**Reviewers’ Comments**

Reviewer 1:

In the description of figure 1, it is mentioned that the study intersection has 6 incoming lanes (two each in the north-south orientation, and one each in the east-west orientation). However, the figure does not match with the description.

* **The figure matches the description. The reviewer did not take into account that there are also outgoing lanes. (!)**

In the equation of Queue PLN based reward, the whole queue length function should be divided by the phase length (1/(t-t^p )).

* **This is exactly what is done in the function. (!)**

The authors did not mention about the computational process in the paper. The whole data was collected for 3.5 months from Vivacity vision-based sensors. But the data processing techniques and time for data processing and overall computational complexity is not specified in the manuscript.

* **Calibration of traffic models is beyond the scope of the paper. Done by a different team (company). (!)**

The author has to pay attention on spelling and grammatical mistakes of the manuscript.

* **Reviewed paper, did not find anything else after visual inspection nor after running auto-correct. (!)**

The introduction and literature review section of the paper is not well organized. It does not give a proper understanding of the limitations of the previous work. The nobility of the paper is not well defined as well. Overall, the paper has limited things to contribute.

* **Highlighted the discussion that covers complex state representations and why we do not take it.**
* **Highlighted more the introduction of pedestrians and the few concessions made in terms of simplification of the model topology and underlying systems.**
* **As a secondary effect the section now reads a bit repetitive.**

Reviewer 2:

Some of my major concerns are listed as follows:

1. In recent years, many similar papers that use RL to control intersections are published. What is the main difference between the proposed one and existing models?

* **Already highlighted the pedestrian introduction in the intro.**
* **Added “simultaneous introduction to abstract line 4.**

2. The paper spent a lot of effort discussing the reward functions. Are all of them used in one single RL?

* **Does not deserve an answer. (!)**

3. If I understand correctly, the study used SUMO for model evaluation. SUMO is open-source software and contains many flaws. Has the simulation network been calibrated? If no, the results are not convincing.

* **Calibration is already named in 6 different occasions throughout the paper. (!)**

Reviewer 3:

The paper presents a reinforcement learning approach for managing an urban intersection where both cars and pedestrians are present. The paper is well-written and contributes to the literature on managing multimodal operations at urban intersections explicitly accounting for pedestrian operations and performance. I offer the following comments that can hopefully further improve the paper:

1. The literature review is comprehensive. It would be good to explain the limitations of the research cited ([2]) as using genetic algorithm. Why is the proposed approach better?

* **Unclear it is better, this statement is not made. (!)**

2. In 2nd page top: It is not clear which one is the “second approach” that is taken. Please rephrase to clarify

* **Rephrased to per-lane representations.**

3. The computation time for the different reward functions should be reported. Are some of those calculations more computationally intense than others?

* **Is this worth answering? This is a series (O(10)) floating points operations vs running a whole simulator with O(1000) simultaneous agents. Negligible but explanation hard to fit anywhere. (!)**

4. It is also not clear how frequently the algorithm is called; is it every 12 seconds?

* **Added text in Fig. 2.**

5. Does the notion of a cycle length exist? If so, how is it taken into account? If not, why not? There should at least be maximum green requirements to ensure that no approaches/movements wait for prolonged periods.

* **Not the case. Already explained in 4B. (!)**

6. The explanation of the need for Stage 1 is not clear.

* **Changed text indicating that it is not just defined by transport authority, it is defined as a requirement for this intersection.**

7. The authors might want to consider presenting Fig. 2 using NEMA signal phasing.

* **This is not a US paper. (!)**

8. The paper could also clarify the specifics of the data obtained for both cases and pedestrians and the frequency of those observations. The location of the sensors should also be provided.

* **Indicated that data is logged with the same update frequency as the controller uses in 4C.**
* **This information about the configuration is proprietary. (!)**

9. I would also recommend describing the time interval over which the reward functions are calculated.

* **Already explained in the Algorithm**
* **Added extra statement on 4E.**

10. Are the queue reward functions calculated for only cars?

* **Not worth answering. (!)**

11. The paper could benefit from some discussion on the selection of the rewards and the motivation behind them as well as their limitations as identified through previous research and this work.

* **Simply not enough space, this should be addressed in a journal review, not a conference paper. Some of this is done on reward descriptions. (!)**

12. Were the extreme values for the prioritization weights tested? If so, what were the results/conclusions?

* **Comment not well thought. Extremes mean only a single reward (veh or ped) is considered. Would not work, did not tests. (!)**

13. The delay calculations for pedestrians should be

explained in more detail.

* **Added detail in 5C-1.**

14. The authors claim that the rewards that do not require prioritization are more practical due to the fact that they do not need calibration; yet a maximum waiting time, maximum desirable capacity, and maximum speeds are required for some of the other ones. How are those

calculated? Do they also need to be calibrated for different sites?

* First question: **Indicated they are heuristically set.**
* Second question: **No need for answer. (!)**

15. In the tests performed, are all approaches experiencing the same demand?

* **This is thoroughly explained in 6A and 6B. (!)**

16. It’s not clear how the comparison against the

reference systems comes into play in the presented results.

Are all of the reward functions presented in Fig. 3 performing better than the reference ones? A comparison with the reference systems should be presented.

* **This is already presented in 7, paragraph 2. (!)**

17. Fig. 3 abbreviations should be explained, e.g., VA.

* **Added abbreviations in 6A.**

18. What are the “vanilla reference” methods?

* **Removed “vanilla”. Methods are already specified in the following section. (!)**

19. Can the green time extension be changed to be shorter? How would that affect the results?

* **Extensions are already 0.6s. Indicated this is in line with the sensor sampling frequency. Unclear if theoretical maximum, as it depends on the sensors themselves which are out of the scope of the paper. (!)**

20. Why aren’t the wait time vs delay reward functions have the same/similar results?

* **Because they are different functions encoding different information. (!)**

21. Many results are presented, and it is hard to decipher what is important/what can guide future implementations of RL at other sites.

* **Not a question. (!)**

22. Would the problem related to the instantaneous snapshot be resolved if average queue were to be included in the reward function?

* **No. (!)**