

Estimating vehicular emissions by applying deep learning on video camera scenes

Corresponding Author: Dr Paolo Santi

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

This paper presents an interesting and timely approach that combines computer vision with an emission modeling framework to estimate vehicular pollutants from moving cameras. The use of a large-scale dataset (DRIVE) and the case study in Amsterdam make the paper highly relevant to both the research community and urban policy stakeholders. The writing is clear, the figures are informative, and the interdisciplinary perspective is commendable. I believe this work has strong potential, and with a few refinements it could become even more impactful. Below are some suggestions that the authors can readily address in a revision.

- 1、 Add a short paragraph in the Introduction explicitly stating what is new compared with previous camera-based emission studies (e.g., larger dataset, improved speed estimation, case study application).
- 2、 You should include a few more recent references on vehicle emission monitoring using AI and remote sensing, to situate this work more firmly in the literature.
- 3、 Some figures (e.g., emission maps, dataset samples) could benefit from more detailed captions so that they can be understood independently of the main text.
- 4、 You should briefly explain how the DRIVE dataset was curated, especially how non-relevant images (e.g., interiors) were filtered, to increase reader confidence.
- 5、 Provide approximate runtime for detection, classification, and emission estimation, to demonstrate feasibility for near-real-time applications.
- 6、 You need to study how the method could be transferred to other cities with different vehicle fleets and road conditions.

Reviewer #2

(Remarks to the Author)

Thank you for the interesting and innovative paper. It is a valuable topic, and the paper is written clearly.

The major points are would raise are:

- * It would be usual to have a literature review at the start of the paper, but I don't see this.
- * Equally, there is are no Conclusions/Implications, Limitations, and Future Work sections, which would be customary.
- * The results would benefit from some method of validating the accuracy of the emissions estimates. Ideally, this would performing some real-world testing. As an alternative, could some existing data or forecasts from other models be used to test on a subset of your estimates?
- * Not being able to differentiate the fuel, vehicle age and Euro stage is the single biggest limitation of your method, which limits the general accuracy of forecasts. For example, can it differentiate between a ten-year-old Ford Focus diesel from a three-year old gasoline For Focus? The emissions profiles of these two vehicles are completely different.

More minor points are:

- * Please give an exact definition of "top-1" and "top-5" in the main paper.
- * Figure 3 - please summarise how exactly the three observed routes are combined for these plots.

- * Figure 6 - please clarify which route(s) this chart relates to.
- * Page 8/25 - worth commenting that assuming constant vehicle speed over each video frame probably creates a downwards bias in our forecasts.
- * Equation 9 - I think this is distance-specific emissions - please clarify.

Reviewer #3

(Remarks to the Author)

Urban traffic emission modeling is of crucial importance in our days. Accordingly, this study, utilizing computer vision to estimate vehicular emissions from moving camera footage, offers data in urban pollution monitoring. The results are based on real-world data set making it a credible research. However, validation is fully missing. In all, the paper is technically sound but needs clarification in limitation and some further explanation is needed. See the comments below.

1) The paper mentions the scalability of the proposed method. That could be a really important feature. However, I cannot see this. It is unclear how the city-wide emission monitoring would be feasible exactly. Now, the method estimates emission along dedicated routes using moving sensor technique. There are plenty of papers about network wide emission (using vehicle level emission models). I think you should have a look on them first, e.g.,
<https://dx.doi.org/https://doi.org/10.1016/j.trip.2025.101513>
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0305481>
<https://iopscience.iop.org/article/10.1088/1748-9326/ac8b21>
<https://ieeexplore.ieee.org/document/10819051>
 etc.

So, do you have any idea for scalability?

2) You do validation for the car type recognition only, but validation for the emission modeling is omitted in the research. I miss PEMs based investigation (<https://ieeexplore.ieee.org/document/10819857>) or any other real car level emission measurement data for validation of the emission. Or at least the estimated punctuality of proposed technique (end-to-end modeling validation) should be provided in some way. But if you want to fully avoid it, you should declare as limitation/assumption that you only rely on an existing (COPERT model (which itself has some ambiguity)).

3) What about the electric and hybrid vehicles? Do you have any idea, how to filter them out from the footage data?

4) Start-stop systems are very common in vehicles. Could you consider this in the modeling or in future work? (I see that street parking is also considered). Are you able to detect cars in the queue at stop lines?

5) Figure 6: there is a straight part (regressed?) of the data at around time "2415:30". What is the reason for this? Additionally, the time scale labeling is unclear for me. Is it hour:min? Should be corrected to be precise.

6) Subfigs in Fig. 7 are very small, hard to see and understand. The more, when superzooming in, they look the very same, having no added value separately.. You should present in an another way or add levels to the pics (e.g. color codes of normalized emission value).

7) In case of multilane roads, the vehicles may cover each other. Is it a problem for your method? Can you estimate the error due to this covering?

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

I have no more comment.

Reviewer #2

(Remarks to the Author)

Thanks very much for the responses to my comments, which in the main are satisfactory and improve the papers.

On the use of the COPERT model, I realise it is a standard model, but it is also one about which there are well-know reservations to use for this purposes, and especially for pollutants that are significantly acceleration dependent (CO, NOx, PM). I would recommend that this is very clearly stated as a known limitation.

The only significant remaining issue is the inability of the system to differentiate between fuel type. The orders-of-magnitude differences in CO, NOX, PM and VOCs between gasoline and diesel could render your predictions almost meaningless. Is

there any of information from you can gather from your system to resolve this issue?

Reviewer #3

(Remarks to the Author)

Most of my comments were neglected and avoided being fully addressed (especially points 2, 3, 4, and 7).

The main point is the lack of novelty of this research in this form. Again, please note that a plethora of papers exist for network-wide traffic emission modeling based on Copert, HBEFA, etc.

The simplest way is to collect loop data or any other sensor, and directly put them into traffic simulators equipped with fully validated emission models (e.g., VISUM, VISSIM, SUMO).

The authors simply avoided showing the real scientific contribution of their paper.

Also, the authors avoided considering numerous relevant papers from transportation journals (I also suggested a few ones, but you can google much more).

Considering the fact that this journal is an upcoming series and the authors did a lot of work (even if it is not fully proper in this field), I suggest acceptance.

Sorry to be frank with you, but I am directly working in this field, and so I am probably more critical than the other reviewers. (Still, if you try to better address my comments, I would be happy.)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

Thank you for the further revisions to the manuscript. The only reminaing issue is the consequence of not being able to identify the fuel of the vehicle. I accept that such identification is difficult using the camera.

Although I am not an expert in COPERT, I understand it provides emissions factors split by fuel type. Your solution - I think - is to create some average factor across the different fuel types, weighted by the mix of fuel tyres in the vehicle population in question. This would resolve the accuracy of your forecasts *at the fleet level* but not at the vehicle level. This would immediately limit the value of the approach, if the emissions forecast for an individual vehicle were subject to large uncertainties.

At the fleet level, my concern is that, if you already have the fleet fuel mix, you don't need the cameras at all - you would just weight COPERT values by this mix you already know. In this scenario, your model doesn't add any value.

Put another way, I can see that you are acknowledging the limitations of your system transparently, but nevertheless, if you don't know the fuel, How can you ever forecast the emissions of an individual vehicle with any accuracy?

I apologise if I am missing something fundamental, but I cannot see how it is possible to compensate for not knowing the fuel tyre by statistical techniques.

Reviewer #3

(Remarks to the Author)

Authors properly revised the paper.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response Letter

Dear editors,

We thank the reviewers for their thoughtful and constructive comments, which have helped us improve the clarity and focus of our manuscript. We appreciated the comments on the scalability of the method; we addressed this by conducting another experiment in Abu Dhabi. We also carefully considered each suggestion and revised the paper accordingly to better articulate our motivation, strengthen our methodological explanations, and clarify the interpretation of our results. Below, we provide detailed responses to each comment.

Best regards,
All authors

Reviewer #1

This paper presents an interesting and timely approach that combines computer vision with an emission modeling framework to estimate vehicular pollutants from moving cameras. The use of a large-scale dataset (DRIVE) and the case study in Amsterdam make the paper highly relevant to both the research community and urban policy stakeholders. The writing is clear, the figures are informative, and the interdisciplinary perspective is commendable. I believe this work has strong potential, and with a few refinements, it could become even more impactful. Below are some suggestions that the authors can readily address in a revision.

- 1、 Add a short paragraph in the Introduction explicitly stating what is new compared with previous camera-based emission studies (e.g., larger dataset, improved speed estimation, case study application).
- 2、 You should include a few more recent references on vehicle emission monitoring using AI and remote sensing, to situate this work more firmly in the literature.

Response

Thank you for your comments. In the revised manuscript, we have extended the Introduction and added a short paragraph reporting references to recent work on emission monitoring using AI and remote sensing, and more clearly stating the novelty of our approach vs. existing literature.

3、 Some figures (e.g., emission maps, dataset samples) could benefit from more detailed captions so that they can be understood independently of the main text.

Response

Thank you for your comments. We have added more details to the captions in the paper

4、 You should briefly explain how the DRIVE dataset was curated, especially how non-relevant images (e.g., interiors) were filtered, to increase reader confidence.

Response

In the revised manuscript, we have added a paragraph in Section XX mentioning how the DRIVE data set was curated. More specifically, DRIVE trains a make/model classifier on detected vehicle crops; in both our test set and real-world use, the pipeline first detects a vehicle and then classifies the crop. Therefore, interior-only images never reach the classifier in real-world test case scenarios and do not introduce label noise. While we can remove interiors by passing all images through a pre-trained vehicle detector (which would not detect interiors as vehicles), our results would not be affected by an extra filter.

5、 Provide approximate runtime for detection, classification, and emission estimation, to demonstrate feasibility for near-real-time applications.

Response:

We have added a paragraph in Section xx of the revised manuscript reporting computational times of the proposed method. In particular, on an NVIDIA H100 processing 1080p video, detection averages 15.1 ms per frame, with a 90th-percentile latency of 15.6 ms. Classification averages 36.4 ms per frame, with a 90th-percentile latency of 78.2 ms, while emission estimation averages 2.6 ms per frame and, on frames where it is invoked, takes 42 ms with a 90th-percentile latency of 43 ms. The end-to-end latency is 73.4 ms per frame (≈ 13.6 FPS), demonstrating near-real-time feasibility.

6、 You need to study how the method could be transferred to other cities with different vehicle fleets and road conditions.

Response

Thank you for the helpful comment. To demonstrate the model's transferability, we conducted an additional driving experiment in Abu Dhabi, a city characterized by distinct vehicle fleets and road conditions, and repeated the same analysis. The new results have been added to the manuscript and discussed in both the main text and the appendix (see Figure 22). Overall, the model continues to perform well, capturing both the aggregate and fine-grained patterns of driving behavior and emissions. In addition, we are conducting further large-scale experiments in New York City (reported in a separate paper: <https://www.researchsquare.com/article/rs-7030366/v1>), which will provide further evidence on the scalability and generalizability of the approach.

Reviewer #2

Thank you for the interesting and innovative paper. It is a valuable topic, and the paper is written clearly.

The major points I would raise are:

1. It would be usual to have a literature review at the start of the paper, but I don't see this.
2. Equally, there are no Conclusions/Implications, Limitations, and Future Work sections, which would be customary.

Response:

Thank you for your comment. We have followed the journal's template (Nature-affiliated) in which there is no specific section for literature review, which is included in the *Introduction* section. Similarly, the discussion of conclusions/implications, limitations, and future work is contained in the Discussion section at the end of the manuscript. While not changing the overall manuscript structure to remain aligned with Nature-style formatting guidelines, we have worked on the content of the mentioned sections to better discuss related literature and novel contributions of our work, in the Introduction, and limitations/conclusions/future work in the Discussion section.

* The results would benefit from some method of validating the accuracy of the emissions estimates. Ideally, this would performing some real-world testing. As an alternative, could some existing data or forecasts from other models be used to test on a subset of your estimates?

Response:

Thank you for your comment. We agree that direct emission measurements would provide the strongest validation. However, the ultimate motivation of this work is to use point-wise, road-level emission estimations to provide city-scale estimations, which has led us to use a macroscopic emission model in the framework; namely, the widely known COPERT model. Within this framework, whose macro-level accuracy has been extensively validated in the literature, vehicle speed and acceleration are the primary operational inputs determining emission factors. Our validation, therefore, focused on ensuring the accuracy of speed estimates, which directly govern the reliability of modeled outputs. We have revised the paper to better discuss the motivation of our work and the importance of assessing the accuracy of speed/acceleration estimations.

* Not being able to differentiate the fuel, vehicle age and Euro stage is the single biggest limitation of your method, which limits the general accuracy of forecasts. For example, can it differentiate between a ten-year-old Ford Focus diesel from a three-year old gasoline For Focus? The emissions profiles of these two vehicles are completely different.

Response:

We fully acknowledge this limitation. Visual data alone cannot distinguish engine variants (e.g., diesel vs. gasoline) or Euro standards within visually identical model generations. To address this, we apply average emission factors derived from the COPERT model, which reflect standard EU fleet compositions. While this introduces uncertainty for individual vehicles compared to registration-based lookups, the primary contribution of our method is capturing high-resolution dynamic behavior (speed and acceleration profiles) at scale data which static fleet inventories lack.

More minor points are:

* Please give an exact defintion of "top-1" and "top-5" in the main paper.

Response:

We have added explicit definitions for both metrics in the "Vehicle make and model classification" section to ensure clarity.

Top-1 accuracy refers to the percentage of samples where the model's highest-probability prediction matches the ground truth label. Top-5 accuracy refers to the percentage of samples where the correct label appears within the model's five highest-probability predictions.

* Figure 3 - please summarise how exactly the three observed routes are combined for these plots.

Response:

We have updated the caption for Figure 3 to clarify that the data represent an aggregation of detections from all three observer routes. We combined these datasets to ensure a sufficient sample size across the entire speed range (0–140 km/h) and to capture diverse driving behaviors. A single route would not provide enough data points at specific speed intervals to generate a smooth, robust curve; therefore, aggregation was necessary to provide a more accurate estimate of the emission factors.

* Figure 6 - please clarify which route(s) this chart relates to.

Response:

We have updated the caption for Figure 6 (Emission Maps) to explicitly state that the visualization combines the spatial coverage of all three observer vehicles.

* Page 8/25 - worth commenting that assuming constant vehicle speed over each video frame probably creates a downwards bias in our forecasts.

Response:

We agree with the reviewer that this assumption likely leads to conservative estimates. We have added a statement addressing this specific bias in the "Discussion" section.

* Equation 9 - I think this is distance-specific emissions - please clarify.

Response:

We have clarified in the text following Equation 9 that the calculated emissions (E_m^{CALC}) represent distance-specific emissions measured in g/km.

Reviewer #3

Urban traffic emission modeling is of crucial importance in our days. Accordingly, this study, utilizing computer vision to estimate vehicular emissions from moving camera footage, offers data in urban pollution monitoring. The results are based on real-world data set making it a credible research. However, validation is fully missing. In all, the

paper is technically sound but needs clarification in limitation and some further explanation is needed. See the comments below.

1) The paper mentions the scalability of the proposed method. That could be a really important feature. However, I cannot see this. It is unclear how the city-wide emission monitoring would be feasible exactly. Now, the method estimates emission along dedicated routes using moving sensor technique. There are plenty of papers about network wide emission (using vehicle level emission models). I think you should have a look on them first, e.g., <https://dx.doi.org/https://doi.org/10.1016/j.trip.2025.101513> <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0305481> <https://iopscience.iop.org/article/10.1088/1748-9326/ac8b21> <https://ieeexplore.ieee.org/document/10819051> etc. So, do you have any idea for scalability?

Response

Thank you for the helpful comment. To demonstrate the model's scalability, we conducted an additional driving experiment in Abu Dhabi, a city characterized by distinct vehicle fleets and road conditions, and repeated the same analysis. The new results have been added to the manuscript and discussed in both the main text and the appendix (see Figure 22). Overall, the model continues to perform well, capturing both the aggregate and fine-grained patterns of driving behavior and emissions.

2) You do validation for the car type recognition only, but validation for the emission modeling is omitted in the research. I miss PEMs based investigation (<https://ieeexplore.ieee.org/document/10819857>) or any other real car level emission measurement data for validation of the emission. Or at least the estimated punctuality of proposed technique (end-to-end modeling validation) should be provided in some way. But if you want to fully avoid it, you should declare as limitation/assumption that you only rely on an existing (COPERT model (which itself has some ambiguity)).

Response:

Thank you for your comment. We agree that direct emission measurements would offer the strongest validation, but our goal is to scale point-wise, road-level estimates to city-level emissions using the well-established COPERT macroscopic model. Since this framework primarily depends on vehicle speed and acceleration, our validation focuses on ensuring the accuracy of these estimates, which determine the reliability of the modeled emissions; we have clarified this motivation and focus in the revised paper.

3) What about the electric and hybrid vehicles? Do you have any idea, how to filter them out from the footage data?

Response:

We thank the reviewer for pointing this out. The inclusion of electric and hybrid vehicles in the emission estimates is a limitation. We have specified this limitation in the discussion, while also recognizing that the DRIVE dataset presented in the paper brings us closer to classifying electric and hybrid vehicles - future work could extend the DRIVE dataset to include more varied vehicle makes and models, and evaluate whether this solves the issue of hybrid and electric vehicles.

4) Start-stop systems are very common in vehicles. Could you consider this in the modeling or in future work? (I see that street parking is also considered). Are you able to detect cars in the queue at stop lines?

Response:

In the manuscript, we changed the title of the “Detection of parked vehicles” to “Detection of parked and stopped vehicles”. We clarified that vehicles estimated to be stationary are not considered in the COPERT emissions estimate. Consequently, this effectively simulates a scenario where all stopped vehicles have Start-Stop systems active. This comprises both parked and temporarily stopped vehicles.

5) Figure 6: There is a straight part (regressed?) of the data at around time "2415:30". What is the reason for this? Additionally, the time scale labeling is unclear for me. Is it hour:min? Should be corrected to be precise.

Response:

Thank you for noticing this. We updated the figure using better normalization. As a result of this, the gaps are now separated.

6) Subfigs in Fig. 7 are very small, hard to see and understand. The more, when superzooming in, they look the very same, having no added value separately.. You should present in an another way or add levels to the pics (e.g. color codes of normalized emission value).

Response:

Thank you for your comment. We have updated Figure 7 as suggested.

7) In case of multilane roads, the vehicles may cover each other. Is it a problem for your method? Can you estimate the error due to this covering?

Response:

This is a relevant and practical question. There are several works on the re-identification of vehicles after a brief period of occlusion, many of which we cited. Our tracking algorithm can deal with short-term occlusions. However, for vehicles that are hidden for longer, we acknowledge the limitation. With the inclusion of Abu Dhabi as a case study in the revised manuscript, we were able to showcase how the method works, as many of the roads traversed are multi-lane.

Reviewer #1

I have no more comment.

Response: Thank you for your review and for your helpful feedback in the earlier round. We appreciate your positive assessment of the revised manuscript.

Reviewer #2

Thanks very much for the responses to my comments, which in the main are satisfactory and improve the papers.

Response: We thank the reviewer for the constructive feedback and are pleased that the revisions were largely satisfactory.

On the use of the COPERT model, I realise it is a standard model, but it is also one about which there are well-known reservations to use for this purposes, and especially for pollutants that are significantly acceleration dependent (CO, NOx, PM). I would recommend that this is very clearly stated as a known limitation.

We agree and have revised the manuscript to state this limitation explicitly and prominently. While COPERT is a standardized and widely used inventory framework, it is not a direct tailpipe measurement model and its accuracy depends on calibration assumptions and how well transient/engine-load effects are represented. This is particularly important for pollutants that are strongly acceleration- and transient-operation dependent (e.g., CO, NOx, PM, VOC). We now clearly state that our pollutant estimates inherit model-form uncertainty from COPERT, and that this limitation is most relevant for these acceleration-sensitive pollutants. We also clarify that we partially address uncertainty by validating video-derived speed against LiDAR and propagating kinematic uncertainty into emissions via Monte Carlo, while full end-to-end validation against on-road measurement campaigns (e.g., PEMS/remote sensing) remains future work.

The only significant remaining issue is the inability of the system to differentiate between fuel type. The orders-of-magnitude differences in CO, NOX, PM and VOCs between gasoline and diesel could render your predictions almost meaningless. Is there any of information from you can gather from your system to resolve this issue?

We fully agree that fuel type is a major driver of pollutant emissions (especially NOx, PM, CO, VOC), and we have strengthened the manuscript to make this a central limitation and to clarify the interpretation of our outputs. In general, fuel type and detailed powertrain variants cannot be reliably inferred from RGB video alone because

visually similar trims can correspond to different engines and after-treatment systems across years and markets. Therefore, we now explicitly state that our pollutant-specific estimates should be interpreted as conditional on the fleet-average assumptions embedded in the COPERT Tier 3 parameterization used in this study.

We also describe practical mitigation pathways for deployments: (i) incorporating external fleet priors (e.g., region- and road-class-specific fuel-type distributions from registrations or fleet statistics) to compute expected emissions under a probabilistic fuel-type mixture and report uncertainty bounds, and/or (ii) fusing with registration-derived attributes where legally and ethically permissible. We clarify that without such external priors or attribute fusion, pollutant estimates represent fleet-average COPERT outputs rather than trim- or fuel-specific emissions for each detected vehicle, and we present this explicitly as a limitation and an opportunity for future work.

Reviewer #3

Most of my comments were neglected and avoided being fully addressed (especially points 2, 3, 4, and 7). The main point is the lack of novelty of this research in this form. Again, please note that a plethora of papers exist for network-wide traffic emission modeling based on Copert, HBEFA, etc. The simplest way is to collect loop data or any other sensor, and directly put them into traffic simulators equipped with fully validated emission models (e.g., VISUM, VISSIM, SUMO). The authors simply avoided showing the real scientific contribution of their paper. Also, the authors avoided considering numerous relevant papers from transportation journals (I also suggested a few ones, but you can google much more). Considering the fact that this journal is an upcoming series and the authors did a lot of work (even if it is not fully proper in this field), I suggest acceptance. Sorry to be frank with you, but I am directly working in this field, and so I am probably more critical than the other reviewers. (Still, if you try to better address my comments, I would be happy.)

We appreciate the reviewer's candid assessment and agree that network-wide emission modeling using established frameworks (e.g., COPERT/HBEFA) and traffic simulation pipelines (e.g., VISUM/VISSIM/SUMO with loop or fixed sensors) is a mature and important area.

Our intent is not to reclaim novelty in network-wide emission modeling itself, nor to replace these established approaches. Instead, we position our contribution as complementary: a vehicle-resolved, mobile sensing pipeline that derives (i) fleet composition at high granularity (make/model, via large-scale visual classification) and (ii) kinematics (speed/acceleration) directly from moving-camera video, and aggregates

these to link-level, time-stamped emission estimates that can support and enrich network-wide modeling—a contribution that has been noticed by the other reviewers.

In the revised manuscript, we have strengthened the framing of the scientific contribution in three ways. First, we clarify that our main novelty is the vehicle-level inference of fleet mix and kinematics from ubiquitous video, rather than the emission-factor model itself. Second, we explain how these probe-derived link-level estimates can be integrated with conventional network-wide frameworks, providing higher-resolution, observation-based inputs (fleet composition and speed–acceleration distributions) that are often assumed or only coarsely measured in simulation-based workflows. Third, we expanded the contextualization and citations to better reflect the transportation emissions literature and to explicitly situate our work relative to prior network-wide approaches, emphasizing complementarity and avoiding over-claiming.

Finally, we thank the reviewer for the recommendation of acceptance and for pushing us to more clearly articulate the contribution and scope of the paper.

Thank you for the further revisions to the manuscript. The only remaining issue is the consequence of not being able to identify the fuel of the vehicle. I accept that such identification is difficult using the camera.

Although I am not an expert in COPERT, I understand it provides emissions factors split by fuel type. Your solution - I think - is to create some average factor across the different fuel types, weighted by the mix of fuel types in the vehicle population in question. This would resolve the accuracy of your forecasts *at the fleet level* but not at the vehicle level. This would immediately limit the value of the approach, if the emissions forecast for an individual vehicle were subject to large uncertainties.

At the fleet level, my concern is that, if you already have the fleet fuel mix, you don't need the cameras at all - you would just weight COPERT values by this mix you already know. In this scenario, your model doesn't add any value.

Put another way, I can see that you are acknowledging the limitations of your system transparently, but nevertheless, if you don't know the fuel, How can you ever forecast the emissions of an individual vehicle with any accuracy?

I apologise if I am missing something fundamental, but I cannot see how it is possible to compensate for not knowing the fuel type by statistical techniques.

Response:

We thank the reviewer for this important comment. We agree that fuel type is an important determinant in COPERT and that RGB imagery alone does not always allow deterministic identification of the exact fuel or emissions-control subclass of an individual vehicle. To address this concern, we have revised the manuscript to clarify the intended scope of the framework and to soften statements that could be interpreted as claiming certification-grade accuracy for a single passing vehicle.

Specifically, we now clarify that our method should not be interpreted as providing exact homologation-grade emissions for an individual vehicle when fuel type is not directly observable. Rather, the framework provides vehicle-resolved, model-based emission estimates whose main analytical value emerges after aggregation across vehicles, road links, and time intervals.

We also clarify why the moving-camera approach remains valuable even when fleet fuel shares are known. Fleet fuel composition alone cannot identify which vehicles are present on a given road segment at a given time, nor can it recover speed, acceleration, congestion episodes, or stop-and-go dynamics. These activity variables are essential inputs for high-resolution urban emissions estimation within the COPERT framework. Our system contributes precisely this missing spatiotemporal activity information, enabling link-level and time-resolved emissions mapping rather than relying only on static fleet averages.

In response to the reviewer's comment, we therefore revised the abstract, introduction, methods, discussion, and conclusion to make this scope explicit, soften the wording around single-vehicle accuracy, and emphasize the intended use of the method for high-resolution urban emissions monitoring.