

**Authors' response to Reviewers of
"Enhancing Urban Mobility through Traffic Management with UAVs and VLC Technologies"
Javier Garau Guzman, and Victor Monzon Baeza
(Manuscript Drones-2724071)**

REVIEWER 1:

Authors: We sincerely thank the Reviewer for their valuable feedback and insightful suggestions. We truly appreciate your detailed review and are committed to addressing the concerns raised. Here are our responses to each point:

Reviewer: Although the paper is well written with proper organization, I still feel that to step ahead, the authors must improve the following weaknesses.

This work presents enhancing urban mobility through traffic management with UAVs and VLC technologies. whereas:

1. The details of Visible Light Communication (VLC) are missing in the introduction section.

Authors: We have reviewed the introduction, where in the last paragraph VLC technology is mentioned. We highlight in the manuscript the following paragraph:

***Manuscript:** Furthermore, the emergence of Visible Light Communication (VLC) technology has introduced a transformative communication framework. This technology leverages the visible light spectrum, relieving congestion produced in the radio frequency spectrum and opening new avenues for reliable, high-speed data transmission. While both UAVs and VLC technologies individually hold promise, their combined application in urban traffic management remains an emerging frontier with immense potential.*

In this prototype, VLC serves as a novel medium for transmitting signals between UAVs and traffic lights, capitalizing on VLC's advantages, such as high bandwidth, a free spectrum, and more efficient technology. While this introduction touches on VLC, we believe delving into additional technical details at this point is unnecessary, as our focus here is to provide preliminary context about which technologies are considered. On the other hand, the background section examines the current state of the art, including various VLC-related works. Please note that specific details about VLC beyond what is presented here are not included in our work because the objective is not to design a VLC link. Technical details of VLC are not considered in this proposal. However, if the Reviewer has specific inquiries in mind, we invite them to communicate those to us.

2. YOLO processes frames at the rate of 45 fps (larger network) to 150 fps (smaller network), which is better than real-time. In contrast, it has comparatively low recall and more localization error compared to Faster R_CNN. So, what motivates authors this selection in the presence of SSD, Faster R-CNN, RetinaNet, and MobileNet-SSD?

Authors: Thank you very much for this indication. On the one hand, the motivation for selecting the YOLO model over other popular object detection models such as SSD, Faster R-CNN, RetinaNet, and MobileNet-SSD is motivated by specific characteristics and requirements of the proposed system. Here are the reasons for selecting YOLO:

- 1) Real-Time Processing Speed: as Reviewer said, YOLO processes frames at a high speed, ranging from 45 frames per second (fps) for larger networks to 150 fps for smaller networks. This real-time processing capability is crucial for applications like traffic management, where timely detection and response are essential. YOLO's speed advantage ensures swift processing of video frames, enabling rapid decision-making in dynamic traffic scenarios.
- 2) Efficiency in Resource Usage: YOLO is known for its efficiency in resource utilization, making it well-suited for deployment on resource-constrained devices like UAVs. While it may have comparatively lower recall and increased localization error, the trade-off is acceptable in scenarios where real-time processing and efficiency are prioritized.
- 3) Single-Pass Object Detection: YOLO follows a single-pass object detection approach, dividing the input image into a grid and directly predicting bounding boxes and class probabilities. This design contrasts with two-pass methods like Faster R-CNN. The single-pass architecture aligns to minimize processing time and resource usage.
- 4) Simplicity and Ease of Integration: YOLO's simplicity and straightforward architecture make implementing and integrating into the overall system easier. The streamlined design contributes to faster inference and facilitates deployment on UAVs with limited computational capabilities.
- 5) Suitability for UAV Applications: Considering the use case involving Unmanned Aerial Vehicles (UAVs) for traffic management, YOLO's balance between speed and accuracy aligns with the requirements of real-time processing images captured by UAVs.

While other object detection models may excel in certain aspects, the selection of YOLO is a strategic choice based on the specific needs and constraints of the proposed system. The decision is centered on achieving a balance between real-time processing, efficiency, and suitability for UAV deployment in traffic management scenarios.

Moreover, it is crucial to emphasize that in opting for YOLO as the detection algorithm, the intention is not to streamline the process but rather to employ an algorithm that is widely recognized and extensively tested in numerous studies, facilitating its seamless integration into the proposed system. Subsequent efforts in future works can be directed toward identifying the most suitable version or a more optimized algorithm tailored to specific traffic congestion scenarios. The advantages elucidated in the literature regarding YOLO are deemed sufficient for the objectives outlined in this work.

Additionally, we have incorporated a literature-based comparison to substantiate the selection of YOLO in the context of car identification:

The table illustrates YOLO's superior precision and a higher count of true positives compared to alternative models. Although its frames-per-second processing speed falls between that of SSD and other models, we prioritize detection precision for our system. Since SSD sacrifices precision in our specific context of identifying cars, we emphasize detection accuracy in our evaluation. These data are from an example of car detection on the road in [new reference 1]

Model	Average Precision	True Positive	False Positive	Frame per Second (FPR)
YOLO v4	98.08%	273	25	82.1
R_CNN	93.2%	262	17	36.32
SSD	92.7%	257	34	105.14

In the context of [new reference 2], these identical models underwent a comparison for an alternative image detection scenario, specifically involving version 3 of YOLO, where the latter exhibited superior performance compared to other models. This scenario holds significance as it involves substantial losses in the link due to the considerable distance between the satellite and the ground. Such findings are relevant as they can be extrapolated to situations with low visibility caused by factors such as fog, pollution, or other environmental conditions. On the other hand, YOLO has presented a better performance in terms of processing speed than RetinaNet, according to [new reference 3].

New References:

[new reference 1] J. -a. Kim, J. -Y. Sung and S. -h. Park, "Comparison of Faster-RCNN, YOLO, and SSD for Real-Time Vehicle Type Recognition," 2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), Seoul, Korea (South), 2020, pp. 1-4, doi: 10.1109/ICCE-Asia49877.2020.9277040.

[new reference 2] Li, M.; Zhang, Z.; Lei, L.; Wang, X.; Guo, X. Agricultural Greenhouses Detection in High-Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, YOLO v3 and SSD. *Sensors* 2020, 20, 4938. <https://doi.org/10.3390/s20174938>

[new reference 3] Tan, L., Huangfu, T., Wu, L. et al. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC Med Inform Decis Mak* 21, 324 (2021). <https://doi.org/10.1186/s12911-021-01691-8>

3. In 3.2. Traffic congestion or saturation detection: Figure 7 and Figure 9 are not relevant for Traffic congestion. It would be better if the drone-taken images explained parameters in the presence of congestion scenarios.

Authors: Thank you very much for your suggestion. We have refined and replaced the figures (Figure 7 and Figure 9) to better align with the context of traffic congestion detection, providing clear explanations related to congestion scenarios.

4. NACK-free protocol may assist you in less processing.

Authors: We appreciate the suggestion about the N-ACK protocol in order to have less processing. The decision to use an ACK protocol arises from the fact that we consider it important because the traffic lights have to work correctly, and an unwanted mode change does not occur if the information has not arrived correctly or has been affected. In case any unwanted mode change occurs, instead of managing traffic saturation to decrease it, we will help to do the opposite because more vehicles will arrive on the road and cannot get through it.

5. The authors did not discuss the battery constraint regarding processing and flight maintenance.

Authors: We appreciate your observation regarding the battery constraint and flight maintenance. In this paper, our primary focus is on the communication system, signal processing, and protocols implemented in the prototype. This is the integration from the point of view of system blocks to present a possible system that improves Smart Cities. The aspects related to battery constraints and flight maintenance fall outside the scope of this study. Our intention is to provide a detailed exploration of the communication technologies employed in our prototype for traffic management using UAVs and VLC. Regarding the battery aspect, it's noteworthy to mention that the consideration of VLC usage is attributed to its energy-efficient features. However, beyond this point, this study does not consider the system.

However, to enhance the clarity of our contribution, we will explicitly mention in the revised manuscript that certain operational considerations, such as battery constraint and flight maintenance, are beyond the immediate scope of this work. We value your feedback and are committed to refining our manuscript to accurately represent the study's scope and objectives.

6. In figures, the font style is not the same.

Authors: We have checked the manuscript to ensure consistent font styles in all figures for improved readability and presentation.

7. How authors manage traffic using Unmanned Aerial Vehicles (UAVs) in foggy environments.

Authors: Thank you for bringing up this valuable point. We acknowledge your comment regarding the potential impact of foggy environments on the traffic management system using UAVs. In this specific paper, we have focused on scenarios with clear skies, and the impact of foggy conditions is not explicitly addressed.

However, we acknowledge the importance of considering environmental factors like fog in future developments of our system. It is an interesting avenue for future research, and we plan to explore the challenges and adaptations needed for traffic management using UAVs in varying weather conditions, including foggy environments.

In the revised manuscript, we have made it clearer that our current study focuses on clear-sky scenarios, and we will emphasize the potential directions for future research, including considering adverse weather conditions like fog.

8. I could not find the result section. The inclusion of a results section holds significant importance in any research paper. Lacking a results section creates a sense of ambiguity for the reader, leaving them uncertain about the study's conclusions. Authors should recognize the pivotal role of the results section in contributing to the scientific discourse and take care to include this crucial element in their work.

Authors: Thank you for your insightful feedback, and we apologize for the confusion. We appreciate your concern regarding the paper's absence of a dedicated results section. While we acknowledge the

conventional expectation for research papers to include a results section that bolsters the proposal, this is not entirely applicable to our current work. The primary objective of this paper is to present a prototype and propose a novel approach to urban traffic management using mainly UAVs; in the second plane using VLC links. Therefore, our aim is to pave the way for a new avenue of research within the realm of Smart Cities development. We encourage fellow researchers to center their theses on the evolution of this application. To compensate for the absence of a traditional results section, we have incorporated a section with an evaluation detailing the advantages, needs, problems, and potential research directions arising from our work. Through this analysis and evaluation, our primary objective is to establish the groundwork for a research trajectory centered on this application. We hope for your understanding regarding our unique approach and the overarching goal of this work, irrespective of the absence of a dedicated results section.

In the context of a prototype proposal, the results are inherently embedded in the description of the prototype itself. In addition to the evaluation, this paper introduces the conceptual framework or system architecture and integrates technologies to address urban traffic challenges. Since the work is centered around developing and proposing a prototype, a traditional standalone results section, as seen in more mature studies, may not be applicable.

Nevertheless, we understand the importance of clarity in presenting our findings. To address this concern, we ensure that the manuscript explicitly highlights and discusses the outcomes and implications of our prototype throughout the document, providing a clear narrative of the system's performance and capabilities. The actual implementation is reserved for future endeavors, constituting a significant challenge that we are actively addressing. However, achieving this implementation requires additional time. Hence, our current focus is to solidify this proposal as a prototype. This means that the culmination of this paper lies in the tangible form of a prototype—an integrated amalgamation of diverse technologies complemented by the introduction of a novel communication protocol and messaging system, collectively embodying the proposed solution for enhanced urban traffic management.

9. This could be either a review article enabling authors to compare various existing models or a technical article allowing authors to present proposed results in various scenarios with tests and validation through a comparison with existing ones.

Authors: Thank you for your valuable input. While we appreciate the suggestions for either a review article or a technical article, our current direction does not align with these options. As we covered in previous comments, our focus is on a distinct approach that doesn't involve comparing existing models or presenting proposed results in various scenarios. Our specific objective deviates from presenting or proposing a new architecture for urban mobility in Smart Cities with sustainable technologies. If you have any more concerns or considerations regarding our chosen approach also those exposed here, please feel free to share them for further discussion.

On the other hand, based on your previous comments and from other reviewers, to address them, we have included more references in the background (state of the art).

10. There are some grammatical mistakes i.e., interjections and conjunctions.

Authors: Thank you very much for bringing to us this typo. We have carefully proofread the manuscript to address any grammatical mistakes, including interjections and conjunctions.

We appreciate your insights, and these revisions will significantly enhance the clarity and completeness of our work. Thank you for guiding us toward improving the quality of our manuscript.

We trust that the absence of specific technical results does not diminish the significance of introducing this prototype concept, which aims to enhance traffic management and facilitate the progression of Smart Cities.

Sincerely,
The Authors