Response to the Reviewers Comments

Author reply to RC1 egusphere-2025-77 (NHESS)

Paper title: Identification of nighttime urban flood inundation extent using deep learning

Authors: Jiaquan Wan et al.

GENERAL COMMENTS

This study proposes a method for recognizing flooded urban areas by analyzing night-time images taken by urban surveillance cameras during a rainy event. The method is based on combining deep learning model (using advanced semantic segmentation techniques) with video images as a data source to effectively perform real-time urban flood recognition, which addresses common challenges in the field of research into the identification of night-time urban flooding. The proposed Model - nighttime waterlogging recognition model —NWseg - outperforms the other models tested in this study for detecting urban nighttime flooding. The ability of the models to recognize flooding on a series of night-time flood images was used as a basis for validating the experimental results. My minor and major comments are presented in the next paragraphs.

Response: Sincerely thank the reviewer for the recognition and detailed evaluation of this study. We look forward to your feedback and commit to revising the manuscript earnestly based on your specific suggestions to further enhance its scientific rigor and readability. Below, the reviewer's comments are presented in black, and the authors' responses are provided in blue.

Minor comments

- Line 22: Please add space
- Line 33: It would be interesting to add references about complex climate change
- Line 219: Please remove comma repetition
- Line 240: Please add space
- Line 317: Please add space (Check along the text)

Response: We sincerely appreciate the reviewer's careful review and valuable comments. We have fully revised the manuscript according to the suggestions, including formatting corrections, removing redundant punctuation, and adding relevant references. Additionally, we checked the entire text to ensure consistency. In Line 31, we have added references about complex climate change as follows.

References:

Burn, D. H., Whitfield, P. H.: Climate related changes to flood regimes show an increasing rainfall influence, Journal of Hydrology, 617, 13, http://doi.org/10.1016/j.jhydrol.2023.129075, 2023.

Kim, H., Villarini, G., Wasko, C., Tramblay, Y.: Changes in the Climate System Dominate Inter-Annual Variability in Flooding Across the Globe, Geophysical Research Letters, 51(6), http://doi.org/10.1029/2023gl107480, 2024.

Major Comments

1. Line 44 : ‘‘… remote sensing techniques have difficulty in capturing subtle topographic changes within cities …’’. Can the proposed model do it?

Response: We sincerely appreciate the reviewer’s valuable comments and extend our sincere apologies for any potential misunderstanding caused by our wording. We have improved this sentence as “*Due to the limitations in temporal resolution and the impact of cloud cover and atmospheric variations, remote sensing technology struggles to capture the dynamic changes of urban flooding, making real-time monitoring of rapidly evolving flood events challenging.*”

2. Line 70 : ‘‘… under complex weather conditions …”. Let's imagine that a heavy or exceptional nighttime rainstorm caused a power cut and consequently darkness in the city. What would be the performance (flood recognition) of the proposed NWseg model?

Response: Thanks to your suggestion, we have added the description about model performance in all-black scenarios in the new section “5 Discuss” as follows:

“*On the other hand, in nighttime scenarios with extremely low illumination or even complete power outage (e.g., the case of city blackout triggered by heavy rainfall), the model has difficulty in extracting effective edge and texture information, which leads to a significant degradation of the recognition performance.*”

3. Line 220 : ‘‘… GPU acceleration libraries…’’. What are the calculation times? Compatible with real-time forecasting? Please specify. Indeed the ability to provide results quickly to facilitate decision-making is essential in urban environments, given the economic, social, and other issues at stake.

Response: Thanks to your suggestion, we have added the description about the actual operating efficiency of the model in the new section “5 Discuss” as follows:

“*In addition, NWseg achieves an inference speed of 37.8 FPS (i.e., approximately 26.5 milliseconds per image) under the NVIDIA GeForce RTX 3080 environment, demonstrating its potential for real-time applications in high-performance computing platforms.*”

4. Line 274 : An important element in the characterization of a flood is its depth, velocity and extent. It is not clear to me whether the model detects only the extent or both extent, velocity, and depth.

- How does the NWseg model calculate depth? velocity?
- To what spatial scale is the model applicable (Street? District? Whole city?)

Response: Thanks to your suggestion, we have added the description about the applicable spatial scale of the model in the section “1 Introduction” as follows: “*Meanwhile, Given that the data are mainly sourced from urban road surveillance systems, the method is particularly suitable for street (Street) and local area (District) scale flood detection.*” Furthermore, it is further clarified in the full text that this model is only used for flood extent identification tasks.

5. Line 329 : ‘‘… Future work will focus on reducing the model’s parameters…”. The various model parameters have not been detailed.

Response: Thanks to your suggestion, we have added the parameter information of each model and provided a detailed analysis in the section "4 Result".

6. The study should robustly integrate a discussion section providing the advantages, limitations, and improvements of the proposed model.

Response: Thanks to your suggestion, we have added a new section "5 Discuss" as follows:

"*5 Discuss*

In this study, a state-of-the-art model named NWseg is proposed to address the challenges of nighttime urban flood extent identification. Through a series of experimental validations, the NWseg model demonstrates superior performance with 95.99%, 94.8%, 95.39%, and 91.46% in Precision, Recall, F1 score, and MIoU, respectively. In the prediction comparison of real scenarios, the model also shows high accuracy and robustness, and effectively recognizes flooded areas in complex nighttime environments. In addition, NWseg achieves an inference speed of 37.8 FPS (i.e., approximately 26.5 milliseconds per image) under the NVIDIA GeForce RTX 3080 environment, demonstrating its potential for real-time applications in high-performance computing platforms. This study bridges the current research gap in flood extent recognition in nighttime scenarios, providing a technical reference for flood monitoring and emergency response.

Nevertheless, this study still has some limitations. First, the overall structure of NWseg is relatively complex, and the model parameters are large in scale, which limits its deployment capability on resource-constrained edge devices. On the other hand, in nighttime scenarios with extremely low illumination or even complete power outage (e.g., the case of city blackout triggered by heavy rainfall), the model has difficulty in extracting effective edge and texture information, which leads to a significant degradation of the recognition performance. In the future, we will further optimize the network structure to reduce the computational complexity of the model and improve deployment flexibility. In addition, we consider combining infrared thermal imaging, low-light image enhancement, or multimodal fusion methods to improve the robustness and generalization ability of the model under extreme low-light conditions."

Author reply to RC2 egusphere-2025-77 (NHESS)

The paper presents an interesting contribution on the use of deep learning to identify flood extents in urban areas from night-time images. As such it is relevant to NHESS, but it has a strong focus on the deep learning methods. Many readers of the journal, like myself, will not be familiar with the detailed language and concepts used in deep learning and the paper needs to be re-written with this audience in mind. Sections 2.1 and 2.2 need particular attention.

Response: We sincerely appreciate the reviewer's recognition of our study and the constructive feedback provided. We have revised Sections 2.1 and 2.2 to simplify the language, reduce the use of technical terms, and add relevant background information to enhance readability.

Much of the language in the paper is opaque and uses terms that are not common in scientific discourse (some of these are set out below). There also needs to be greater clarification of what the authors set out to do in the research, which of the methods they developed themselves, and what the conclusions mean for those working on urban flooding. For example, on Line 82 the aims are listed, but these are in fact a description of what was done without a justification. Further Line 74 says that NWseg is "proposed" and Line 91 says that the NWseg is "contributed", but was it developed by the authors or taken from other research work?

Response: We sincerely appreciate the reviewer's in-depth critique and constructive suggestions regarding the language and content of this paper. In the revised manuscript, we have comprehensively improved the language, clarified the research objectives and methods, and further elaborated on the practical significance of our conclusions.First, we simplified the complex deep learning terms in the paper. At the same time, it was made clear that NWseg was proposed by us. Secondly, we stated our research objectives in the introduction as follows:

"*(1) Contributed a method for nighttime urban flooding extent identification based on urban surveillance cameras, aiming at realizing efficient assessment of nighttime urban flooding areas and filling the gaps of research in this field at this stage.*

(2) To support the generalization ability of the model in complex nighttime environments, this study constructs a nighttime flood inundation dataset covering a variety of nighttime scenarios (e.g., different weather, illumination intensity, and urban structure), which provides diverse sample resources required for training and testing.

(3) Replace the original DeepLabv3+ model network backbone with MobilenetV2 and ResNet101 networks and verify the effect of different network backbones on the performance of the Deeplavv3+ model.

(4) An urban flood identification model NWseg for nighttime scenarios is proposed, and the significant advantages of the model in terms of robustness, effectiveness and practicality are verified by comparing with other existing models, which advances the research and development of nighttime urban flooding extent identification."

In addition, the significance of the research conclusions to urban flood researchers is clearly stated in the conclusion section as follows: "*This study successfully demonstrates the superior performance of the NWseg model in nighttime urban flood detection, filling the research gap in nighttime flood range identification. Our work not only promotes the*

development of the field of nighttime urban flood identification, but also provides a reference for future deep learning applications under extreme lighting conditions."

Significantly more than half of the references that I tried to read online returned an error message or a message in Chinese characters. Further, many of these are to non-peer reviewed sources. Whilst the latter is acceptable in a few cases, the former is not at all acceptable in an international journal.

Response: We sincerely appreciate the reviewer's meticulous examination and constructive criticism regarding the references. We have revised the references with invalid links or those pointing to Chinese pages, and added new references.

Specific points are:

Line 50: a more scientific term that "remarkable" would be more appropriate.

Response: Thanks to your suggestion, we have replaced "remarkable results" with "*significant performance*".

Line 54: it is stated that such surveillance is "ubiquitous", but whilst this may be true in the authors' experience is not true in all countries. This should be acknowledged and it limits the usefulness of these methods.

Response: Thanks to your suggestion, we adjusted the formulation of "ubiquitous" by revising it as follows: "*particularly in highly urbanized areas*".

Line 55: of these three references two (Cheng, Yang) have links that do not work and one is to a publication that is not peer-reviewed. The remaining one does not mention whether this methodology has been tested in more than one country. Please clarify.

Response: Thanks to your suggestion, we have added related references as follows.

Reference:

Hao, X., Lyu, H., Wang, Z., Fu, S., and Zhang, C.: Estimating the spatial-temporal distribution of urban street ponding levels from surveillance videos based on computer vision, Water Resources Management, 36(6), 1799-1812, https://doi.org/10.1007/s11269-022-03107-2, 2022.

Wang, Y., Shen, Y., Salahshour, B., Cetin, M., Iftekharuddin, K., Tahvildari, N., Huang, G., Harris, D., Ampofo, K., and Goodall, J.: Urban flood extent segmentation and evaluation from real-world surveillance camera images using deep convolutional neural network, Environmental Modelling & Software, 173, 105939, http://doi.org/10.1016/j.envsoft.2023.105939, 2024.

Line 89: the term "ablation" is common in machine learning, but is an example of a term that needs explaining to a different audience.

Response: Thanks to your suggestion, we have added the description about ablation in the section "2.2 Typical semantic segmentation model" as follow: "*this study designs a series of controlled experiments, systematically modifying or removing network components to verify the effectiveness of different backbone networks (i.e., ablation studies) and compared the results with the NWseg model.*"

Line 99: there is no subject in this sentence so we cannot see who proposed the model.
Response: Thanks to your suggestion, we have revised the manuscript to clearly state that the NWseg model is proposed in this study.

Line 102: many terms in this paragraph need clarifying for a non-expert in machine learning, some examples are: SOD, illumination-independent reflectance, semantically supervising the training of the de-entanglement module, Retinex, Illumination-Aware Parser (IAParser), pyramid pooling module and a convolutional layer to construct an attention mask.
Response: Thanks to your suggestion. In the section "2.1 Nighttime Urban Segmentation Model", we revised and clarified many deep learning terms.

Line 115: a diagram showing how all these methods fit together would help readers understand what you are doing.
Response: We sincerely appreciate the reviewer's valuable suggestion. Line 115 of the manuscript primarily introduces other mainstream segmentation models to provide background for the subsequent performance comparison with our proposed NWseg model. The reviewer suggested using a diagram to illustrate the relationships among these methods for better comprehension. We have carefully considered this suggestion; however, since this study focuses on the performance comparison between NWseg and other models rather than an integrated framework of different approaches, we believe that the textual description is already sufficiently clear and that an additional diagram is unnecessary.

Line 189: what is a "Labelme tool"?
Response: Thanks to your suggestion, we have added the description about Labelme in the section "3.1 Construction of dataset" as follows: "*In addition, we employed Labelme, an open-source image annotation tool widely used in the field of computer vision, to manually annotate the flooded regions in the images. Through its graphical interface, annotators can polygonally map the inundated areas in an image and assign corresponding category labels to each area, thus generating high-quality semantic segmentation data that can be used for deep learning model training.*"

Line 191 says that the work was done by three graduate students. Rather than describing who did the work, it is necessary to explain how they did it and how the quality of the analysis was checked and ensured.
Response: Thanks to your suggestion, we have added the description about how the graduate students checked and ensured the quality of the annotations in the section "3.1 Construction

of dataset" as follow: "*Specifically, each flood image was annotated separately by all three annotators, followed by a cross-review process to identify potential discrepancies in the flood boundaries. In cases of inconsistency, the annotators engaged in multiple rounds of collaborative discussion and iterative refinement, optimizing the boundaries based on image details.*"

Line 193: "waterlogged" refers to soil saturation. I think "inundated" would be a better word.
Response: Thanks to your suggestion, we have replaced "waterlogged" with "*inundated*".

Line 196: this and other figures captions need more details.
Response: We are very grateful for the reviewer's detailed suggestions on figure captions. We have improved all figures captions and added more details.

Section 4.1: I think there should be less discussion of the three inferior methods as the differences between them are minor compared to their differences to NWseg.
Response: Thanks to your suggestion, we have revised the section "4.1 Ablation Study" as follows: "*In this section, we present a comparative analysis of the DeepLabV3+ model with different backbone networks and compare it with the NWseg model. As shown in Table 2 and Figure 3, all evaluation metrics are improved after replacing the original backbone network of DeepLabV3+ with MobileNetv2 and ResNet101, respectively. Notably, when ResNet101 was used as the backbone, the model achieved the best performance, with Precision, F1 score, Recall, and MIoU increasing by 14.4%, 10.11%, 6.63%, and 5.91%, respectively, compared to the baseline model. However, all DeepLabV3+ variants still exhibited a significant performance gap when compared to NWseg. The NWseg model achieved 95.99% in Precision, 94.80% in Recall, 95.39% in F1-score, and 91.46% in MIoU, demonstrating its superior capability in nighttime urban flood extent recognition. Although NWseg has a relatively large number of parameters, it delivers outstanding accuracy and robustness.*"

Line 250: I don't see any experimental results in the text i.e. data that was collected through physical measurements on site.
Response: We sincerely appreciate the reviewer's thorough examination and valuable suggestions regarding Line 250. The original text states, 'Training results of NWseg, ResNet50-FCN, LRASPP, and U-Net models on the training set,' which describes the performance of these models on a night-time flood image dataset rather than data obtained through physical field measurements. Additionally, the experimental data in this study are derived from an annotated surveillance image dataset rather than on-site physical measurements, aiming to assess the models' capability in identifying flood-affected areas. This aligns with the design objective of the NWseg model, which is to perform semantic segmentation based on surveillance imagery.

Line 252: what is "social inundation"?

Response: We sincerely appreciate the reviewer's careful examination and inquiry regarding the wording in Line 252. We have replaced “social inundation dataset” with “*nighttime flood inundation dataset*”.

Line 254: "exceptional" is too strong a word here.
Response: Thanks to your suggestion, we have replaced “exceptional” with “*superior*”.

Line 285: these images, and later one, are too small.
Response: Thanks to your suggestion, we have adjusted the sizes of Figures 5 and 6.

Line 328: if the conclusions state that there is a high computational demand, this should be investigated and reported on in the results section. How much greater is it? How long did it take? What sort of computer was used? Does this allow for practical use of NWseg?
Response: Thanks to your suggestion, we have added the description about the actual operating efficiency of the model in the new section “5 Discuss” as follows:

“*In addition, NWseg achieves an inference speed of 37.8 FPS (i.e., approximately 26.5 milliseconds per image) under the NVIDIA GeForce RTX 3080 environment, demonstrating its potential for real-time applications in high-performance computing platforms.*”