Mediterranean Smart Cities Conference (MSCC 2024)
2-4 May, 2024, Martil, Morocco

Integrating UAV-Based Aerial Monitoring and SSD for Enhanced Traffic Management in Smart Cities

Murat Bakirci
Unmanned/Intelligent Systems Lab
Faculty of Aeronautics and Astronautics
Tarsus University
Mersin, Turkiye
muratbakirci@tarsus.edu.tr

Irem Bayraktar
Unmanned/Intelligent Systems Lab
Faculty of Aeronautics and Astronautics
Tarsus University
Mersin, Turkiye
irem_bayraktar@tarsus.edu.tr

Abstract—**In the context of smart cities, transportation stands out as a critical aspect, showcasing innovative solutions to tackle congestion, emissions, and overall mobility challenges. From intelligent traffic management systems to the proliferation of shared mobility services, these cities prioritize efficiency and sustainability. Moreover, they foster the development of electric and autonomous vehicles, enhancing safety and reducing carbon footprints. Intelligent transportation systems play a pivotal role in addressing transportation challenges within smart cities. These systems harness cutting-edge technology to optimize efficiency, safety, and sustainability by enabling dynamic traffic management and promoting multimodal connectivity. Object detection technology further enhances traffic management by providing real-time data for proactive intervention and optimization of traffic patterns. Object detection algorithms revolutionize traffic management by accurately identifying and tracking vehicles in real-time. Additionally, unmanned aerial vehicles (UAVs) serve as powerful tools for aerial traffic monitoring, providing valuable real-time imagery for object detection algorithms. This study focuses on enhancing ITS within smart cities through the integration of UAV-based aerial monitoring and the SSD detection algorithm for vehicle detection. By compiling a large dataset using vehicle images acquired through UAV aerial monitoring, the study demonstrates robust and real-time vehicle detection capabilities. The integration of UAV-based aerial monitoring with SSD vehicle detection holds immense potential for enhancing traffic management and improving overall mobility within smart cities, contributing to safer, more efficient, and sustainable urban environments.**

Keywords—smart city; transportation management; UAV; SSD; vehicle detection; intelligent transportation systems

I. Introduction

In smart cities, transportation emerges as a cornerstone aspect, showcasing innovative solutions to mitigate congestion, reduce emissions, and enhance overall mobility [1]. Leveraging advanced technologies such as Internet of Things (IoT), Artificial Intelligence (AI), and data analytics, smart cities optimize transportation networks, offering seamless integration of various modes of transit [2]. From intelligent traffic management systems that dynamically adjust signals based on real-time traffic flow to the proliferation of shared mobility services like ride-hailing and bike-sharing, these cities prioritize efficiency and sustainability [3]. Additionally, smart cities foster the development of electric and autonomous vehicles, further reducing carbon footprints and enhancing safety. By prioritizing accessibility and connectivity, the transportation infrastructure in smart cities facilitates smoother commutes, fosters economic growth, and cultivates a higher quality of life for residents and visitors alike [4].

Intelligent Transportation Systems (ITS) play a pivotal role in addressing transportation challenges within smart cities by harnessing cutting-edge technology to optimize efficiency, safety, and sustainability [5, 6]. Through real-time data collection and analysis, ITS enables dynamic traffic management, allowing for the timely adjustment of signals and routes to alleviate congestion and reduce travel times. Integrated with smart sensors and cameras, these systems facilitate proactive incident detection and management, enhancing overall safety on roadways [7, 8]. Moreover, ITS facilitates the seamless coordination of various modes of transportation, promoting multimodal connectivity and encouraging the use of sustainable options such as public transit, biking, and walking [9]. By providing travelers with accurate and reliable information through mobile apps and digital displays, ITS empowers them to make informed decisions, reducing reliance on personal vehicles and mitigating environmental impacts.

Object detection technology holds immense potential to revolutionize traffic management within smart cities by enhancing situational awareness and response capabilities [10]. Through the deployment of advanced sensors, cameras, and machine learning algorithms, object detection systems can accurately identify and track vehicles in real-time, providing invaluable data to traffic management centers [11, 12]. By continuously monitoring traffic flow and detecting anomalies such as accidents, breakdowns, or road obstructions, these systems enable prompt intervention and optimization of traffic patterns [13]. Furthermore, object detection can facilitate the implementation of adaptive traffic signal control, dynamically adjusting signal timings based on current traffic conditions to minimize congestion and improve overall efficiency [14]. Additionally, the data collected through object detection can inform long-term planning strategies, enabling authorities to identify recurring congestion hotspots and prioritize infrastructure investments accordingly [15]. Thus, by leveraging object detection technology, smart cities can achieve more proactive, responsive, and effective traffic management, leading to reduced emissions.

Object detection algorithms utilize advanced machine learning techniques such as convolutional neural networks (CNNs) to accurately identify and track vehicles in real-time from video feeds or sensor data [16]. One commonly employed approach is the region-based convolutional neural network (R-CNN), which divides the image into regions of interest and then classifies each region to detect vehicles [17]. Another popular method is the You Only Look Once (YOLO) algorithm, which performs object detection by directly predicting bounding boxes and class probabilities for each detected object in a single pass through the network, enabling faster processing speeds suitable for real-time applications [18]. These algorithms not only enable traffic management systems to detect and track vehicles but also support various other transportation-related tasks such as pedestrian detection, traffic sign recognition, and road condition monitoring.

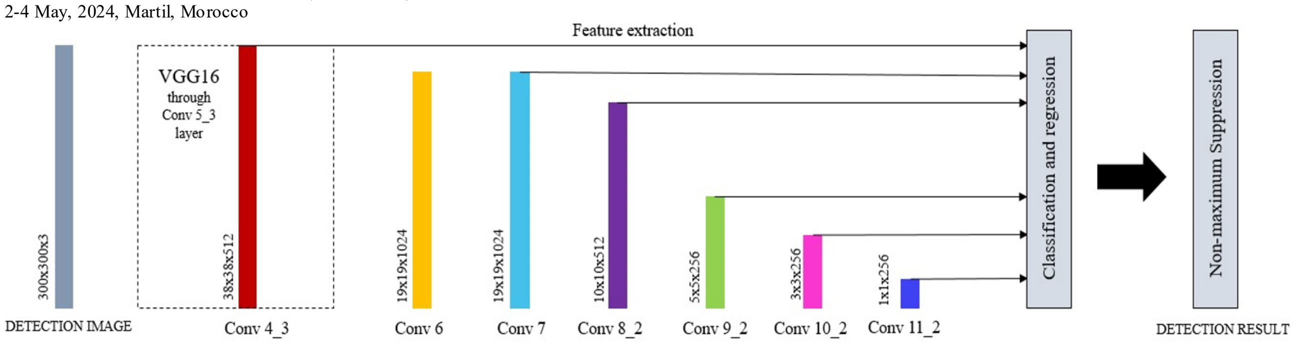
On the other hand, aerial traffic monitoring via UAVs has emerged as a powerful tool in the arsenal of smart cities for transportation management. These systems equipped with high-resolution cameras can provide valuable real-time aerial imagery of traffic conditions, infrastructure, and events [19]. By deploying UAVs for surveillance over roadways, intersections, and public spaces, authorities gain a comprehensive and dynamic perspective on traffic flow, congestion, and potential incidents. Moreover, UAVs serve as an ideal image source for object detection algorithms utilized in transportation applications [20]. With their bird's-eye view and agile maneuverability, they capture detailed imagery of vehicles, pedestrians, and other objects on the ground, facilitating accurate and efficient object detection. Integrating UAV imagery with object detection algorithms enables rapid identification and tracking of vehicles, pedestrians, and other relevant objects, empowering traffic management systems with timely and actionable insights [21, 22]. Through the synergy of UAVs and object detection algorithms, smart cities can enhance traffic monitoring capabilities, optimize transportation networks, and ultimately improve the overall safety and efficiency of urban mobility.

In this study, the focus lies on harnessing cutting-edge technologies to enhance ITS within the context of smart cities, particularly through the utilization of UAVs for aerial monitoring and the implementation of the SSD detection algorithm for vehicle detection. A large dataset was compiled using vehicle images acquired through aerial monitoring conducted by a UAV equipped with a wide-angle camera. Utilizing the UAV system facilitated efficient coverage of expansive areas, thereby enabling the collection of diverse datasets encompassing vehicle images from various perspectives and angles. These vehicle images captured by the UAV serve as the input for the SSD detection algorithm. The study aims to achieve robust and real-time vehicle detection capabilities by inputting vehicle images into SSD, ensuring precise tracking and monitoring of vehicles on highways. It is demonstrated that the integration of UAV-based aerial monitoring with SSD vehicle detection holds immense potential for enhancing traffic management, optimizing transportation networks, and improving overall mobility within smart cities. Thus, this integration contributes to safer, more efficient, and sustainable urban environments.

II. NETWORK ARCHITECTURE

SSD, a single-stage object detection algorithm rooted in deep learning, has garnered significant attention following YOLO. Prior to SSD, both YOLO and Fast R-CNN encountered certain limitations. SSD aims to surpass YOLO in detection accuracy while maintaining a faster pace than Faster R-CNN. To achieve this, SSD was developed by addressing the shortcomings of YOLO and Fast R-CNN. It excels in detecting objects swiftly compared to Fast R-CNN and with greater precision than YOLO, particularly in lower resolution images. SSD embraces the regression concept from YOLO while employing a streamlined end-to-end network for simultaneous bounding box detection and classification. Additionally, SSD incorporates the feature pyramid from Faster R-CNN however omits fully connected layer utilized in Fast R-CNN. This design choice renders SSD more adept at detecting objects than YOLO, with the added benefit of omitting predictions for candidate regions to enhance speed, outperforming Fast R-CNN in this aspect.

The SSD network architecture could be dissected into 3 main components: the backbone network, the original bounding box creation, and the convolution prediction. The backbone network comprises two layers: the basic network layer and the feature extraction layer as shown in Fig. 1. We adopt a modified VGG16 network as the base network, originally trained for image classification on ILSVRC, with its final 2 interconnected layers, substituted with Conv6 and Conv7. This modification aims to preserve feature location information, as fully connected layers may hinder it. Additional convolutional layers, Conv8, Conv9, Conv10, and Conv11, are then appended for further dimension reduction and feature extraction. Utilizing 1x1 and 3x3 convolutional kernels for feature extraction, a multi-scale feature extraction network is constructed by integrating the feature maps from Conv4_3 and Conv7 with those from Conv8_2, Conv9_2, Conv10_2, and Conv11_2. Subsequently, convolution operations are applied to each feature map in the detection network using two 3x3 convolutional kernels. This enhancement allows for the extraction of more comprehen



$$L(x,c,l,g)=\frac{1}{N}[L_{conf}(x,c)+\alpha L_{loc}(x,l,g)] \qquad L_{loc}(x,l,g)=\sum_{i\epsilon Pos}^{N}\sum_{m\epsilon\{cx,cy,\omega,h\}} x_{ij}^{k}\cdot smooth_{L1}(l_i^m-\hat{g}_j^m)$$

Fig. 1. Simplified architecture of SSD.

sive features during object detection. The results of these calculations are amalgamated and passed to the loss layer. The ultimate outcome is generated through the non-maximum suppression (NMS) approach.

The total loss function in the SSD network model comprises a weighted combination of localization errors and confidence errors. This multitask loss function emphasizes both accurate object positioning and reliable confidence estimation. Mathematically, the total loss formula, L, can be represented as shown on Fig. 1. In the expression, N denotes the count of detection boxes aligning with the actual object box, where L_{loc} signifies the location loss, L_{conf} represents the confidence loss, and x denotes the match status of the object box with the detection boxes. If a match occurs between the object box and the detection box, z equals 1; otherwise, it equals 0. The variable c denotes the confidence of the multi-class object, while l provides details regarding the position of the detection box. The coefficient α signifies the weighting factor, establishing the balance between position and confidence loss, typically adjusted to be one via cross-validation. Lastly, g furnishes information about the position of the real object box.

To compute the position loss, we calculate the $smooth_{L1}$ loss between the actual object box and the detection box. This involves regressing the width (w), height (h), and offset of the bounding box's center coordinates (x, y) to minimize positional discrepancies. The position loss equation is also expressed on Fig. 1. The careful execution of the training phase profoundly impacts the success of an object detection algorithm. It necessitates scrupulous preparation of the dataset used during training, which should encompass a rich variety of examples relevant to the target object. Images sourced from remote sensing present unique challenges, diverging from typical natural scene captures. They often exhibit distinct object distributions and viewing perspectives owing to variations in shooting angles, resulting in larger aspect ratios. Object density can vary significantly across different regions within these images, with some areas densely packed and others sparsely populated, sometimes rendering objects minuscule in appearance. Moreover, complex backgrounds prevalent in remote sensing imagery can obscure object clarity, particularly when objects are in close proximity to background elements. To mitigate the algorithm's susceptibility to such challenges, the training dataset should be enriched with diverse instances of the target object across varied scenarios. This includes incorporating images depicting the object from different angles, perspectives, and under diverse lighting conditions throughout the day. By incorporating such comprehensive training data, the algorithm can attain heightened performance levels conducive to real-world applications.

In this study, we conducted stable flights using a manually controlled drone equipped with a high-resolution camera to generate the dataset. The drone, selected for its suitability in aerial surveillance and data collection missions, was programmed to capture numerous videos and photographs from diverse heights and angles across a predefined area. To ensure consistency in image features, we coordinated flights to capture images of specific areas from varying angles and at different times of the day. This deliberate approach enabled us to provide more nuanced assessments of the algorithm's performance. Subsequently, from the collected footage, we extracted between 200 to 250 images from each video at different intervals, resulting in a comprehensive dataset. This dataset comprised 2570 samples for training the algorithm, 530 samples for validation, and 1380 samples for testing algorithm performance. Following training on the training dataset and subsequent validation on the validation dataset, the algorithm underwent rigorous testing using the independent test dataset to evaluate its performance thoroughly.

III. RESULTS

Fig. 2 illustrates 12 sample instances showcasing the detection capabilities of the algorithm, deliberately selected from cases where the algorithm encountered partial success and detection errors to highlight its limitations and deficiencies. In the first image, the algorithm fails to detect a van (pickup truck) and a car obscured by a directional sign on the road, thus only partially visible. Additionally, a fuel tanker is erroneously classified as a bus due to its top view resembling that of a bus. Another car in the upper right corner escapes detection as only a small portion is visible. In the second image, dense shadows obscure two vehicles—one about to enter a roundabout and another already on it. At this juncture, it's crucial to underscore that the dark coloration of the vehicles exacerbates this issue. Moreover, it is observed that two additional black vehicles entering the roundabout from the south, moving northwards, and

Fig. 2. Some examples of vehicle detection, highlighting instances where the algorithm encountered difficulty and made errors.

executing a left turn are successfully detected, but their shadows factored into the identification process. The third image presents a white van misclassified as a car, alongside a vehicle partially obscured by tree branches and nearby motorcycles that go undetected.

The fourth image exemplifies how dense shadows adversely affect detection accuracy. Here, two cars remain undetected due to the shadows cast by trees. Notably, as with the scenario in the second image, these undetectable cars are of a black hue. In shadowy conditions, the dark coloration of vehicles compounds the challenge of detection. This inference becomes even more apparent in the subsequent image. In the fifth image, two distinct white cars occupying identical positions are effortlessly identified, underscoring the significance of vehicle coloration in shadowy environments. Additionally, it's evident that a partially visible car in the upper right corner remains undetected. In the sixth image, while the bus was accurately detected, the vehicle to its right escaped detection and was erroneously included in the bounding box alongside the bus. Furthermore, an approaching car intending to make a right turn eluded detection due to partial obstruction by a signaling pole. Similarly, the black car visible at the bottom left of the image remained undetected due to partial obstruction from both tree branches and the road lighting pole.

In the seventh image, notably, the algorithm erroneously identified an object by the roadside as a car. In the subsequent image, number eight, several detection challenges are evident: one car remains undetected due to partial obstruction by a signaling pole, another is obscured by roadside bushes, and a third in the bottom right corner is only partially visible. Additionally, a white van traveling northward is misclassified as a car, likely due to the lack of discernible details from the image's angle. An unusual instance arises with a black car that eludes detection on the road, situated directly above asphalt traffic warning signs. The algorithm likely struggles due to the multitude of background textures in this scenario. In the ninth image, echoing the preceding scenario, a van in the upper left is misclassified as a car. Furthermore, two blue objects observed by the roadside were erroneously identified as cars.

It is evident in the tenth image that a truck passing under the viaduct, with only part of its body visible, remains undetected. Similarly, akin to the preceding image, a motorcycle traversing directly over the asphalt's traffic warning symbols goes unnoticed. Additionally, an object observed by the roadside was incorrectly identified as a car. The eleventh image highlights the dark car in the upper left remains undetected due to shadowing, while the black car in the lower portion is correctly detected along with its shadow. It's worth noting that the latter was misclassified as a van, likely influenced by the presence of the shadow. In the twelfth and final image, it is noticeable that two of the stationary vehicles went undetected. The car positioned at the right end of the image likely escaped detection due to being partially visible, echoing patterns observed in prior examples. Furthermore, it is anticipated that the van remained undetected due to its exposed rear chassis, compounded by the presence of variously colored and shaped objects within it.

IV. CONCLUSION

This study has focused on leveraging cutting-edge technologies to enhance ITS within the framework of smart cities, with a particular emphasis on integrating UAVs for aerial monitoring and implementing the SSD detection algorithm for vehicle detection. By compiling a comprehensive dataset using vehicle images acquired through UAV aerial monitoring, the study aimed to achieve robust and real-time vehicle detection capabilities using SSD. The integration of UAV-based aerial monitoring with the SSD detection algorithm demonstrated significant potential for enhancing traffic management, optimizing transportation networks, and ultimately improving overall mobility within smart cities. However, the study also highlights certain limitations and challenges encountered in the implementation of the SSD detection algorithm, as evidenced by the sample instances. These instances underscore the algorithm's susceptibility to detection errors, particularly in scenarios involving partial visibility, dense shadows, and complex backgrounds. The detection accuracy is influenced by factors such as vehicle coloration, obstructions, and lighting conditions. Additionally, the algorithm occasionally misclassifies objects, leading to erroneous detections. Despite these challenges, the study provides valuable insights into the potential applications of UAV-based aerial monitoring and SSD for traffic management within smart cities. Future research endeavors should focus on refining the algorithm's performance under challenging conditions, improving its robustness and accuracy, and exploring additional avenues for enhancing ITS capabilities in smart urban environments.

REFERENCES

[1] Reddy, D.S., Babu, K.V.G., Murthy, D.L.N., "Transportation planning aspects of a smart city–case study of GIFT city, Gujarat," *Transp. Res. Proc.*, vol. 17, pp. 134-144, 2016.

[2] Bubelíny, O., Kubina, M., "Impact of the concept Smart City on public transport," *Transp. Res. Proc.*, vol. 55, pp. 1361-1367, 2021.

[3] Zawieska, J., Pieriegud, J., "Smart city as a tool for sustainable mobility and transport decarbonisation," *Transp. Policy*, vol. 63, pp. 39-50, 2018.

[4] Yan, J., Liu, J., Tseng, F.M., "An evaluation system based on the self-organizing system framework of smart cities: A case study of smart transportation systems in China," *Tech. Forecas. Social Change*, vol. 153, no. 119371, 2020.

[5] Lv, Z., Shang, W., "Impacts of intelligent transportation systems on energy conservation and emission reduction of transport systems: A comprehensive review," *Green Tech. Sustain.*, vol. 1, no. 100002, 2023.

[6] Saharan, S., Bawa, S., Kumar, N., "Dynamic pricing techniques for Intelligent Transportation System in smart cities: A systematic review," *Comp. Comm.*, vol. 150, pp. 603-625, 2020.

[7] Salazar-Cabrera, R., de la Cruz, A.P., Molina, J.M.M., "Sustainable transit vehicle tracking service, using intelligent transportation system services and emerging communication technologies: A review," *J. Traff. Transp. Eng.*, vol. 7, no. 6, pp. 729-747, 2020.

[8] Waqar, A., Alshehri, A.H., Alanazi, F., Alotaibi, S., Almujibah, H.R., "Evaluation of challenges to the adoption of intelligent transportation system for urban smart mobility," *Res. Transp. Busin. Manag.*, vol. 51, no. 101060, 2023.

[9] Visan, M., Negrea, S.L., Mone, F., "Towards intelligent public transport systems in Smart Cities; Collaborative decisions to be made," *Proc. Comp. Sci.*, vol. 199, pp. 1221-1228, 2022.

[10] Wang, L., Fan, X., Chen, J., Cheng, J., Tan, J., Ma, X., "3D object detection based on sparse convolution neural network and feature fusion for autonomous driving in smart cities," *Sustain. Citi. Soci.*, vol. 54, no. 102002, 2020.

[11] Rathore, M.M., Paul, A., Rho, S., Khan, M., Vimal, S., Shah, S.A., "Smart traffic control: Identifying driving-violations using fog devices with vehicular cameras in smart cities," *Sust. Citi. Soci.*, vol. 71, no. 102986, 2021.

[12] Chandrakar, R., Raja, R., Miri, R., Sinha, U., Kushwaha, A.K.S., Raja, H., "Enhanced the moving object detection and object tracking for traffic surveillance using RBF-FDLNN and CBF algorithm," *Exp. Syst. Appl.*, vol. 191, no. 116306, 2022.

[13] Wang, K., Chen, C.M., Hossain, M.S., Muhammad, G., Kumar, S., Kumari, S., "Transfer reinforcement learning-based road object

detection in next generation IoT domain," *Comp. Netw.*, vol. 193, no. 108078, 2021.
[14] Gan, H.M., Fernando, S., Molina-Solana, M., "Scalable object detection pipeline for traffic cameras: Application to Tfl JamCams," *Exp. Syst. Appl.*, vol. 182, no. 115154, 2021.
[15] Li, D., Deng, L., Cai, Z., "Intelligent vehicle network system and smart city management based on genetic algorithms and image perception," *Mech. Syst. Sign. Proces.*, vol. 141, no. 106623, 2020.
[16] Swapna, G., Soman, Kp., Vinayakumar, R., "Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals," *Proc. Comp. Sci.*, vol. 132, pp. 1253-1262, 2018.
[17] Li, J., Zhang, D., Zhang, J., Zhang, J., Li, T., et al., "Facial expression recognition with Faster R-CNN," *Proc. Comp. Sci.*, vol. 107, pp. 135-140, 2017.
[18] Shinde, S., Kothari, A., Gupta, V., "YOLO based human action recognition and localization," *Proc. Comp. Sci.*, vol. 133, pp. 831-838, 2018.
[19] Wang, L., Chen, F., Yin, H., "Detecting and tracking vehicles in traffic by unmanned aerial vehicles," *Autom. Constr.*, vol. 72, pp. 294-308, 2016.
[20] Garcia-Aunon, P., Roldán, J.J., Barrientos, A., "Monitoring traffic in future cities with aerial swarms: Developing and optimizing a behavior-based surveillance algorithm," *Cogn. Syst. Res.*, vol. 54, pp. 273-286, 2019.
[21] Bakirci, M., "A novel swarm unmanned aerial vehicle system: Incorporating autonomous flight, real-time object detection, and coordinated intelligence for enhanced performance," *Traitement du Signal*, vol. 40, no. 5, pp. 2063-2078, 2023.
[22] Sutheerakul, C., Kronprasert, N., Kaewmoracharoen, M., Pichayapan, P., "Application of unmanned aerial vehicles to pedestrian traffic monitoring and management for shopping streets," *Transp. Res. Proc.*, vol. 25, pp. 1717-1734, 2017.