotemporal trajectory prediction," in *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, 2024, pp. 1418–1426.

[139] Z. Mo, R. Shi, and X. Di, "A physics-informed deep learning paradigm for car-following models," *Transportation research part C: emerging technologies*, vol. 130, p. 103240, 2021.

[140] Z. Mo and X. Di, "Uncertainty quantification of car-following behaviors: physics-informed generative adversarial networks," in *the 28th ACM SIGKDD in conjunction with the 11th International Workshop on Urban Computing (UrbComp2022)*, 2022.

[141] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[142] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision–ECCV 2020*. Springer, 2020, pp. 683–700.

[143] G. Jocher, A. Chaurasia, and J. Qiu, "YOLO by Ultralytics," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[144] Unity Technologies, "Unity (version 2023)," https://unity.com/, 2024.

[145] K. Sun, J. Wu, Q. Pan, X. Zheng, J. Li, and S. Yu, "Leveraging digital twin and drl for collaborative context offloading in c-v2x autonomous driving," *IEEE Transactions on Vehicular Technology*, 2023.

[146] K. Wang, Z. Li, K. Nonomura, T. Yu, K. Sakaguchi, O. Hashash, and W. Saad, "Smart mobility digital twin based automated vehicle navigation system: A proof of concept," *IEEE Transactions on Intelligent Vehicles*, 2024.

[147] M. Bagheri, M. Siekkinen, and J. K. Nurminen, "Cloud-based pedestrian road-safety with situation-adaptive energy-efficient communication," *IEEE Intelligent transportation systems magazine*, vol. 8, no. 3, pp. 45–62, 2016.

[148] S. Y. Gelbal, B. Aksun-Guvenc, and L. Guvenc, "Vulnerable road user safety using mobile phones with vehicle-to-vru communication," *Electronics*, vol. 13, no. 2, p. 331, 2024.

[149] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.

[150] M. Conti, A. Passarella, and F. Pezzoni, "Opportunistic networking: Data forwarding in disconnected mobile ad hoc networks," *IEEE Communications Magazine*, vol. 44, no. 11, pp. 139–145, 2010.

[151] E. Shi, T.-H. H. Chan, E. Rieffel, R. Chow, and D. Song, "Privacy-preserving aggregation of time-series data," in *NDSS*, 2011.

[152] G. P. Stein, "Tracking from multiple view points: Self-calibration of space and time," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, vol. 1. IEEE, 1999, pp. 521–527.

[153] C. Albl, Z. Kukelova, A. Fitzgibbon, J. Heller, M. Smid, and T. Pajdla, "On the two-view geometry of unsynchronized cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4847–4856.

[154] M. Douze, J. Revaud, J. Verbeek, H. Jégou, and C. Schmid, "Circulant temporal encoding for video retrieval and temporal alignment," *International Journal of Computer Vision*, vol. 119, pp. 291–306, 2016.

[155] L. Baraldi, M. Douze, R. Cucchiara, and H. Jégou, "Lamv: Learning to align and match videos with kernelized temporal layers," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7804–7813.

[156] X. Wu, Z. Wu, Y. Zhang, L. Ju, and S. Wang, "Multi-video temporal synchronization by matching pose features of shared moving subjects," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

[157] M. Smid and J. Matas, "Rolling shutter camera synchronization with sub-millisecond accuracy," *arXiv preprint arXiv:1902.11084*, 2019.

[158] R. Latimer, J. Holloway, A. Veeraraghavan, and A. Sabharwal, "Socialsync: Sub-frame synchronization in a smartphone camera network," in *Computer Vision-ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part II 13*. Springer, 2015, pp. 561–575.

[159] G. Litos, X. Zabulis, and G. Triantafyllidis, "Synchronous image acquisition based on network synchronization," in *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. IEEE, 2006, pp. 167–167.

[160] L. Ahrenberg, I. Ihrke, and M. Magnor, "A mobile system for multi-video recording," in *1st European Conference on Visual Media Production (CVMP)*, 2004, pp. 127–132.

[161] S. Ansari, N. Wadhwa, R. Garg, and J. Chen, "Wireless software synchronization of multiple distributed cameras," in *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2019, pp. 1–9.

[162] Y. Liu, J. James, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7751–7763, 2020.

[163] A. El Ouadrhiri and A. Abdelhadi, "Differential privacy for deep and federated learning: A survey," *IEEE access*, vol. 10, pp. 22 359–22 380, 2022.

[164] Y. Fu and X. Di, "Federated reinforcement learning for adaptive traffic signal control: A case study in new york city," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 5738–5743.

[165] F. Liu, Y. Cui, C. Masouros, J. Xu, T. X. Han, Y. C. Eldar, and S. Buzzi, "Integrated sensing and communications: Toward dual-functional wireless networks for 6g and beyond," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1728–1767, 2022.

[166] T. Wild, V. Braun, and H. Viswanathan, "Joint Design of Communication and Sensing for Beyond 5G and 6G Systems," *IEEE Access*, vol. 9, pp. 30 845–30 857, 2021.

[167] A. Paidimarri, A. Tzadok, S. G. Sanchez, A. Kludze, A. Gallyas-Sanhueza, and A. Valdes-Garcia, "Eye-Beam: A mmWave 5G-compliant Platform for Integrated Communications and Sensing Enabling AI-based Object Recognition," *IEEE Journal on Sel, Areas in Comm.*, 2024.

[168] X. Gu, A. Paidimarri, B. Sadhu, C. Baks, S. Lukashov, M. Yeck, Y. Kwark, T. Chen, G. Zussman, I. Seskar *et al.*, "Development of a compact 28-GHz software-defined phased array for a city-scale wireless research testbed," in *Proc. IEEE IMS*, 2021.

[169] H. Kamal, W. Yánez, S. Hassan, and D. Sobhy, "Digital-twin-based deep reinforcement learning approach for adaptive traffic signal control," *IEEE Internet of Things Journal*, 2024.

[170] S. Dasgupta, M. Rahman, and S. Jones, "Harnessing digital twin technology for adaptive traffic signal control: Improving signalized intersection performance and user satisfaction," *IEEE Internet of Things Journal*, 2024.

[171] V. K. Kumarasamy, A. J. Saroj, Y. Liang, D. Wu, M. P. Hunter, A. Guin, and M. Sartipi, "Integration of decentralized graph-based multi-agent reinforcement learning with digital twin for traffic signal optimization," *Symmetry*, vol. 16, no. 4, p. 448, 2024.

[172] S. Khadka, P. Wang, P. Li, and S. P. Mattingly, "Automated traffic signal performance measures (atspms) in the loop simulation: A digital twin approach," *Transportation Research Record*, p. 03611981241258985, 2024.

[173] H. Zhu, F. Sun, K. Tang, H. Wu, J. Feng, and Z. Tang, "Digital twin-enhanced adaptive traffic signal framework under limited synchronization conditions," *Sustainability*, vol. 16, no. 13, p. 5502, 2024.

[174] Z. Mo, W. Li, Y. Fu, K. Ruan, and X. Di, "Cvlight: Decentralized learning for adaptive traffic signal control with connected vehicles," *Transportation research part C: emerging technologies*, vol. 141, p. 103728, 2022.

[175] W. Li, T. Zhu, and Y. Feng, "A cooperative perception based adaptive signal control under early deployment of connected and automated vehicles," *Transportation Research Part C: Emerging Technologies*, vol. 169, p. 104860, 2024.

[176] T. Chu, J. Wang, L. Codecà, and Z. Li, "Multi-agent deep reinforcement learning for large-scale traffic signal control," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 3, pp. 1086–1095, 2019.

[177] J. Guo, L. Cheng, and S. Wang, "Cotv: Cooperative control for traffic light signals and connected autonomous vehicles using deep reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, pp. 10 501–10 512, 2023.

[178] Y. Ye, W. Zhao, T. Wei, S. Hu, and M. Chen, "Fedlight: Federated reinforcement learning for autonomous multi-intersection traffic signal control," in *2021 58th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 2021, pp. 847–852.

[179] N. Hudson, P. Oza, H. Khamfroush, and T. Chantem, "Smart edge-enabled traffic light control: Improving reward-communication trade-offs with federated reinforcement learning," in *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2022, pp. 40–47.

[180] J. Bao, C. Wu, Y. Lin, L. Zhong, X. Chen, and R. Yin, "A scalable approach to optimize traffic signal control with federated reinforcement learning," *Scientific Reports*, vol. 13, no. 1, p. 19184, 2023.

[181] S. Fang, R. Ye, W. Wang, Z. Liu, Y. Wang, Y. Wang, S. Chen, and Y. Wang, "Fedrsu: Federated learning for scene flow estimation on roadside units," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[182] Q. Liu, S. Sun, M. Liu, Y. Wang, and B. Gao, "Online spatio-temporal correlation-based federated learning for traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, 2024.

[183] J. Akram, A. Anaissi, R. S. Rathore, R. H. Jhaveri, and A. Akram, "Digital twin-driven trust management in open ran-based spatial crowdsourcing drone services," *IEEE Transactions on Green Communications and Networking*, 2024.

[184] C. Y. Yiu, K. K. Ng, C.-H. Lee, C. T. Chow, T. C. Chan, K. C. Li, and K. Y. Wong, "A digital twin-based platform towards intelligent automation with virtual counterparts of flight and air traffic control operations," *Applied Sciences*, vol. 11, no. 22, p. 10923, 2021.

[185] M. ElSayed and M. Mohamed, "Robust digital-twin airspace discretization and trajectory optimization for autonomous unmanned aerial vehicles," *Scientific Reports*, vol. 14, no. 1, p. 12506, 2024.

[186] M. Zhou, W. Zhou, J. Huang, J. Yang, M. Du, and Q. Li, "Stealthy and effective physical adversarial attacks in autonomous driving," *IEEE Transactions on Information Forensics and Security*, 2024.

[187] Y. Deng, X. Zheng, T. Zhang, C. Chen, G. Lou, and M. Kim, "An analysis of adversarial attacks and defenses on autonomous driving models," in *2020 IEEE international conference on pervasive computing and communications (PerCom)*. IEEE, 2020, pp. 1–10.

[188] H. Wu, S. Yunas, S. Rowlands, W. Ruan, and J. Wahlström, "Adversarial driving: Attacking end-to-end autonomous driving," in *2023 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2023, pp. 1–7.

[189] B. Giovanola and S. Tiribelli, "Beyond bias and discrimination: redefining the ai ethics principle of fairness in healthcare machine-learning algorithms," *AI & society*, vol. 38, no. 2, pp. 549–563, 2023.

[190] S. Tiribelli, B. Giovanola, R. Pietrini, E. Frontoni, and M. Paolanti, "Embedding ai ethics into the design and use of computer vision technology for consumer's behavior understanding," *Computer Vision and Image Understanding*, p. 104142, 2024.

[191] R. A. Waelen, "The ethics of computer vision: an overview in terms of power," *AI and Ethics*, vol. 4, no. 2, pp. 353–362, 2024.

[192] J. Edstedt, Q. Sun, G. Bökman, M. Wadenbäck, and M. Felsberg, "Roma: Robust dense feature matching," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 19 790–19 800.

[193] N. Rabbani and A. Bartoli, "Unsupervised confidence approximation: Trustworthy learning from noisy labelled data," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4609–4617.

[194] F. Guillaro, D. Cozzolino, A. Sud, N. Dufour, and L. Verdoliva, "Trufor: Leveraging all-round clues for trustworthy image forgery detection and localization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 20 606–20 615.

[195] M. Althoff, "Reachability analysis and its application to the safety assessment of autonomous cars," Ph.D. dissertation, Technische Universität München, 2010.

[196] S. Zhao and K. Zhang, "A distributionally robust stochastic optimization-based model predictive control with distributionally robust chance constraints for cooperative adaptive cruise control under uncertain traffic conditions," *Transportation Research Part B*, vol. 138, pp. 144–178, 2020.

[197] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2016.

[198] B. Li, S. Wen, Z. Yan, G. Wen, and T. Huang, "A survey on the control lyapunov function and control barrier function for nonlinear-affine control systems," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 3, pp. 584–602, 2023.

[199] M. Selim, A. Alanwar, S. Kousik, G. Gao, M. Pavone, and K. H. Johansson, "Safe reinforcement learning using black-box reachability analysis," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 665–10 672, 2022.

[200] X. Wang and M. Althoff, "Safe reinforcement learning for automated vehicles via online reachability analysis," *IEEE Transactions on Intelligent Vehicles*, 2023.

[201] D. Boetius, S. Leue, and T. Sutter, "A robust optimisation perspective on counterexample-guided repair of neural networks," in *International Conference on Machine Learning*. PMLR, 2023, pp. 2712–2737.

[202] A. Ferdowsi, U. Challita, W. Saad, and N. B. Mandayam, "Robust deep reinforcement learning for security and safety in autonomous vehicle systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 307–312.

[203] B. Gangopadhyay, P. Dasgupta, and S. Dey, "Safe and stable rl (s 2 rl) driving policies using control barrier and control lyapunov functions," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 2, pp. 1889–1899, 2022.

[204] Y.-C. Chang, N. Roohi, and S. Gao, "Neural lyapunov control," *Advances in neural information processing systems*, vol. 32, 2019.

[205] R. J. Somers, J. A. Douthwaite, D. J. Wagg, N. Walkinshaw, and R. M. Hierons, "Digital-twin-based testing for cyber–physical systems: A systematic literature review," *Information and Software Technology*, vol. 156, p. 107145, 2023.

[206] Z. Mo, X. Di, and R. Shi, "Robust data sampling in machine learning: A game-theoretic framework for training and validation data selection," *Games*, vol. 14, no. 1, p. 13, 2023.

[207] Z. Mo, H. Xiang, and X. Di, "Diffirm: A diffusion-augmented invariant risk minimization framework for spatiotemporal prediction over graphs," *Transportation Science*, 2025.

[208] X. Di and H. X. Liu, "Boundedly rational route choice behavior: A review of models and methodologies," *Transportation Research Part B: Methodological*, vol. 85, pp. 142–179, 2016.

[209] W. Yin, C. Chai, Z. Zhou, C. Li, Y. Lu, and X. Shi, "Effects of trust in human-automation shared control: A human-in-the-loop driving simulation study," in *IEEE ITSC*. IEEE, 2021, pp. 1147–1154.

[210] Y. Li, Y. Su, X. Zhang, Q. Cai, H. Lu, and Y. Liu, "A simulation system for human-in-the-loop driving," in *IEEE ITSC*. IEEE, 2022, pp. 4183–4188.

[211] K. Kuru, "Metaomnicity: Toward immersive urban metaverse cyberspaces using smart city digital twins," *IEEE Access*, vol. 11, pp. 43 844–43 868, 2023.

[212] D. Xue, J. Cheng, X. Zhao, and Z. Wang, "A vehicle-in-the-loop simulation test based digital-twin for intelligent vehicles," in *IEEE DASC/PiCom/CBDCom/CyberSciTech*. IEEE, 2021, pp. 918–922.

[213] X. Di, S. Qian, and C. Osorio, "Special issue on machine learning methods for urban passenger mobility," *Transportation Science*, vol. 59, no. 4, pp. iii–vi, 2025.

[214] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.

[215] D. J. Beutel, T. Topal, A. Mathur, X. Qiu, J. Fernandez-Marques, Y. Gao, L. Sani, K. H. Li, T. Parcollet, P. P. B. de Gusmão *et al.*, "Flower: A friendly federated learning research framework," *arXiv preprint arXiv:2007.14390*, 2020.

[216] J. Terry, B. Black, N. Grammel, M. Jayakumar, A. Hari, R. Sullivan, L. S. Santos, C. Dieffendahl, C. Horsch, R. Perez-Vicente *et al.*, "Pettingzoo: Gym for multi-agent reinforcement learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 15 032–15 043, 2021.

[217] S. Liu, Y. Wang, X. Chen, Y. Fu, and X. Di, "Smart-eflo: An integrated sumo-gym framework for multi-agent reinforcement learning in electric fleet management problem," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3026–3031.

[218] E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. Jordan, and I. Stoica, "Rllib: Abstractions for distributed reinforcement learning," in *International conference on machine learning*. PMLR, 2018, pp. 3053–3062.

[219] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu *et al.*, "Grounding dino: Marrying dino with grounded pre-training for open-set object detection," *arXiv preprint arXiv:2303.05499*, 2023.

[220] M. Minderer, A. Gritsenko, A. Stone, M. Neumann, D. Weissenborn, A. Dosovitskiy, A. Mahendran, A. Arnab, M. Dehghani, Z. Shen *et al.*, "Simple open-vocabulary object detection," in *European Conference on Computer Vision*. Springer, 2022, pp. 728–755.

[221] T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, and Y. Shan, "Yolo-world: Real-time open-vocabulary object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16 901–16 911.

[222] A. Pitkevich and I. Makarov, "A survey on sim-to-real transfer methods for robotic manipulation," in *2024 IEEE 22nd Jubilee International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, 2024, pp. 000 259–000 266.

[223] G. Peyré, M. Cuturi *et al.*, "Computational optimal transport: With applications to data science," *Foundations and Trends® in Machine Learning*, vol. 11, no. 5-6, pp. 355–607, 2019.

[224] E. Salvato, G. Fenu, E. Medvet, and F. A. Pellegrino, "Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning," *IEEE Access*, vol. 9, pp. 153 171–153 187, 2021.

[225] X. Liu, C. Gong, and Q. Liu, "Flow straight and fast: Learning to generate and transfer data with rectified flow," *arXiv preprint arXiv:2209.03003*, 2022.

[226] L. Da, J. Turnau, T. P. Kutralingam, A. Velasquez, P. Shakarian, and H. Wei, "A survey of sim-to-real methods in rl: Progress, prospects and challenges with foundation models," *arXiv preprint arXiv:2502.13187*, 2025.

[227] J. Abou-Chakra, L. Sun, K. Rana, B. May, K. Schmeckpeper, M. V. Minniti, and L. Herlant, "Real-is-sim: Bridging the sim-to-real gap with a dynamic digital twin for real-world robot policy evaluation," *arXiv preprint arXiv:2504.03597*, 2025.

[228] Z. Xie, Z. Liu, Z. Peng, W. Wu, and B. Zhou, "Vid2sim: Realistic and interactive simulation from video for urban navigation," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 1581–1591.

[229] Z. Hong, "Effective learning mechanism based on reward-oriented hierarchies for sim-to-real adaption in autonomous driving systems," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[230] D. Li and O. Okhrin, "A platform-agnostic deep reinforcement learning framework for effective sim2real transfer towards autonomous driving," *Communications Engineering*, vol. 3, no. 1, p. 147, 2024.

[231] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2017, pp. 23–30.

[232] T.-D. Truong, P. Helton, A. Moustafa, J. D. Cothren, and K. Luu, "Conda: Continual unsupervised domain adaptation learning in visual perception for self-driving cars," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 5642–5650.

[233] S. Xi, Z. Liu, Z. Wang, Q. Zhang, H. Ding, C. C. Kang, and Z. Chen, "Autonomous driving roadway feature interpretation using integrated semantic analysis and domain adaptation," *IEEE Access*, 2024.

[234] M. E. Khan and F. Khan, "A comparative study of white box, black box and grey box testing techniques," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 6, 2012.

[235] F. Chinesta, E. Cueto, E. Abisset-Chavanne, J. L. Duval, and F. El Khaldi, "Virtual, digital and hybrid twins: a new paradigm in data-based engineering and engineered data," *Archives of Computational Methods in Engineering*, pp. 1–30, 2018.

[236] S. Abburu, A. J. Berre, M. Jacoby, D. Roman, L. Stojanovic, and N. Stojanovic, "Cognitwin–hybrid and cognitive digital twins for the process industry," in *IEEE ICE/ITMC*. IEEE, 2020, pp. 1–8.
[237] S. Ma, K. A. Flanigan, and M. Bergés, "Bridging the reality gap in digital twins with context-aware, physics-guided deep learning," *arXiv preprint arXiv:2505.11847*, 2025.
[238] T. Dai, J. Wong, Y. Jiang, C. Wang, C. Gokmen, R. Zhang, J. Wu, and L. Fei-Fei, "Acdc: Automated creation of digital cousins for robust policy learning," *arXiv preprint arXiv:2410.07408*, 2024.
[239] T. Li, Q. Long, H. Chai, S. Zhang, F. Jiang, H. Liu, W. Huang, D. Jin, and Y. Li, "Generative ai empowered network digital twins: Architecture, technologies, and applications," *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–43, 2025.
[240] W. Xu, H. Yang, Z. Ji, and M. Ba, "Cognitive digital twin-enabled multi-robot collaborative manufacturing: Framework and approaches," *Computers & Industrial Engineering*, vol. 194, p. 110418, 2024.
[241] X. Liu, P. Xu, J. Wu, J. Yuan, Y. Yang, Y. Zhou, F. Liu, T. Guan, H. Wang, T. Yu *et al.*, "Large language models and causal inference in collaboration: A comprehensive survey," *Findings of the Association for Computational Linguistics: NAACL 2025*, pp. 7668–7684, 2025.
[242] Y. Shen, H. Ding, L. Seenivasan, T. Shu, and M. Unberath, "Position: Foundation models need digital twin representations," *arXiv preprint arXiv:2505.03798*, 2025.
[243] C. Zhao, R. Zhang, J. Wang, G. Zhao, D. Niyato, G. Sun, S. Mao, and D. I. Kim, "World models for cognitive agents: Transforming edge intelligence in future networks," *arXiv preprint arXiv:2506.00417*, 2025.
[244] S. Gebreab, A. Musamih, K. Salah, R. Jayaraman, and D. Boscovic, "Accelerating digital twin development with generative ai: A framework for 3d modeling and data integration," *IEEE Access*, 2024.
[245] Y. Wang, S. Xing, C. Can, R. Li, H. Hua, K. Tian, Z. Mo, X. Gao, K. Wu, S. Zhou *et al.*, "Generative ai for autonomous driving: Frontiers and opportunities," *arXiv preprint arXiv:2505.08854*, 2025.
[246] L. Yang, S. Luo, X. Cheng, and L. Yu, "Leveraging large language models for enhanced digital twin modeling: Trends, methods, and challenges," *arXiv preprint arXiv:2503.02167*, 2025.
[247] J. Wen, J. Nie, J. Kang, D. Niyato, H. Du, Y. Zhang, and M. Guizani, "From generative ai to generative internet of things: Fundamentals, framework, and outlooks," *IEEE Internet of Things Magazine*, vol. 7, no. 3, pp. 30–37, 2024.
[248] R. Zhang, K. Xiong, H. Du, D. Niyato, J. Kang, X. Shen, and H. V. Poor, "Generative ai-enabled vehicular networks: Fundamentals, framework, and case study," *IEEE Network*, 2024.
[249] H. Chai, H. Wang, T. Li, and Z. Wang, "Generative ai-driven digital twin for mobile networks," *IEEE Network*, 2024.
[250] J. Zhang, J. Pu, J. Xue, M. Yang, X. Xu, X. Wang, and F.-Y. Wang, "Hivegpt: Human-machine-augmented intelligent vehicles with generative pre-trained transformer," *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 3, pp. 2027–2033, 2023.
[251] Z. Yang, X. Jia, H. Li, and J. Yan, "A survey of large language models for autonomous driving," *arXiv preprint arXiv:2311.01043*, 2023.
[252] B. Fang, Z. Yang, S. Wang, and X. Di, "Travellm: Could you plan my new public transit route in face of a network disruption?" *arXiv preprint arXiv:2407.14926*, 2024.
[253] J. Dong and L. Ren, "A digital twin modeling code generation framework based on large language model," in *IECON 2024-50th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2024, pp. 1–4.
[254] S. Ali, P. Arcaini, and A. Arrieta, "Foundation models for the digital twin creation of cyber-physical systems," *arXiv preprint arXiv:2407.18779*, 2024.
[255] Y. Li, Z. Mo, and X. Di, "Safeaug: Safety-critical driving data augmentation from naturalistic datasets," in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2024, pp. 3251–3256.
[256] Y. Fu, Y. Li, and X. Di, "Gendds: Generating diverse driving video scenarios with prompt-to-video generative model," *arXiv preprint arXiv:2408.15868*, 2024.
[257] Y. Fu, A. Jain, X. Chen, Z. Mo, and X. Di, "Drivegenvlm: Real-world video generation for vision language model based autonomous driving," in *2024 IEEE International Automated Vehicle Validation Conference (IAVVC)*. IEEE, 2024, pp. 1–6.
[258] T. Vilas Samak, C. Vilas Samak, B. Li, and V. Krovi, "When digital twins meet large language models: Realistic, interactive, and editable simulation for autonomous driving," *arXiv e-prints*, pp. arXiv–2507, 2025.
[259] T. Wang, J. Li, Z. Kong, X. Liu, H. Snoussi, and H. Lv, "Digital twin improved via visual question answering for vision-language interactive mode in human–machine collaboration," *Journal of Manufacturing Systems*, vol. 58, pp. 261–269, 2021.
[260] M. Ghasemi, Z. Kostic, J. Ghaderi, and G. Zussman, "Edgecloudai: Edge-cloud distributed video analytics," in *Proc. ACM MobiCom*, 2024, pp. 1778–1780.
[261] S. Lundberg, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.
[262] X. Di, R. Shi, Z. Mo, and Y. Fu, "Physics-informed deep learning for traffic state estimation: A survey and the outlook," *Algorithms*, vol. 16, no. 6, p. 305, 2023.
[263] N. Makke and S. Chawla, "Interpretable scientific discovery with symbolic regression: a review," *Artificial Intelligence Review*, vol. 57, no. 1, p. 2, 2024.
[264] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "Kan: Kolmogorov-arnold networks," *arXiv preprint arXiv:2404.19756*, 2024.
[265] K. Ruan and X. Di, "Learning human driving behaviors with sequential causal imitation learning," *the 36th AAAI Conference on Artificial Intelligence*, 2022.
[266] K. Ruan, J. Zhang, X. Di, and E. Bareinboim, "Causal imitation learning via inverse reinforcement learning," in *The Eleventh International Conference on Learning Representations*, 2023.
[267] ——, "Causal imitation for markov decision processes: A partial identification approach," *Advances in Neural Information Processing Systems*, vol. 37, pp. 87 592–87 620, 2024.
[268] T. Kaufmann, P. Weng, V. Bengs, and E. Hüllermeier, "A survey of reinforcement learning from human feedback," 2024.
[269] W. Huang, H. Liu, Z. Huang, and C. Lv, "Safety-aware human-in-the-loop reinforcement learning with shared control for autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, 2024.
[270] Y. Cao, B. Ivanovic, C. Xiao, and M. Pavone, "Reinforcement learning with human feedback for realistic traffic simulation," in *2024 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2024, pp. 14 428–14 434.
[271] Y. P. Chetnani, "Evaluating the impact of model size on toxicity and stereotyping in generative llm," Master's thesis, State University of New York at Buffalo, 2023.
[272] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang, "A survey on large language model (llm) security and privacy: The good, the bad, and the ugly," *High-Confidence Computing*, p. 100211, 2024.
[273] J. C. L. Ong, S. Y.-H. Chang, W. William, A. J. Butte, N. H. Shah, L. S. T. Chew, N. Liu, F. Doshi-Velez, W. Lu, J. Savulescu *et al.*, "Ethical and regulatory challenges of large language models in medicine," *The Lancet Digital Health*, vol. 6, no. 6, pp. e428–e432, 2024.
[274] S. Morales, R. Clarisó, and J. Cabot, "A dsl for testing llms for fairness and bias," in *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems*, 2024, pp. 203–213.

Yongjie Fu received the B.S. degree in Automotive Engineering from Tsinghua University, Beijing, China, in 2019.

He is currently pursuing the Ph.D. degree in the Department of Civil Engineering and Engineering Mechanics at Columbia University. His research interest includes reinforcement learning, digital twin and smart city.

Mehmet Kerem Turkcan received his PhD degree in Electrical Engineering from Columbia University in 2022. He is currently an Associate Research Scientist in the Department of Civil Engineering and Engineering Mechanics at Columbia University.

His research interests include computer vision, deep learning and their applications.

Mahshid Ghasemi received her B.S. degree in Electrical Engineering and a Minor in Computer Science from Sharif University of Technology, Tehran, Iran, in 2019.

She is currently pursuing the Ph.D. degree in the Department of Electrical Engineering at Columbia University, New York, NY, USA. Her research interests are real-time video analytics optimization, edge cloud computing, distributed inference, and artificial intelligence applications.

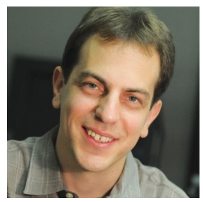
Zhaobin Mo received the B.S. degree in Automotive Engineering from Tsinghua University, Beijing, China, in 2017.

He is currently pursuing the Ph.D. degree in the Department of Civil Engineering and Engineering Mechanics at Columbia University. His research interest includes deep learning, reinforcement learning, and connected and automated vehicle.

Chengbo Zang received the B.S. degree in Automation in Tongji University, Shanghai, China, in 2021. He then received his M.S. in the Department of Electrical Engineering at Columbia University, where he is currently pursuing a Ph.D. His research involves machine learning and deep learning.

Abhishek Adhikari Abhishek Adhikari received the B.S. degree in computer engineering (Magna Cum Laude) from Drexel University in 2021. He is currently an M.S./Ph.D. student in the Department of Electrical Engineering at Columbia University. His research interests include integrated sensing and communication for Beyond-5G millimeter-wave and sub-terahertz wireless networks. Abhishek was the recipient of the Evergreen Fellowship in 2021.

Zoran Kostic Zoran Kostic completed his Ph.D. in Electrical Engineering at the University of Rochester and his Dipl. Ing. degree at the University of Novi Sad. He spent most of his career in industry where he worked in research, product development and in leadership positions. Dr. Kostic is an active member of the IEEE, and he has served as an associate editor of the IEEE Transactions on Communications and IEEE Communications Letters.



Gil Zussman Gil Zussman received a B.Sc. in Industrial Engineering and Management and a B.A. in Economics, both summa cum laude, from the Technion – Israel Institute of Technology in 1995. He earned an M.Sc. in Operations Research from Tel-Aviv University in 1999 (summa cum laude) and a Ph.D. in Electrical Engineering from the Technion in 2004. From 1995 to 1998, he served as an engineer in the Israel Defense Forces. He was a Postdoctoral Associate at MIT's LIDS and CNRG from 2004 to 2007.

Xuan Di received the Ph.D. degree from University of Minnesota, Twin Cities in 2014. She is currently an Associate Professor in the Department of Civil Engineering and Engineering Mechanics and co-chairs the Smart Cities Center in the Data Science Institute at Columbia University.

Her research interests include data-driven urban mobility modeling and cyberphysical systems, leveraging optimization, game theory, and AI. Dr. Di is a member of the IEEE and ACM.