

**Title:** Automated urban flood level detection based on flooded bus dataset using YOLOv8

**Manuscript Number:** egusphere-2024-4053

**Authors:** Yanbin Qiu, Xudong Zhou, Jiaquan Wan, Tao Yang, Lyfei Zhang, Yuanzhuo Zhong, Leqi Shen, and Xinwu Ji

Dear Referee,

We are truly grateful for your meticulous review and valuable suggestions on our manuscript. Based on your comments, we have revised the manuscript point-by-point and provided detailed explanations for the relevant issues. We have strived to enhance the rigor and clarity of the paper through these improvements and sincerely hope to gain your approval. The specific responses are as follows:

### **Referee #1**

#### **1) Regarding the validation set:**

##### **Comment:**

*"Can the authors please clarify whether the 10% validation set is synonymous with a 'test set'? Specifically, was this set completely withheld during training—including during data augmentation—such that no original or augmented images in this set were seen by the model?"*

##### **Response:**

Thank you for your valuable comment. We would like to further clarify that the 10% validation set mentioned in our study can indeed be considered equivalent to a "test set" as you interpreted. Throughout the entire training process, the validation set was strictly separated from the training set, and the model never accessed any original or augmented images from the validation set.

During data processing, data augmentation techniques were applied exclusively to the training set, while the validation set remained as untouched original images. The validation set was used solely for performance evaluation after each training epoch, primarily to monitor the training process, prevent overfitting, and ensure a fair assessment of the model's generalization ability.

##### **Changes in manuscript:**

In the revised manuscript, we have supplemented relevant explanations on page 10, lines 225–230.

## **2) Regarding hyperparameters tuning:**

### **Comment:**

*"Did the authors adjust any hyperparameters of the YOLOv8 algorithm? If so, could they describe the tuning process?"*

### **Response:**

Thank you for raising this point. In this study, we employed the official default hyper parameters settings of YOLOv8 without any additional tuning. Parameters such as learning rate, batch size, and confidence threshold were maintained at their default values to ensure the consistency and reproducibility of the model results.

### **Changes in manuscript:**

In the revised manuscript, we have supplemented relevant explanations on page 10, lines 230–231.

## **3)Regarding the evaluation of complex scenes:**

### **Comment:**

*"The authors discuss two example images as case studies for 'complex' scenes. Can the authors elaborate whether there is a larger dataset of such complex scenes on which the model performance was evaluated? If not, how were these two specific examples selected? How do the authors anticipate the model will generalize to similar complex scenarios?"*

### **Response:**

Thank you for your insightful question regarding the evaluation of complex scenes. Currently, there is no publicly available large-scale dataset of bus flood inundation images, and the images retrieved from social media predominantly depict regular scenes rather than extreme complex scenarios. Due to the scarcity and dispersion of such complex scene images, we were unable to construct an independent large-scale evaluation dataset.

The two examples were selected based on their representative difficulty and relevance to the target application. The selection criteria included: (1) ensuring diversity by covering common interference factors such as low-light nighttime conditions, object occlusion, and multiple object overlaps, thereby avoiding bias caused by a single disturbance type; (2) prioritizing scenes presenting compound challenges that were not

sufficiently covered during training, to better assess the model's adaptability to unseen complex environments.

During the training process, we employed multi-level and diverse data augmentation strategies to significantly increase the complexity and diversity of the training data, encouraging the model to learn more robust feature representations. Experimental results demonstrate that the augmented model achieved notable improvements in complex scene detection. Based on the current results, we believe that the model has good generalization potential and can effectively handle unseen complex scenes. In future work, we plan to further validate the model performance on a larger-scale dataset.

#### 4) Suggestion to introduce YOLOv8 earlier

##### Comment:

*"To improve accessibility of the manuscript for a broader audience, I would suggest introducing the YOLOv8 algorithm with a short description earlier in the manuscript."*

##### Response:

We agree with this suggestion and have added a concise description of the YOLOv8 algorithm in the Introduction section to help readers unfamiliar with the model.

##### Changes in manuscript:

We add a brief introduction to YOLOv8 in the Introduction section on page 4, lines 108–117

#### 5) Specific line edits

Below we address each line-specific issue and corresponding corrections:

	Comment	Response	Page & Line(s)
1	Line 14 - have been emerged	Corrected to "have emerged".	Page 1, Line 14
2	L19 - YOLOv8 is referenced without a preceding description.	Added a short description	Page 1, Lines 16–17
3	L23 - as they remain *in* service	Grammar corrected for clarity.	Page 1, Line 25
4	L69 - Park et al. 2021 is cited twice for the same statement.	Removed duplicate citation.	Page 3, Lines 87–88
5	L70 - Suggest describing YOLO as a CNN-based CV model prior to first usage.	Added description at first mention.	Page 2, Lines 68–73

	<b>Comment</b>	<b>Response</b>	<b>Page &amp; Line(s)</b>
6	<i>L85 - submerged states of buses *is* categorized</i>	Corrected to “ <i>are categorized</i> ”.	Page 4, Line 105
7	<i>L88 - Missing source citation</i>	Added appropriate source.	Page 4, Line 116
8	<i>L101 - configurations, *and* explains the experimental design *and* model evaluation metrics</i>	Revised for consistent phrasing.	Page 5, Line 129
9	<i>L113 - images in exhibit</i>	Rephrased to “ <i>images used in the experiment vary in</i> ”.	Page 5, Line 141
10	<i>L122 - What does “instances” refer to?</i>	The instances labeled with Labeling refer to the buses and their submersion statuses that have been annotated in the images.	Page 5, Line 159
11	<i>L165 - Suggest expanding acronyms at first usage.</i>	Acronyms fully expanded at first use.	Page 8 Line 201
12	<i>L189 - Based on the sentence, it appears that the 90-10 split was done after data augmentation, while the correct approach will be to perform data augmentation only on the training set, to avoid data leakage. Please clarify in text.</i>	Clarified the procedure.	Page 10, Lines 225–230
13	<i>L198 - two particularly demanding *scenes*</i>	Improved wording.	Page 10, Line 239
14	<i>L206 -The statement is unclear.</i>	Rewritten for clarity.	Page 11, Lines 247–249
15	<i>Eq 4 - The parameters - n, AP, P, R - are undefined.</i>	Defined all terms in the text.	Page 11, Lines 270–272
16	<i>L228 - Suggest introducing IoU prior to first usage.</i>	Added short definition at first use.	Page 12, Lines 273–277
17	<i>L426 - all four YOLOv8 models may exhibited</i>	The revised content has been updated	Page 24, Line 460

## **Referee #2**

### **1) On the Practical Significance of Water Depth Estimation:**

#### **Comment:**

*"The manuscript should clarify the practical implications of the estimated flood depths. For instance, the difference between 20 cm and 45 cm of water may not significantly affect early disaster response decisions. In contrast, a depth corresponding to the average floor height might indicate a risk of people being trapped and requiring rescue. I recommend the authors provide more context on how the predicted water depths contribute to emergency decision-making."*

#### **Response:**

We are grateful for your constructive suggestion. The practical significance of water depth estimation in emergency response certainly deserves further clarification. In the original manuscript, our primary focus was on estimating flood water depth distributions from social media images to help identify potential flood risk areas in urban settings.

In practical applications, even a difference of just a few centimeters in water depth can have varying impacts. For example, 20–30 cm of flooding may cause vehicle stalls or obstruct pedestrian movement, while water levels approaching or exceeding 40–50 cm are more likely to enter residential or commercial areas, posing a significant threat to personal safety and property. By referencing objects like buses and bus door steps in the images, we attempt to provide a visual risk indicator. We hope that the predicted depth ranges will assist emergency response teams in assessing flood severity, identifying high-risk areas, and planning priority responses. In the revised manuscript, we will carefully consider your suggestions and provide further clarification on this point. Thank you again for your valuable feedback.

#### **Changes in manuscript:**

In the revised manuscript, we have supplemented relevant explanations on page 2, lines 43–61.

### **2) On Using Buses as Reference Objects:**

#### **Comment:**

*"The authors argue that using buses as reference objects improves accuracy. However, it is unclear whether potential buoyancy or floating of the buses was considered. Did the authors verify that the buses used as reference points remained stationary during the flood? Also, please specify the assumed bus height used in the model calibration or estimation."*

**Response:**

We appreciate your attention to this issue. During the model development process, we paid particular attention to the reliability of the reference objects. All images used for training and testing were manually screened to ensure that the buses in the images were either stationary or moving normally in the floodwaters, and that there was no buoyancy or floating involved, ensuring that buses remained stable reference points for accurate measurements.

We standardized the height of the buses in our model to approximately 3 meters, referencing typical urban bus models seen in current social media images. Based on a systematic evaluation of the images, we classified flood water depth into several levels, with the majority of images showing water levels between 0–50 cm. Only in a few rare cases did the water level exceed 100 cm. Given the rarity of high-water scenes (i.e., >100 cm) in urban environments, we chose to focus on the more common water levels for classification, without setting higher water levels, as we believe this is more aligned with realistic application needs and image distribution characteristics.

**Changes in manuscript:**

In the revised manuscript, we have supplemented relevant explanations on page 5, lines 146–148 and on page 6, lines 150–157.

**3) On the Description of Data Augmentation:**

**Comment:**

*"While data augmentation is widely recognized to improve model performance, this is already a well-established practice. It may not warrant substantial emphasis in the discussion and conclusions unless the authors offer a novel or particularly insightful implementation."*

**Response:**

Thank you for your comment regarding data augmentation. We plan to modify and simplify this section in the revised manuscript.

**Changes in manuscript:**

Reduced emphasis on augmentation in the discussion and conclusions.

**4) On Providing a Brief Introduction to the YOLO Model:**

**Comment:**

*"The Introduction should include a short explanation of YOLO models, especially considering that not all NHESS readers are familiar with machine learning or object detection frameworks."*

**Response:**

We appreciate your suggestion. We understand that some readers of NHESS may not be very familiar with machine learning and object detection frameworks. Therefore, in the revised manuscript, we will add a brief introduction to the YOLO model in the introduction section, as you suggested.

**Changes in manuscript:**

A brief description of YOLO models has been added in the Introduction (page 4, lines 108–117).

**5) On Established Methods for Estimating Flood Water Depth:**

**Comment:**

*"Are there other established methods for estimating flood depth beyond analyzing social media imagery? If so, a short overview in the Introduction would help situate the proposed approach within the broader context."*

**Response:**

We appreciate your suggestion for expanding the background of the study. In fact, we briefly reviewed the main methods for urban flood water depth estimation in the original manuscript, including water level gauges, remote sensing technologies, and hydrodynamic models.

We pointed out that while water level gauges provide accurate point measurements, their high deployment and maintenance costs limit their widespread use. Remote sensing methods are mainly used to identify flood inundation areas, but they do not yet have the

capability to directly estimate water depth. While hydrodynamic models can estimate water depth, they require high-quality input data, are computationally complex, and have slower response times, making them less suitable for real-time emergency response. We believe that the background provided supports the validity of our method using social media images for flood water depth estimation and shows that it can serve as an effective complement to traditional methods. Due to space limitations, the discussion of these methods in the original manuscript was relatively brief, but it covers the core points.

**Changes in manuscript:**

We have expanded the Introduction (page 3, lines 62–69) to briefly mention other methods.

**6) On the Discussion Section:**

**Comment:**

*"The Discussion needs to be strengthened. I encourage the authors to include a critical evaluation of their method, its limitations, and a comparison with related studies—particularly those using alternative reference objects."*

**Response:**

Thank you for your valuable suggestion regarding the discussion section. We fully agree with your point that we need to more comprehensively assess the limitations of our method and compare it with existing research, particularly studies using other reference objects.

In the revised manuscript, we will modify the discussion section in accordance with your suggestion, providing a more detailed analysis of the method's limitations. We will also add comparisons with other studies to help readers better understand the relative strengths and weaknesses of our method.

**Changes in manuscript:**

We have expanded Discussion with limitations and comparative analysis on page 25–27, lines 491–538.

**7) Minor comment:**

**Comment:**



*"Line 329: Please rephrase the sentence "The numbers on the image represent...".  
The current wording is unclear and may confuse readers."*

**Response:**

Thank you for pointing this out. We have revised this sentence for clarity.

**Changes in manuscript:**

We have rephrased the sentence for clarity (page 17, line 337).

We sincerely thank both referees again for their detailed and constructive feedback.

We believe that these revisions have substantially strengthened the manuscript.