

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

**Manuscript Number:** egosphere-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) Comment:**

*"Line 57: In addition, in Texas alone... The language makes it appear that the ratio of evacuation related fatalities would be larger if other states were included, which is likely untrue. Suggest changing the sentence to: For example, in Texas during the XXX flooding event, it was estimated that approximately 75% of flood-related fatalities occurred during evacuation efforts via local roadways, primarily due to the lack of awareness regarding inundation depth in the surrounding areas. (Alizadeh Kharazi and Behzadan, 2021)"*

#### **Response:**

Thank you for this helpful suggestion. We have revised the sentence.

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

We have rephrased the sentence for clarity (page 2, line 58).

#### **2) Comment:**

*"Line 154: The authors state that images with water depth exceeding 100 cm are rare, although Table 1 shows roughly equal distribution across the 4 levels. Can the authors clarify if this is because they intentionally constructed a fairly balanced set?"*

#### **Response:**

Thank you for raising this point. We did not intentionally construct a class-balanced dataset. Images with water depth exceeding 100 cm are indeed fewer in our raw

collection; however, within those few images we collected, there is a relatively high number of bus instances, which makes the instance-level counts for this depth range comparable to the other levels and thus appear approximately balanced in Table 2. Our annotations were performed objectively without deliberate class balancing.

**Changes in manuscript:**

In the revised manuscript, we have added a clarification on page 6, lines 167–169.

**3)Comment:**

*"Line 227: What does it mean by the validation set expanded from 108 to 198 mages? If the authors are referencing their previous manuscript version, the readers would have no information about those previous versions, therefore only the final experiment setup should be described."*

**Response:**

Thank you for the helpful comment. We have revised the sentence to improve clarity.

**Changes in manuscript:**

In the revised manuscript, we have rephrased the sentence for clarity on page 10, line 238.

**4) Comment:**

*"Validation set: Can the authors describe the split of validation set across the 4 levels? It could be included within Table 1, or a similar table could be added."*

**Response:**

Thank you for the suggestion. We have added the level-wise distribution of the validation set. This information is now presented in Table 2 and referenced in the Methods section.

**Changes in manuscript:**

We added a sentence in the Methods section (page 10, line 239) describing the distribution of the validation set across the four levels.

**5) Comment:**

*"Section 2.2.3: I suggest the authors add a statement about the difficulty in collecting a larger dataset of complex scenes, reflecting their response to my previous comments. This would help address this question for future readers."*

**Response:**

We appreciate this suggestion. The relevant statement has been added to Section 2.2.3.

**Changes in manuscript:**

In the revised manuscript, we have added a relevant statement on page 10, lines 251–255.

**6) Comment:**

*"Line 247: The statement seems to convey that the authors are assessing the general performance of YOLOv8 compared to previous models, however, they are only assessing the performance for their current dataset. Suggest changing the statement to: Although the introduction states that YOLOv8 is the latest algorithm in the YOLO series and has been known to perform better than earlier versions on a general image dataset, comparative analysis with earlier versions was performed for this dataset to quantify performance differences."*

**Response:**

Thank you for this helpful suggestion. We have revised the sentence.

**Changes in manuscript:**

We have rephrased the sentence for clarity (page 11, lines 264-266).

**7) Comment:**

*"Line 273: For improved clarity, suggest adding a statement like the following prior to defining IOU: Additionally, the metric Intersection over Union (IOU) was also calculated."*

**Response:**

Thank you for this helpful suggestion. We have revised the sentence.

**Changes in manuscript:**

We have rephrased the sentence for clarity (page 12, line 290).

**8) Comment:**

*"Line 342: The statement is unclear without a qualifier for what improved Level 3 and 4 performance. Suggest changing the statement to: Notably, the detection results for higher-risk categories (Level 3 and Level 4) show improved AP values in all models trained with augmented images, as evidenced by the Precision-Recall curves shifting closer to the upper-right corner."*

**Response:**

Thank you for this helpful suggestion. We have revised the sentence.

**Changes in manuscript:**

We have rephrased the sentence for clarity (page 15, lines 360-361).

**9) Comment:**

*"Line 461: The following statement does not appear to address the specific advantage of data augmentation, and can be removed since the next statement already addresses it: This improvement can be credited to the enhanced learning and generalization abilities of models."*

**Response:**

Thank you for this helpful suggestion. We have removed the sentence.

**Changes in manuscript:**

We have removed the sentence for clarity (page 24, line 479).

**10) Comment:**

*"Line 505: The clause - they are usually deployed only in limited areas - is unclear. What does deployment refer to? Suggest clarifying/removing the clause."*

**Response:**

Thank you for this helpful suggestion. We have removed the clause.

**Changes in manuscript:**

We have removed the clause for clarity (page 26, line 522).

**11) Comment:**

*"Line 517: Since edge-device deployment wasn't tested by the authors in this study, suggest providing a reference for this claim."*

**Response:**

Thank you for this helpful suggestion. We have provided a reference for this claim.

**Changes in manuscript:**

In the revised manuscript, we have provided a relevant reference. (page 26, line 533).

**12) Comment:**

*"Line 531: Suggest providing references for illumination-guided transformers and light-aware attention mechanisms."*

**Response:**

Thank you for this helpful suggestion. We have provided a reference for illumination-guided transformers and light-aware attention mechanisms.

**Changes in manuscript:**

In the revised manuscript, we have provided a relevant reference. (page 27, line 548).

## Referee #2

### 1) Comment:

*"The details of data collection in Section 2.1.1 are insufficient: The article mentions that images are from "Baidu, Google, Douyin, and WeChat" but fails to specify the specific search keywords, time range, and screening criteria. These pieces of information are crucial for evaluating the representativeness and reproducibility of the data. It is necessary to supplement the spatiotemporal distribution statistics of data sources (such as the proportion of each platform, year distribution) and screening rules."*

### Response:

We are grateful for your constructive suggestion. In the revised manuscript (Section 2.1.1), we have supplemented the specific time range for data acquisition, stating that the dataset covers the most recent five years. In addition, we have explicitly described the screening criteria, which ensures that buses remain valid, stable reference objects. These additions improve the clarity, reproducibility, and representativeness of our dataset.

### Changes in manuscript:

In the revised manuscript, we have supplemented relevant explanations on lines 144–156.

### 2) Comment:

*"In Section 2.2.2, the hyperparameter settings use default hyperparameters (e.g., batch size = 16, epochs = 100) without explaining the basis. It is recommended to supplement the reasons for choosing the default values."*

### Response:

Thank you for your comment regarding our use of default hyperparameters. We adopted the official default settings of YOLOv8 because these values were established by the Ultralytics team through large-scale pretraining and benchmarking on COCO ((Common Objects in Context)) and other datasets, and are recommended as a configuration that achieves a robust balance among detection accuracy, convergence stability, and computational efficiency. Designed for general downstream object-detection tasks, this configuration has been widely used in recent detection and recognition studies; we have added supporting references in the revised manuscript. In line with this common

practice—and considering our hardware resources and the need for stable training—we therefore used the official defaults.

**Changes in manuscript:**

In the revised manuscript, we have supplemented relevant explanations on page 10, lines 241–244.

**3) Comment:**

*"Some research work can be useful. Water inrush mechanism and variable mass seepage of karst collapse columns based on a nonlinear coupling mechanical model. Multiphysics modeling of thermal-fluid-solid interactions in coalbed methane reservoirs: Simulations and optimization strategies. Diffusion evolution rules of grouting slurry in mining-induced cracks in overlying strata."*

**Response:**

Thank you for providing the three related studies. We have added citations to these works in the Introduction of the revised manuscript.

**Changes in manuscript:**

Citations to the relevant literature have been added in the Introduction of the revised manuscript (lines 7 and 53).

**4) Comment:**

*"Section 3.1.1 mentions that mAP50-95 has increased by 20%, but there is no analysis of the performance of specific categories under different risk levels, making the significance of the indicator improvement unclear. It is suggested to decompose the mAP improvement ratio by risk level."*

**Response:**

Thank you for your valuable suggestion. In response, we have added a detailed analysis of the mAP50-95 performance across different flood risk levels in Section 3.1.1, as shown in the newly added Table 5.

**Changes in manuscript:**

A new table (Table 5) and corresponding analysis were added to Section 3.1.1 to present mAP50–95 improvements by flood risk level. (page 18, lines 389–401).

We sincerely thank both referees again for their detailed and constructive feedback.  
We believe that these revisions have substantially strengthened the manuscript.