Reviewer 2:

Thank you for reviewing our manuscript! Here are our responses:

**Feedback 1:** Page 2, L41-42: The authors should add a citation indicating what studies are available.

**Author’s Response:** *Thank you for your feedback!*

*We have added these two citations related to machine learning based odor source localization in lines 41-42 of Section 1.*

*Kim, H., Park, M., Kim, C. W., & Shin, D. (2019). Source localization for hazardous material release in an outdoor chemical plant via a combination of LSTM-RNN and CFD simulation. Computers & Chemical Engineering, 125, 476-489.*

*Hu, H., Song, S., & Chen, C. P. (2019). Plume tracing via model-free reinforcement learning method. IEEE transactions on neural networks and learning systems, 30(8), 2515-2527.*

**Feedback 2:** Page 2, L43-57: This paragraph claims that organisms generally use vision and smell, but it is known that animals other than humans have poor eyesight. Hence, animals should use other modalities such as wind, tactile, heat, and so on. The authors should make a convincing argument as to why you should use vision among the multiple candidate modalities. The current paragraph is not fair.

**Author’s Response:** *Thank you for your feedback!*

*To find an unknown odor source location, the sense of smell, i.e., olfaction, is the primary modality for animals to rely on. Wind sensing is included in the animal olfactory behaviors, such as the mate-seeking behaviors of male moths, where a male moth flies against the wind direction to approach the odor source location when it senses high odor concentrations. Somatosensory sensing (i.e., tactile sensing), and thermal sensing are not widely used in the task of finding odor source location.*

*Vision, on the other hand, is important in the task of finding the odor source location. Usually, vision is working with olfaction for animals to pinpoint the exact odor source location. In the event of sensing high odor concentration, animals will first use vision to search for possible odor source targets, and if such targets are not directly visible, animals will rely on olfaction to trace odor plumes and approach the odor source. The final odor source confirmation is usually relying on vision. Thus, both vision and olfaction are vital modalities in the task of odor source localization.*

*According to your feedback, we have updated lines 49-61 of Section 1.*

**Feedback 3:** Page 3, L124-126: Recently, a 3D algorithm in turbulent environments have been demonstrated with a palm-sized drone.

Shigaki, S., Yoshimura, Y., Kurabayashi, D., & Hosoda, K. (2022). Palm-Sized Quadcopter for Three-Dimensional Chemical Plume Tracking. IEEE Transactions on Instrumentation and Measurement, 71, 1-12.

**Author’s Response:** *Thank you for your feedback!*

*We have added the citation in line 134 of section 2.*

**Feedback 4:** Page 3-4, L127-136: It has been reported that OSL can be efficiently achieved by switching between the Infotaxis and Dijkstra algorithms in the obstacle domain depending on the situation, and I recommend that you should cite its research.

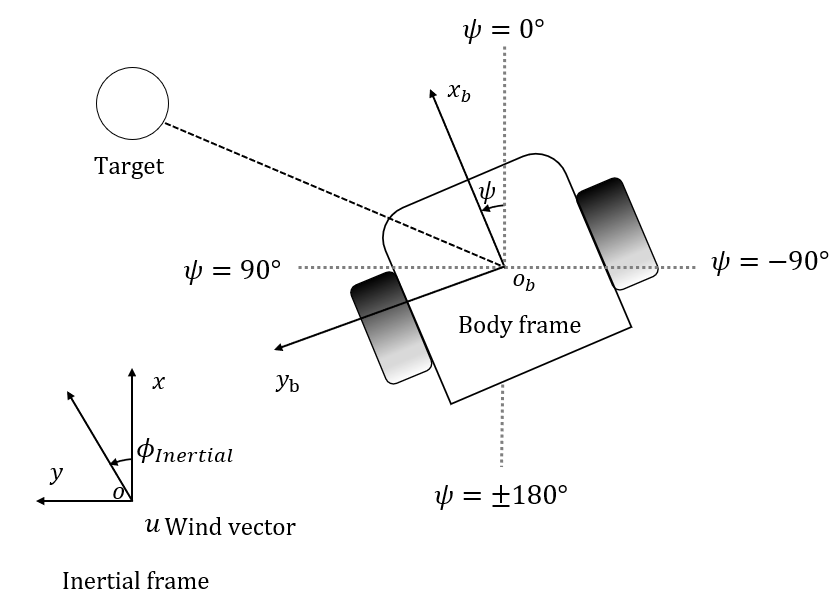
Luong, D. N., & Kurabayashi, D. (2023). Odor Source Localization in Obstacle Regions Using Switching Planning Algorithms with a Switching Framework. Sensors, 23(3), 1140.

**Author’s Response:** *Thank you for your feedback!*

*We have added the citation in line 143 of section 2.*

**Feedback 5:** Page5, Eq. (1): Please explain in more detail. What does the angle of the output represent? Is it the attitude in the robot coordinate system? And what exactly is the angle in the inertial frame?

**Author’s Response:** *Thank you for your feedback!*

**

*Fig: robot coordinate system.*

*Crosswind movement, where the robot heading is perpendicular to the wind direction, increases the chance of the robot detecting odor plumes. In the figure, is the inertial frame, which represents the fixed global frame. And is the body frame, which is the local frame fixed on the robot. Denote that the wind direction in the inertial frame is . Following the `Crosswind maneuver' behavior the robot target heading in the inertial frame psi is defined as degrees.*

*The "Crosswind maneuver" behavior of the proposed fusion navigation algorithm outputs target heading angle in the inertial frame. The robot then changes the angular velocity to match the target heading.*

*We have added figure 3 explaining the robot notations and added lines 194-197 of subsection 3.2.*

**Feedback 6:** Page5, Chap3.3: The big question here is why discrete behavior control even though LiDAR is installed? As you know, with LiDAR, it is easy to create spatial maps from point cloud data, and obstacle avoidance can be done more smoothly. Therefore, it should also be possible to avoid behind obstacles, which would be a disadvantage for olfactory navigation. Despite the system having rich data, it is necessary to justify the meaning of using very classical control.

**Author’s Response:** *Thank you for your feedback!*

*Q: Why we don’t use classic obstacle avoidance:*

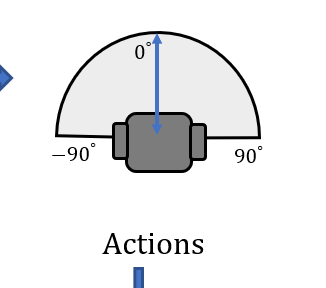
*A: It should be mentioned that the proposed navigation algorithm does not have access to the global map of the search environment, the locations of obstacles and the odor source. Therefore, planning-based obstacle avoidance algorithms like Artificial Potential Field algorithm (APF), A-star algorithm, Dijkstra (DJ) algorithm, etc. are not applicable in our problem. A similar approach was employed in [1].*

*[1] Kahn, G., Abbeel, P., & Levine, S. (2021). Badgr: An autonomous self-supervised learning-based navigation system. IEEE Robotics and Automation Letters, 6(2), 1312-1319.*

*In such partially observable environments, using discrete behavior control would be the best option. Comparing to classic motion-planning algorithms (like DJ, A-star, APF, etc.), our proposed obstacle avoidance algorithm has following advantages:*

1. *It does not rely on prior knowledge of the global map, or location of obstacles or destination.*
2. *Compared to most deep learning-based navigation planners, it requires less inference time.*

*In the proposed work, our robotic system only moves forward. A possible future heading selection is from -90 to 90 degrees, as presented in the following figure.*

**

*Fig: robot heading range.*

*Therefore, it only needs to detect and avoid obstacles in its approximate field of view.*

*For clarification, we have added lines 229-240 of subsection 3.3.*

**Feedback 7:** Page5, Chap3.4: This time, the authors are using yolo to recognize objects, which is supervised learning. This time, what are you using as an odor source to make the robot learn? Dose the robot learn vapor? Or does it learn the humidifier in Fig. 4? Considering the practical aspect, it is most likely an unknown odor source, and supervised learning is difficult to use. Moreover, the authors mentioned that the odor source is a vapor-emitting object, as you know, there are also many colorless and transparent chemicals. In addition, in your case, you employed something like a scent diffuser, but it is quite possible that it is in a completely different container and is not emitting vapor, or that the paper is coated with chemicals. Given this, this technique is quite limited and is not expected to work in situations beyond laboratory-level experiments. Based on the current explanation, the reviewer cannot agree with this method at all. The paper should be rewritten to justify the method, and the limitations of the research should be made clear.

**Author’s Response:** *Thank you for your feedback! We break down your question into sub-questions and answer all of them in the following:*

*Q: This time, what are you using as an odor source to make the robot learn?*

*A: In subsection 3.4 of the paper, we have discussed the training of the vision model. We use* ***a humidifier*** *as the odor source to emit odor plumes, and the vision model is trained to detect visible odor plumes.*

*Q: Does the robot learn vapor?*

*A: Yes, the proposed vision model is trained to detect vapors.*

*Q: Or dose it learns the humidifier in Fig. 4?*

*A: No, the proposed vision model is trained to detect vapors, not the humidifier.*

*Q: Considering the practical aspect, it is most likely an unknown odor source, and supervised learning is difficult to use.*

*A: Visible odor plumes are a common feature of many real-world odor sources (e.g., smoke sources, chemical sources, hydrothermal vents, etc.). Therefore, we trained the vision model to detect odor plumes (or vapors) instead of the shape of the odor source, since the shape of odor source is unknown and flexible. This setting makes our proposed vision model generalized to detect various kinds of odor sources, regardless of their shapes.*

*Q: As you know, there are also many colorless and transparent chemicals.*

*A: In the case of transparent odor plume, vision is not effective in finding the odor source this scenario. However, it should be mentioned that our proposed navigation algorithm* ***combines vision and olfaction****. In the event of vision cannot provide valid odor source information, olfaction is employed to guide the robot to approach the odor source location.*

*Q: In addition, in your case, you employed something like a scent diffuser, but it is quite possible that it is in a completely different container and is not emitting vapor, or that the paper is coated with chemicals.*

*A: It is true that odor sources could have different shapes, but the proposed vision model is trained to learn a common feature of odor sources, i.e., vapors, not trained to learn specific odor source shapes. Thus, our proposed vision model can be generalized to detect odor sources in different shapes.*

*In the event of colorless or transparent vapors, the proposed navigation algorithm is still valid in guiding the robot to find the odor source since it employs both vision and olfaction. When the vision model cannot detect valid odor vapors, the olfaction can be employed to detect odor concentrations to guide the robot to find the odor source location. Our experiment results verified that the proposed navigation algorithm, which combined vision and olfaction, was more effective than vision-only or olfaction-only navigation algorithms.*

**Feedback 8:** Page8, Chap3.6: This is a condition for the end of the experiment, not an automatic "source declaration" by the robot. Therefore, it is inappropriate to put it here. It should be placed in the design of the experiment.

**Author’s Response:** *Thank you for your feedback!*

*We have shifted the ‘Source Declaration’ to subsection 4.4 in section 4.*

**Feedback 9:** Page8, Figure 5: The robot picture is small and difficult to see. It is recommended that the picture of the robot be displayed independently.

**Author’s Response:** *Thank you for your feedback!*

*We have added an independent picture of the mobile robot in subsection 4.1.*

**Feedback 10:** Page13, Figure 10: Fig. 10 and Fig. 11 are exactly the same result, so there is no increase in the amount of information. Fig. 10 (or Fig. 11) should be deleted.

**Author’s Response:** *Thank you for your feedback!*

*We have deleted figure 10 and updated subsection 4.6.*

**Feedback 11:** Page13, Chap4.6: This is not a STATISTICAL ANALYSIS at all. What specific statistical method is used? Is there any significant difference in success rate or search time depending on the method used? I think that 5 trials are too small for statistical analysis.

**Author’s Response:** *Thank you for your feedback!*

*Q:* What specific statistical method is used?

*A: We have used descriptive statistics of averaged search time and success rate, two commonly-used criteria, for evaluating the performance of the three odor source localization algorithms.*

Q: Is there any significant difference in success rate or search time depending on the method used?

*A: Since we only used simple descriptive statistics to compare the performances of the three navigation algorithms, we have deleted the subsection ‘Statistical Analysis’, and added the content to subsection 4.6 Repeated Experiment Trials to clarify this question.*

**Feedback 12:** Page13, Chap4.5 and 4.6: There was no discussion of why combining vision and olfaction would improve the results, just showing the results. The authors should discuss how the combination contributes to the function of OSL.

**Author’s Response:** *Thank you for your feedback!*

*In our experiments, Olfaction-only navigation algorithm performed well in laminar airflow environments - the robot followed relatively direct airflow towards the odor source. However, in turbulent airflow environments, the robot got diverted by the complex airflow directions and often failed to reach the odor source by the designated time limit. In comparison, the proposed Vision and Olfaction Fusion Navigation algorithm test runs were consistently successful in both laminar and turbulent airflow environments. The Crosswind maneuver and Olfaction-based Navigation led the robot toward the odor source which allowed the robot to detect plume vision. Once it started to follow Vision-based Navigation, the robot was not affected by turbulent airflow. Thus, the combination of vision and olfaction contributes to better odor source localization performance in turbulent airflow environments.*

*We have added the discussion of the proposed OSL algorithm in lines 382-401 of subsection 4.5.*

**Feedback 13:** Page14, L361-362: It is unfair to calculate the average search time by assuming that a failed trial is 200 seconds. It is intentionally increasing the search time for conditions with many failures. The search time should be calculated only for successful trials.

**Author’s Response:** *Thank you for your feedback!*

*According to your suggestion, we have recalculated average search times only for successful trials and updated the Tabe 3. This time, we only took the actual times taken by successful runs to calculate the Average Search times. For example, Olfaction-only Navigation Algorithm managed to navigate to the odor source twice in turbulent airflow environment. Thus, we added the two times and divided by two to get the Average Search Time taken by Olfaction-only Navigation Algorithm in the turbulent airflow environment.*