The 7th Open Society Conference (OSC) 2025
Faculty of Law, Social and Political Science
Universitas Terbuka
DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

A Conceptual Framework for Inclusive Smart City Monitoring Using YOLOv10

KianLam Tan[1], ChenKim Lim[2]
[1] School of Digital Technology (DiGiT), Wawasan Open University (WOU), Penang, Malaysia
[2] Institution for Environment and Development (LESTARI), Universiti Kebangsaan Malaysia (UKM), Bangi, Malaysia
e-mail: andrewtan@wou.edu.my; kim@ukm.edu.my

Abstract

The shift towards digital economy has made it increasingly clear that the city digital governance needs to be adaptable enough to ensure that the city it serves is as inclusive and sustainable as possible. However, real time object detection powered by deep learning no longer is a straightforward technical problem of model replacement for computational improvements or inference speedups. It needs to become more scalable, contextually more correct, and fairer as the risks of AI-powered city monitoring to marginalize already underserved and underrepresented communities grows with scope of its deployment and range of its applications. To address this shortcoming, this paper attempts to provide a conceptual framework for responsible deployment of state-of-art YOLOv10 object detection models for situational awareness in the smart cities. Linking the strengths of YOLOv10 (multiscale object detection with edge-compatible architectures and improved contextual understanding) and the key principles of digital inclusion, transparency and effective governance, the proposed conceptual framework will help improve both scalability and the contextuality as well as fairness in deployment of AI-based object detection applications towards informed and responsible urban decision-making. A five-layered conceptual model is offered from data collection to ethical considerations and representative applications of inclusive city digital governance such as accessibility mapping, emergency response and smart waste management. Overall, this work attempts to situate the ongoing discussion of responsible AI in object detection on a firmer foundation with ethical design considerations by drawing meaningful parallels to the emerging field of inclusive urban digital communication.

Keywords: *YOLOv10, smart cities, inclusive urban governance, real-time object detection, edge computing*

INTRODUCTION

The rise of smart cities [1, 2] is a prominent trend of 21st-century urbanism. This shift towards digitally enabled, data-driven, and human-centered urban systems is being actively adopted by governments worldwide to achieve more efficient and predictive governance, cost-effective service delivery, and citizen-centric development. At the core of these emerging 21st-century "smart cities" are the advances in artificial intelligence, the Internet of Things, big data analytics, and systems of real-time monitoring. Artificial intelligence (AI) powered video and image recognition systems are being deployed by cities to address core local challenges from traffic congestion and waste management to air quality monitoring, energy management, or civic

DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

security.

As the 21st-century cities transform and adopt new technologies to become smarter, greener, and cleaner in their functions, a growing concern emerges on how digital innovation can address the increasing social inequality rather than aggravating it. Technology can significantly improve service efficiency but it should reach everyone, regardless of their socio-economic position. This means that the emerging smart city tools should not only be intelligent and fast but also inclusive in their design and implementation. A smart CCTV network that detects only cars and ignores mobility aids, street vendors, or pedestrians with umbrellas is effectively excluding a part of the population from representation in smart city infrastructure. The same is true for other classes of technology: software for waste management that does not detect 1-liter water bottles from the field of view of a mobile garbage truck camera, or air quality monitoring systems that misclassify common cultural practices such as burning leaves or smoke from specific religious events as emergencies are both examples of technology failing to reflect the actual diversity of an urban population.

Object detection systems [3], if intelligently designed and implemented, can help serve as a foundational AI-powered digital infrastructure of an inclusive smart city to allow a digital twin of the city to "see" and respond to various urban processes in near-real time. From pedestrian counting to contextual understanding and labeling of potential hazards, obstacles, or mobility assistance systems in public spaces, detection systems provide the situational awareness required for inclusive public service delivery, community engagement, and effective urban planning.

The goal of this paper is to outline a conceptual framework for integrating YOLOv10 [4], a state-of-art real time object detection algorithm, into a smart urban environment. This effort is motivated by several factors, most prominently that YOLOv10 is exceptionally scalable for large deployments at scale into city infrastructure. With its strong real time inference at high levels of accuracy and low computational requirements, this object detection model can be a core part of not only improving service efficiencies but also transparency and inclusion.

THE ROLE OF OBJECT DETECTION IN INCLUSIVE URBAN SYSTEMS

Cities are becoming ever more diverse as newcomers from other cities or regions join the populations already marked by differences in age, mobility, class, gender, and cultural or ethnic background. This increased heterogeneity of city dwellers is not necessarily reflected in the public services and infrastructures, which tend to cater to people with relatively mainstream mobility, location preferences, and spending capacities. Visiting public parks or open spaces, using public transportation, or simply walking on the streets are some of the activities that might be only partially available to all groups of the population. For example, lack of physical accessibility indicators (wheelchair ramps, support rails, ramp sensors), a variety of mobility aids, human-powered vehicles, or built environment indicators (sidewalk widths, steps, street food or vendor stalls) are anecdotally available to people in the areas of disability, gerontology, or ethnicity that are often already underserved. The situation is not getting better for older adults, people with disabilities, and migrants whose interaction with the physical city, however, continues to be sidelined in urban planning and development for want of reliable data.

Object detection technology [5] has already been used in a number of urban planning, resource management, and citizen-facing public services use cases:

- Traffic and pedestrian flows: detecting vehicles on the road and crosswalks to control

DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

traffic lights dynamically, measure traffic volumes, or inform urban mobility planning
- Public safety and emergencies: detection of suspicious/unattended bags, crowd density, or traffic accidents to alert response services in real time
- Sustainability and green city: video-based smart waste management, monitoring of illegal dumping and overflowing bins, and tracking street cleanliness or water wastage
- Accessibility: detection of wheelchairs, walkers, temporary accessibility modifications or tactile tiles for auditing and augmenting public space accessibility and built environment for accessibility
- Informal settlements and community resources: smart mapping of community resources such as benches, lighting, water fountains, or public signage in areas where these are lacking or required

However, in most of these tasks, standard object detection systems were trained on well-balanced but often not entirely representative samples. The result is, however, sometimes failure to detect small, occluded, or otherwise nonstandard but still crucial for underserved communities objects. For example, a traditional object detector will miss a rickshaw parked on a city street or a unique wheelchair design used by a local resident. This lack of awareness of local nuance is both a direct source of limited use of this technology for good (e.g., inferior public service delivery), and an indirect driver of exclusion from representation in digital twin cities (e.g., when policymaking or resource allocation rely on data).

YOLOv10, however, can overcome many of these deficiencies, and enable more inclusive deployment in urban settings by applying its more scalable, multimodal, and context-sensitive systems of object detection.

OVERVIEW OF YOLOv10

YOLOv10 [6, 7] is the latest iteration of one of the most popular and influential series of real time object detection algorithms. Originally developed with the express goal of producing a real-time object detection system that balanced both accuracy and speed, YOLO (You Only Look Once) algorithms have been growing in number since the original YOLO [8] was introduced. Each new version of YOLO is a result of incremental improvement to the existing YOLOv5 to YOLOv9, adding in architectural improvements to better serve the increasing demands of practical deployment.

YOLOv10 is no exception and continues the existing trend of improving on multiscale object detection with additional focus on both reducing the number of required computations and streamlining the architecture to allow for deployment on edge devices.

Complexity of urban environments as spaces in 21st-century smart city systems can be significant. In one single street scene, cities with dense populations and fast traffic at different speeds of mobility interact with open infrastructure. Weather conditions, illumination, sidewalks or street furniture, street vendor locations, children and adults with strollers and pets are just some of the numerous variables that will need to be accounted for when monitoring an urban environment.

Objects of interest in a city environment also go well beyond traditional vehicles and humans, particularly for an object detection system that is supposed to be inclusive in design and implementation. This would require better accuracy in spotting small, occluded or otherwise unrecognizable or atypical objects such as street vendors carts or pop-up street vendors, temporary

DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

barriers or structures, strollers, or even informal signage and objects used in the city that do not always align with the traditional standards and norms that the models have been trained on. YOLOv10 is particularly well-suited to serve in this capacity, with its superior abilities in all of the key components listed above.

YOLOv10 builds on top of recent improvements of the YOLO family, including YOLOv5 [9], YOLOv6 [10], YOLOv7 [11], and YOLOv8 [12] by 1) improving multiscale feature fusion, 2) anchor-free design, 3) deployment optimization for edge devices, and 4) improved labeling strategy and additional attention modules. This work will cover each of these core improvements individually and as applied to the smart cities in practice.

YOLOv10 significantly improves the model's multiscale object detection ability by improving its feature pyramid architecture. This includes changes such as a) cascade feature fusion, b) integrated spatial pyramid pooling (SPP) and c) dynamically adjustable dilation convolution. The net result is that the model will more effectively and accurately detect small objects – such as small vehicles like motorbikes, wheelchairs, traffic cones, trash cans or litter – which were often missed by previous versions in crowded, complex scenes. It also allows for better differentiation of overlapping or tightly clustered objects, which again, is an important use case in dense urban environments.

YOLOv10 uses anchor-free design, moving away from the prior versions which relied on pre-selected anchor boxes to propose potential object bounding boxes for verification. This approach allows for more flexible object detection, more accurate size and aspect ratio estimation, and reduced computational overheads for model training and prediction. This design is more straightforward and generally allows for better generalization to new deployment areas, which is important for any system that may be used in different parts of a city or a collection of cities.

YOLOv10 is designed to work both in the cloud and at the edge, with dedicated model design choices and lightweight (Tiny) versions of the algorithm to help it be as computationally and energy-efficient as possible on the edge. For a surveillance camera, UAV, or mobile vehicle-mounted device, this is important as these edge devices often have much more limited processing power, memory, and energy constraints. YOLOv10 is exceptionally light-weight while still providing competitive accuracy in its (Tiny) variants, making the deployment of this system at massive scale across urban infrastructure much more feasible from a practical standpoint.

YOLOv10 introduces better label assignment strategies that help during training so that the model can localize the appropriate objects more clearly, especially in dense or crowded scenes. There are additional attention modules (AIM) [13] that allow the algorithm to focus on certain areas of the input image or video clip to help provide additional context and signal to the network, which, in turn, can help prioritize more important or foreground objects of relevance to urban systems. For example, the network can be trained to focus on a child crossing the street in the foreground while the moving car behind them is in the background.

YOLOv10 is able to achieve more than 60FPS on a midrange GPU at this time of writing. This would allow it to be used in real-time analytics situations, especially given its considerable size and accuracy advantages over other detectors. This would then allow a city to make smarter, faster responses to any detected emergencies or accessibility violations in urban areas, as well as be used to help dynamically control rapidly changing traffic situations. It is also important to note that this computational performance comes with relatively little drop-off in terms of accuracy, making it a well-rounded model for a wide range of detection use cases.

DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

In a similar way, what makes YOLOv10 a particularly strong fit for the purpose of not only object detection but also inclusive digital governance is that this model, like the prior series, can be fine-tuned on local datasets, which can help adapt it to the socio-cultural and environmental needs of a specific locality. For example, the standard detection model can be further trained on data about local clothing styles, informal economic activities, or non-standard modes of transportation (e.g., tricycles, tuk-tuks) to ensure that the digital twin of the city it is embedded in mirrors the actual diversity of urban residents.

YOLOv10's strong and balanced improvements to accuracy, computational efficiency, and versatility across scales and classes position it to be a valuable cornerstone tool for building an ethical, inclusive, and data-driven smart city [14, 15].

CONCEPTUAL FRAMEWORK FOR YOLOv10

The following five-layered conceptual framework will provide the reader with a proposed approach to deploying YOLOv10 models for real-time object detection [16] in smart city environments with a key principle of inclusion in mind. As such, this framework maps out the technical and governance infrastructure that is necessary to not only collect and process data with the help of object detection algorithms but also ensure that this data is meaningfully filtered and visualized in the manner that ensures that the resulting information will be used to further the cities with fairness and social equity in mind. The layers are set up to complement one another – working in concert to both collect, process, filter, and meaningfully translate data into information and insights that can be effectively acted upon and regularly governed. In this way, the five layers are set up in such a manner that this multi-layered, cyclical architecture will allow for modularity and scalability in the deployment, management, and use of YOLOv10 in both large and small cities of a range of different development, size, and resource.

The first layer of this framework concerns the continuous and ongoing acquisition of data. For object detection, this is of course largely visual data that can then be fed to an inference engine to process. In the case of YOLOv10, this is preferably a near-continuous feed of visual data that has sources in:

- Fixed surveillance cameras from areas such as traffic intersections, at transit hubs or parking lots, at parks and along sidewalks, etc.
- Mobile cameras affixed to public transport buses, garbage or recycling trucks, and other service vehicles
- UAV or "drone" footage from more difficult-to-access areas of the city or under risk of natural or man-made disasters (floods, fires, infrastructure breakdown)
- Citizen-generated data either through mobile phone applications or civic technology platforms

The first principle is already diversity of coverage of data acquisition areas as often urban surveillance systems are more present in downtown or more affluent areas of a city, but tend to leave peripheral neighborhoods, informal settlements, or low-income areas with less monitoring and by extension in our framework, with less intelligence. It is crucial to ensure that camera and sensor coverage or user-incentivized crowd-sourced data generation is established not only in these under-monitored areas but where possible retrofitted to existing infrastructure in these areas.

DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

Visual data can also often be further supplemented by additional context data (or sensor data) from other sources that can further improve the resulting situational awareness of urban systems. For example, ambient air quality and noise sensors, connected or smart traffic counters or dedicated traffic signal video streams might also be integrated.

Collected visual data from the previous layer is now pushed towards processing units where it will be fed into an inference system for YOLOv10 to perform object detection. Depending on the city and use case, this can happen:

- On edge: using the embedded processing units that are co-located with the camera units (advantage of lower latency, reduced network traffic)
- In the cloud: using a more centralized system of servers to both allow for higher computing power and the possibility of larger-scale data fusion
- On a hybrid model: with the balance of cloud flexibility and edge responsiveness

YOLOv10 will perform object detection and classification in real-time, detecting and recognizing a diverse array of objects and entities such as pedestrians, bicycles, wheelchairs, trash bins, road signs, street vendors, construction workers, etc. In particular, YOLOv10's improvements in multiscale feature extraction will allow this system to detect both large prominent items as well as smaller or otherwise minor items in a scene, which is crucial to account for the widest possible set of relevant urban elements in real-time.

Data reduction will be performed at this stage, ensuring that only relevant and important detections are being retained to save on storage and transmission. Video streams will also be compressed and in the case of existing urban CCTV footage, only event-driven or triggered recording (i.e. movement sensors) will be used to keep only relevant data.

Raw detection outputs will be semantically large and unstructured. This next layer will provide semantic filters that can filter the information that passes through them, annotate it, or otherwise prioritize it as per the requirements of a particular use case or set of data governance objectives. The filters may be as simple as:

- Contextual: near a hospital, the system prioritizes detecting mobility aids and older pedestrians or strollers.
- Temporal: at night, the system prioritizes faces in poorly lit public areas.
- Spatial: in a known flood-prone area, the system prioritizes discarded waste or other objects in the way of a storm drain.

Rule-based logic (condition-action statements or spatial-temporal events) as well as machine learning classifiers can be used to manage semantic filters, so the system can learn to flag anomalous situations, policy violation, or other signals that are relevant to inclusion such as the failure to detect consistently blocked wheelchair ramps over a period of time (after being alerted to its existence via semantic filters) may be used to trigger a more general city-wide review of accessibility infrastructure.

Filtered and prioritized information is then organized into structured logs, alerts, or indicators, all of which are use-ready in their turn for humans or automated decision-making layers.

Semantic data is then fed into appropriate visualization platforms, making it consumable to either human or AI decision-making systems. Dashboards are made interactive and customizable for different stakeholders and roles such as:

DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

- Urban planners: monthly/quarterly/annual dashboards that help focus infrastructure planning and budget allocation.
- Emergency services: real-time alerts for various incidents or accidents.
- Community and CBOs: community-based organizations use their own monitoring capacity or visual evidence to demand services or resources.
- Accessibility/Inclusion officers: will be able to check for relevant indicators such as pedestrian mobility, lack of ramps, blind spot mapping, proximity to curb cuts, street furniture use, and crowding in accessibility-sensitive areas such as busy open-air markets, community events, etc.

Dashboards are also interactive, with the system having a drill-down capability to the level of districts or time period of interest, provide predictive forecasting capacity or integrate other layers (such as detected crowds with ambient pollution sensor networks).

Detection data can also be used to retrain YOLOv10 detection models over time, improving accuracy for the local context and correcting any false positive or negative local incidents. For example, failing to properly detect tuk-tuks or mislabeling mobile local food vendors, particularly in informal market areas, will require building local datasets to improve model robustness and representation.

The final and top-most layer of this model is one of governance. It is here that the data (i.e. inputs) and subsequent information (outputs) are governed in a way that both ensures that the system itself does not violate the rights, dignity, or diversity of its residents while also being able to build in the needed feedback loops for intelligent and human-centered control of the broader system.

Key governance considerations of this layer will include:

- Data privacy: anonymization of individuals in the video feeds and data points, secure storage and handling of data, and compliance with relevant data protection regulation such as GDPR (or local equivalents).
- Transparency and auditability: generation of periodic public-facing reports on detection engine performance (false positives/negatives, detection confidence intervals, precision-recall curves, or demographic detection parity metric).
- Community Engagement: community participation in dataset development, feedback mechanisms, and local oversight committees.
- Bias mitigation: ongoing assessment of detection performance on different population segments, i.e. ensuring the system accurately detects mobility devices used by older adults or various types of cultural or ethnic attire in areas with high minority representation.
- Ethical AI design: compliance with national digital ethics guidelines as well as existing international principles such as explainability, accountability, and fairness.

ETHICAL AND TECHNICAL CONSIDERATIONS

The use of object detection systems and smart CCTV networks in urban spaces need to take into account both technical and ethical concerns. Ethical issues include ensuring that citizens' privacy is protected and the risks of surveillance are mitigated by anonymizing faces and other personally identifiable information in both video and data streams and training model datasets

DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

with diverse, representative populations to avoid model bias or failures. Technical considerations for the deployment of such models include the reliability and maintenance of the hardware and software infrastructure including cameras, sensors, storage, network connectivity, and processing units with necessary security measures against cyberattacks and hacks in place. Edge or cloud deployment of inference engines would need to be carefully considered given bandwidth constraints or dynamic offline-online switching in constrained or unstable internet connectivity. Technical deployment will also depend on whether model updates are being pushed from the cloud to the edge units, necessitating continuous training systems on streaming data, or if YOLOv10 is deployed in a fixed "set and forget" manner that is still retrained periodically in the cloud.

APPLICATION SCENARIOS

A range of application scenarios are possible within the YOLOv10-based framework. These can be found across several core areas of service delivery and digital twin of the city across the diverse urban built environment. Monitoring sidewalks for cases of obstructions to pedestrian flow (particularly older adults and disabled mobility) are especially important to enable greater and inclusive civic security to people with mobility issues. Mapping mobile street vendors is another important way to collect evidence for urban planners to understand the use of open space for informal commerce in a way that can be balanced with urban regulation and order without criminalizing informal livelihoods. UAVs can also be mounted with YOLOv10 to more effectively identify stranded people or damaged infrastructure during emergencies. This can also extend to smart waste management solutions by having these models detect overflowing waste bins, illegal dumping in under-serviced areas or the presence of sanitation workers. YOLOv10 can also be deployed in school zones to monitor the behaviour of traffic as well as crowding for child safety.

CONCLUSION AND FUTURE WORK

In this paper, we proposed a five-layered conceptual framework for YOLOv10, a state-of-art real-time object detection algorithm to be deployed and effectively used in smart city environments with a particular focus on fostering inclusion in urban digital governance. By tying together the need for more scalable, robust, and contextually aware object detection with a greater emphasis on digital inclusion, transparency, and open governance, the proposed framework not only attempts to facilitate more efficient public service delivery but also more socially equitable service distribution. In the future, this framework will be assessed in its ability to provide situational awareness for diverse urban settings and allow for inclusive digital communication in a range of smart city environments with varying levels of resources, development, and infrastructural availability. We would also look to assess potential issues with detection fairness by testing YOLOv10 models against different types of objects across the full range of population groups in urban systems in order to further mitigate instances of inadvertent or algorithmic biases in system design.

DOI: 10.33830/osc.v3i1.7915 / ISSN: 3032-2227

REFERENCES

[1] J. Yang, Y. Kwon and D. Kim, "Regional Smart City Development Focus: The South Korean National Strategic Smart City Program," in *IEEE Access*, vol. 9, pp. 7193-7210, 2021, doi: 10.1109/ACCESS.2020.3047139.

[2] B. Kurniawan, S. Supangkat, F. Lumban Gaol and B. Ranti, "Framework for Developing Smart City Models in Indonesian Cities (Based On Garuda Smart City Framework)," *2022 International Conference on ICT for Smart Society (ICISS)*, Bandung, Indonesia, 2022, pp. 01-05, doi: 10.1109/ICISS55894.2022.9915073.

[3] M. Antonakakis *et al*., "Real-Time Object Detection using an Ultra-High-Resolution Camera on Embedded Systems," *2022 IEEE International Conference on Imaging Systems and Techniques (IST)*, Kaohsiung, Taiwan, 2022, pp. 1-6, doi: 10.1109/IST55454.2022.9827742.

[4] H. Sun, G. Yao, S. Zhu, L. Zhang, H. Xu and J. Kong, "SOD-YOLOv10: Small Object Detection in Remote Sensing Images Based on YOLOv10," in *IEEE Geoscience and Remote Sensing Letters*, vol. 22, pp. 1-5, 2025, Art no. 8000705, doi: 10.1109/LGRS.2025.3534786.

[5] J. Li, X. Zhou, Q. Wang and X. Luo, "Lightweight Real-time Object Detection System Based on Embedded AI Development Kit," *2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*, Guangzhou, China, 2022, pp. 10-14, doi: 10.1109/MLISE57402.2022.00010.

[6] P. Zhang, H. Chen, J. Gao, L. Ma and R. He, "Improved YOLOv10 for High-Precision Road Defect Detection," *2024 4th International Conference on Computer Science and Blockchain (CCSB)*, Shenzhen, China, 2024, pp. 79-83, doi: 10.1109/CCSB63463.2024.10735633.

[7] H. Zhang, "Real-time Pedestrian Detection Algorithm of YOLOv10 under Complex Lighting and Occlusion Conditions," *2024 International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, Dalian, China, 2024, pp. 645-650, doi: 10.1109/AEECA62331.2024.00114.

[8] J. Xu, F. Pan, X. Han, L. Wang, Y. Wang and W. Li, "EdgeTrim-YOLO: Improved Trim YOLO Framework Tailored for Deployment on Edge Devices," *2024 4th International Conference on Computer Communication and Artificial Intelligence (CCAI)*, Xi'an, China, 2024, pp. 113-118, doi: 10.1109/CCAI61966.2024.10602964.

[9] L. Zhou, H. Ran, R. Xiong and R. Tan, "NWD-YOLOv5: A YOLOv5 Model for Small Target Detection Based on NWD Loss," *2024 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI)*, Nanjing, China, 2024, pp. 542-546, doi: 10.1109/RICAI64321.2024.10911422.

[10] X. Yang, C. Zhou, J. Yu, B. Chen and R. Zhang, "A New Strategy for Bird Detection in Coastal Wetlands Based on Improved YOLOv6 3.0," *2023 2nd International Conference on Artificial Intelligence and Intelligent Information Processing (AIIIP)*, Hangzhou, China, 2023, pp. 293-296, doi: 10.1109/AIIIP61647.2023.00062.

[11] H. Zhang, A. Xiong, L. Lai, C. Chen and J. Liang, "AMME-YOLOv7: Improved YOLOv7 Based on Attention Mechanism and Multiscale Expansion for Electric Vehicle Driver and Passenger Helmet Wearing Detection," *2023 IEEE International Conference on Smart Internet of Things (SmartIoT)*, Xining, China, 2023, pp. 223-227, doi: 10.1109/SmartIoT58732.2023.00039.

[12] H. Ling, Z. Tu, G. Li and J. Wang, "ED-YOLOv8s: An Enhanced Approach for Passion Fruit Maturity Detection Based on YOLOv8s," *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, Nanjing, China, 2024, pp. 2320-2324, doi: 10.1109/AINIT61980.2024.10581681.

[13] H. Xin and L. Li, "Arbitrary Style Transfer With Fused Convolutional Block Attention Modules," in *IEEE Access*, vol. 11, pp. 44977-44988, 2023, doi: 10.1109/ACCESS.2023.3273949.

[14] A. Tanji, M. Essaaidi and G. H. Merabet, "Towards an Integrated Smart City Platform: A Prototype for Enhancing Urban Services," *2024 Mediterranean Smart Cities Conference (MSCC)*, Martil - Tetuan, Morocco, 2024, pp. 1-6, doi: 10.1109/MSCC62288.2024.10697044.

[15] N. Mohammadi and J. E. Taylor, "Smart city digital twins," *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI, USA, 2017, pp. 1-5, doi: 10.1109/SSCI.2017.8285439.

[16] F. B. K. ArdaÇ and P. Erdogmus, "Car Object Detection: Comparative Analysis of YOLOv9 and YOLOv10 Models," *2024 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Ankara, Turkiye, 2024, pp. 1-6, doi: 10.1109/ASYU62119.2024.10756955.