Nat. Hazards Earth Syst. Sci., 25, 4361–4373, 2025
https://doi.org/10.5194/nhess-25-4361-2025

Natural Hazards and Earth System Sciences
Open Access
EGU

Identification of nighttime urban flood inundation extent using deep learning

Jiaquan Wan[1,2,3], Xing Wang[4], Yannian Cheng[5], Cuiyan Zhang[4], Fengchang Xue[5], Tao Yang[1,2,3], Fei Tong[6], and Quan J. Wang[7]

[1]College of Hydrology and Water Resources, Hohai University, Nanjing, 210098, China
[2]The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing, 210098, China
[3]Yangtze Institute for Conservation and Development, Hohai University, Jiangsu, 210098, China
[4]School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China
[5]School of Remote Sensing and Surveying Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China
[6]China Institute of Water Resources and Hydropower, Beijing, 100048, China
[7]Department of Infrastructure Engineering, Faculty of Engineering and Information Technology, The University of Melbourne, Melbourne, Victoria 3010, Australia

Correspondence: Xing Wang (jwangxing0719@163.com)

Received: 8 January 2025 – Discussion started: 28 January 2025
Revised: 31 July 2025 – Accepted: 12 September 2025 – Published: 5 November 2025

Abstract. With the acceleration of urbanization, the disaster of urban flooding has had a serious impact on urban socio-economic activities and has become one of the important factors restricting social development in China. Accurate and timely identification of urban flooding extents is crucial for decision-making. Traditional remote sensing technologies are often limited by environmental factors, making them less suitable for application in complex urban terrains. With the increase in urbanization and the development of emerging technologies, video imagery has become a significant data source with great potential for urban flood identification. However, existing research has primarily focused on flood extent identification in daytime scenarios, often neglecting the nighttime, a period of high flood occurrence. In this study, we propose an efficient model (NWseg) to identify flood extents in nighttime scenes. Initially, we constructed a nighttime flood inundation dataset consisting of 4000 images. Subsequently, MobilenetV2 and ResNet101 networks were used to replace the DeepLabv3+ backbone network and compared with the NWseg model. Next, the NWseg model was compared with ResNet50-FCN, LRASPP, and U-Net models to evaluate the performance of different models in nighttime urban flooding extent identification. Finally, we verified the applicability and performance differences of each model in specific environments. Overall, this study successfully demonstrates the effectiveness of the NWseg model for nighttime urban flooding extent identification, providing new insights for nighttime flood monitoring in cities.

1 Introduction

In recent years, extreme rainfall events have been occurring frequently in the context of complex climate change (Burn and Whitfield, 2023; Kim et al., 2024). Concurrently, with the acceleration of urbanization processes, the proportion of impervious surfaces has been continuously expanding, resulting in serious urban flooding issues in many cities worldwide (Ghosh et al., 2024; Liu et al., 2023; Kundzewicz et al., 2019). Urban flooding often coincides with multiple compounded disasters and may even trigger secondary disasters, posing serious threats to the safety of urban residents, the normal operation of city functions, and sustainable development. This exacerbates the vulnerability of urban socio-economic systems (Gu et al., 2025; Visser, 2014; Zheng et al., 2014). Therefore, achieving real-time and effective identification of urban flooding extent has become a critical issue that urgently needs to be addressed.

Remote sensing technology has made significant advancements in the field of urban flood identification, providing new perspectives for flood disaster identification through high spatial, temporal, and spectral resolution data (Bofana et al., 2022). However, despite its excellent performance at the macro scale, remote sensing technology has limitations in urban area monitoring. 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 identification of rapidly evolving flood events challenging (Mason et al., 2012). In addition, the complexity of the urban environment, especially the dynamic changes of small-scale water bodies and localized waterlogging, further increases the difficulty of remote sensing technology in urban flooding extent identification. Therefore, an intelligent and real-time urban flood monitoring method is urgently needed to achieve more precise flood identification.

With technological advancements, the emerging fields of deep learning and computer vision have matured and engaged in interdisciplinary collaborations, achieving significant performance that offers new technical approaches for urban flood identification (Choi and Yoo, 2023). Particularly in image segmentation, deep learning's advantages in extracting global features and contextual information make it highly promising for inundation detection (Liu et al., 2020). Simultaneously, the increasing level of urbanization has led to the widespread deployment of video surveillance devices across urban areas, particularly in highly urbanized areas (Muhadi et al., 2024; Hao et al., 2022). During rainfall, these cameras can fully record the flooding process, providing real-time reflections of road inundation changes (Wang et al., 2024a). Therefore, combining deep learning with traffic cameras can effectively identify the extent of urban flooding.

Existing research has demonstrated that deep learning excels in segmenting inundated areas. Sarp et al. (2020) applied the Mask R-CNN model to automatically detect and segment floodwaters in urban, suburban, and natural scenes, achieving 99 % accuracy in the detection phase and 93 % in the segmentation phase. Sazara et al. (2019) used a deep learning approach to detect standing water on urban roads, in which a pre-trained VGG-16 network was used in the classification phase and a full convolutional neural network was used in the segmentation task, and compared it with the traditional classifier and extraction algorithms with manually-designed features, and the results showed that the deep learning approach has a more obvious advantage in both the recognition and segmentation of standing water. Wang et al. (2024a) used a deep convolutional neural network (DCNN) for urban flood extent recognition based on video images acquired from surveillance cameras. Zeng et al. (2024) proposed a DeepLabv3+ based flood image recognition method, which effectively improves the model performance through image enhancement and the introduction of the super-resolution generative adversarial network. However, current research focuses on daytime scenes, and the existing datasets lack diversity to cover flooding scenes at night or under complex weather conditions. Meanwhile, some algorithms underperform when processing images in low-light or adverse conditions, making flood extent identification at night or in challenging weather a technical challenge. This limitation underscores the urgent need for accurate nighttime flood extent identification and the necessity for algorithm improvements and dataset expansion.

To address the above challenges, this study proposes an efficient method for nighttime urban flood extent identification. First, an urban flood inundation dataset for nighttime scenes is constructed to provide sufficient sample support for model training. Subsequently, a NWseg model for nighttime image segmentation is proposed, which combines a Content-Light Splitter with a Dual-Feature Integrator to enhance the model's performance in identifying flooding extent in low-light environments. 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. Finally, the robustness and performance advantages of the NWseg model in nighttime urban flood recognition are verified through experimental comparison with mainstream segmentation models. This study not only promotes the development of nighttime urban flood recognition technology but also provides theoretical support and practical experience for future deep learning research in low-light environments using in nighttime low-light environments.

Totally, the main contributions of this paper are 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 Deeplabv3+ 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.

2 Model

2.1 Nighttime Urban Segmentation Model

Flood segmentation faces significant challenges in nighttime scenes. Insufficient illumination and interference from complex artificial light sources, such as streetlights and headlights, result in blurring of texture, edge, and color information in flooded regions, further exacerbating the difficulty of the segmentation task and severely affecting the robustness and accuracy of the model. To address this challenge, this study proposed a flood extent recognition model specifically designed for low-light nighttime scenes – the NWseg model, which aims to alleviate the impact of low illumination and complex lighting conditions on segmentation performance. As shown in Fig. 1, the NWseg model consists of two key modules: Content-Light Splitter (CLS) and Dual-Feature Integrator (DFI).

The design of the CLS module is based on the Retinex theory, which states that an image can be decomposed into a pixel-by-pixel product between a light-independent reflectance component (reflectance) and a light-related illumination component (illumination) (Land, 1977). Based on this principle, the CLS module decomposes the night image into a "reflectance map" and an "illumination map", which represent the inherent semantic information of the flood area and the lighting distribution in the scene, respectively (Wei et al., 2023). Subsequently, a semantic guidance mechanism is introduced to optimize the semantic segmentation loss during training (i.e., the difference between predicted pixel-level class labels and true labels), enabling the reflectance map to learn clearer boundaries and stronger semantic expression, thereby achieving accurate identification of the true contours of the flood areas. In addition, to addressing the interference from artificial light sources (such as car headlights and traffic lights), NWseg further designs the DFI module to enhance segmentation performance by adaptively fusing reflectance and illumination features. The DFI module first encodes the reflectance and illumination features and then constructs an attention mechanism that learns the degree of dependency between each pixel and the two feature types, enabling adaptive feature-weighted fusion at the pixel level (Li et al., 2024). This process adopts a pixel-wise weighting strategy, effectively enhancing the model's ability to recognize light-dominated categories. Finally, the DFI module introduces a dual semantic supervision mechanism: it not only applies semantic segmentation supervision to the fused output but also imposes semantic loss on the illumination channel separately, to enhance its independent discriminative ability and improve the model's overall generalization capability (Wei et al., 2023).

In summary, NWseg, through the collaborative design of the CLS and DFI modules, demonstrates superior semantic understanding and segmentation ability in complex nighttime lighting scenarios. It shows significant robustness and recognition advantages, particularly in high-reflection, low-contrast, and locally overexposed areas.

2.2 Typical semantic segmentation model

DeepLabv3+ is an advanced model in the field of image segmentation, which significantly improves the accuracy and detail processing ability of image segmentation by introducing an encoder-decoder structure (Bai et al., 2023; Peng et al., 2023). The encoder part is responsible for extracting the high-level features of the image, while the decoder focuses on recovering the details of the image, thus realizing a more fine-grained segmentation effect (Fu et al., 2021). The model also employs the techniques of void convolution and Atrous Spatial Pyramid Pooling (ASPP), which can effectively capture the multi-scale information of the image and improve the processing capability of complex scenes and object boundaries (Wang et al., 2024b; Peng et al., 2024). Cavity convolution enables the model to capture a larger range of image information without increasing the computational effort by introducing voids in the convolution kernel. It is particularly helpful in capturing the relationships between distant objects in an image (Yu et al., 2017). Atrous Spatial Pyramid Pooling (ASPP), on the other hand, enhances the recognition of objects of different sizes by using different scales of null convolution to extract multiple levels of image features, which helps the model to focus on both detailed and global information (He et al., 2014). In addition, DeepLabv3+ uses Xception as the backbone network, combined with depth-separable convolution to improve computational efficiency. Depth separable convolution divides the traditional convolution operation into two steps: first, each image feature is processed independently, and then the results are combined (Zhang et al., 2023). This approach effectively reduces computation and storage requirements, allowing the model to operate more efficiently while maintaining high accuracy. However, due to its relatively complex network structure, DeepLabv3+ is still slow in the inference stage.

To enhance the segmentation performance of DeepLabv3+ in urban flood scenes, 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 compares the results with the NWseg model. First, experiments were conducted on the original, unmodified DeepLabv3+ network as a baseline model. Then, we replaced the original DeepLabv3+ backbone network with the lightweight MobilenetV2, constructing an improved model (denoted as MobilenetV2-DeepLabv3+). MobilenetV2 is structurally optimized to compress the feature information while retaining the key information as much as possible, thus achieving a lightweight model while maintaining high accuracy (Jin et al., 2024). Finally, we replaced the backbone network of DeepLabv3+ with the residual neural network ResNet101 to form another improved model (denoted as ResNet101-

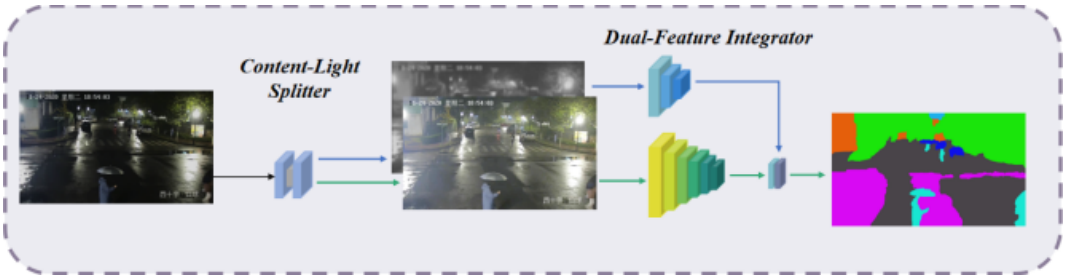

Figure 1. NWseg model inference process.

DeepLabv3+). ResNet101 adopts the residual connection mechanism so that part of the feature information can bypass the intermediate layer and be transmitted directly, which avoids the problems of "learning stagnation" or "training instability" during the training process (Wang et al., 2024b).

The Fully Convolutional Network (FCN) is a deep learning model that divides images into different regions by assigning a specific label to each pixel. Traditional deep learning models typically provide only an overall classification result for an entire image. In contrast, FCNs improve upon these models by replacing the fully connected layers with convolutional operations, enabling the network to handle input images of any size and produce detailed, pixel-level predictions (Yang et al., 2017). FCNs progressively compress the spatial dimensions of the image to extract essential information and then restore the original size to achieve precise localization of different regions (Zhao et al., 2018). Additionally, FCNs combine information from both shallow and deep layers, further enhancing segmentation accuracy in complex areas, such as flood boundaries. In this study, ResNet50 was selected as the backbone network for FCN, referred to as ResNet50-FCN. ResNet50 is a deep neural network that effectively alleviates the gradient vanishing problem during training, improving stability and efficiency. By combining the depth of ResNet50 with the flexibility of FCN, the proposed model enhances the accurate detection of inundated areas in complex environments.

LRASPP (Lightweight Refine Atrous Spatial Pyramid Pooling) is a lightweight model designed for image segmentation tasks. The model adopts MobileNetV3 as the backbone network for extracting the base features of an image and fuses shallow features to enhance the retention of detailed information. To further enhance the inference efficiency, LRASPP reduces the number of convolutional operations in the structural design and streamlines the feature channels to effectively reduce the computational complexity. Ultimately, the model restores the feature map to the same size as the input image through the upsampling operation to achieve accurate prediction of each pixel category (Tang et al., 2024).

U-Net is a deep learning model commonly used for image segmentation tasks, and its structure is mainly composed of two parts: encoder and decoder (Siddique et al., 2021). The encoder is responsible for gradually reducing the image size and extracting key features, while the decoder recovers the detailed information by gradually enlarging the feature map, thus realizing the accurate classification of each pixel. In addition, to compensate for the information lost during the process of reducing the image size, U-Net introduces a jump-join mechanism, which passes the features extracted at different stages in the encoder directly to the corresponding decoder stage (Sengupta et al., 2025). This design enables the model to better preserve the detailed features in the image while maintaining overall semantic understanding (Yadavendra et al., 2022).

We conducted comparative experiments on the FCN, LRASPP, U-Net, and NWseg models, evaluating their performance using metrics such as Precision, Recall, Mean Intersection over Union (MIoU), and F1 Score. All models were initialized with pretrained weights for their backbone networks and trained on the nighttime urban flooding dataset. The models were then evaluated on the test set, with relevant metrics calculated to determine the most suitable model for nighttime urban flood recognition.

3 Design of experiments

3.1 Construction of dataset

In this study, we employed web crawler technology using Google Chrome to construct a comprehensive nighttime urban waterlogging dataset by searching with the keyword "nighttime urban flooding". This dataset contains 4000 images that capture a wide range of nighttime waterlogging scenes, varying in extent and shape. To enhance the dataset's robustness and comprehensiveness, we included images of complex scenes, such as strong lighting conditions and splashes caused by vehicles, ensuring its applicability to diverse nighttime flooding situations. During the data selection process, careful attention was given to the representativeness and balance of waterlogged areas across different scales, ranging from localized ponding to large-scale flood events, to ensure broad coverage of possible urban flooding conditions (Du et al., 2025).

Figure 2. Samples and the flood area labels in the dataset, the white marked range is the flood extent.

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 (Zhang et al., 2023). Using this tool, we precisely labeled the inundated areas in a total of 4000 images. To ensure the accuracy and consistency of the annotations, three graduate students with research backgrounds in urban flooding were recruited to independently perform the annotation work. 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. This process ensured the overall quality and reliability of the dataset. Figure 2 presents a comparison between the original images and the labeled images, where the inundated areas are marked in white and the non-inundated areas are marked in black.

3.2 Evaluation metrics

In validation and testing, mean Intersection over Union (MIoU), $F1$ score, precision, and recall were used to assess the performance of the semantic segmentation models (Munawar et al., 2021).

The MIoU value is defined as the ratio of the intersection area of the predicted bounding box and the real bounding box to the concatenation area, and is calculated by averaging the results for each category. It is used to evaluate the accuracy of the location information of the predicted results of the target detection task. The larger the overlap area between the real and the presumed area of the object, the larger the calculated value of MIoU, and the more accurate the presumed target area. The calculation of the MIoU value follows the following formula:

$$\mathrm{MIoU} = \frac{1}{k+1} \sum_{i=0}^{k} \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

Precision, which is the proportion of samples predicted to be positive that are actually positive, is also known as the check rate, and can be expressed by the following formula:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

Recall, which is the proportion of actual positive samples that are predicted to be positive, is also known as the check all rate, and is given by the following formula:

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$F1$ score is the reconciled mean of precision and recall. The formula for each precision evaluation metric is as follows:

$$F1\ \mathrm{score} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{4}$$

In the above formula, TP is the number of actual situations that are true and predicted to be true; FP is the number of actual situations that are false and predicted to be true; FN is the number of actual situations that are true and predicted to be false; and TN is the number of actual situations that are false and predicted to be false.

Table 1. Configuration table of the experiment.

Project	Model
Operating System	Windows 10
Programming Language	Python 3.13
GPU	NVIDIA GeForce RTX3080
GPU memory	32 GB

3.3 Experimental configuration

All experiments were conducted using an operating system of Windows 10, a CPU model of Intel(R) Core(TM) i712700F@2.10 GHz, a GPU model of NVIDIA GeForce RTX 3080, 32GB of operating memory, a programming language of Python 3.13, and a deep learning framework of PyTorch1.13, GPU acceleration is enabled during model training with CUDA 11.7 and cuDNN 8.4.1 to improve training efficiency. The input image resolution is 512×512 pixels, the training optimizer type is Adam, the weight decay index is 0.0001, and the initialized learning rate is 0.005. Parameters are shown in Table 1.

4 Result

4.1 Ablation study

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 Fig. 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's performance improved even more, 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.

4.2 Model performance experiments

In this section, we present a comparative analysis of the experimental results of the NWseg model against other segmentation models. As shown in Table 3 and Fig. 4, the NWseg model achieved optimal results on the test set of the nighttime flood inundation dataset, with a Precision of 95.99 %, Recall of 94.8 %, $F1$ score of 95.39 %, and MIoU of 91.46 %. These metrics are significantly higher than those of the other models, demonstrating superior accuracy and recall rates. Compared to the ResNet50-FCN model, the NWseg model exhibits superior performance across all indicators, with increases of 10.99 % in Precision, 17.57 % in Recall, 14.46 % in $F1$ score, and an 8.76 % improvement in MIoU. When compared with the U-Net model, while the NWseg's Precision is similar, it outperforms in other metrics, with Recall, $F1$ score, and MIoU higher by 11.23 %, 7.15 %, and 10.96 % respectively. Additionally, compared to the lightweight LRASPP model, the NWseg model shows more pronounced advantages, with Precision increased by 15.82 %, Recall significantly increased by 69.41 %, $F1$ score improved by 56.82 %, and MIoU enhanced by 32.25 %. Although NWseg is higher in the number of model parameters than the other comparative models, it still demonstrates significant advantages in several evaluation metrics. Future research will aim to further optimize the structure of the model while maintaining its performance to achieve a higher degree of lightweighting.

Overall, the NWseg model demonstrates superior performance across all evaluation metrics and also shows strong performance in real scenario tests. In contrast, although the ResNet50-FCN model performs well in precision and detail processing, it lacks efficacy in handling edge regions, leading to slightly insufficient performance in complex scenes. While LRASPP offers advantages in computational efficiency due to its lightweight design, it has limitations in the precise capture of details and boundaries. The U-Net model is comparable to NWseg in accurately detecting target areas but is somewhat less robust and consistent when processing complex scenes.

4.3 Real-world scenes prediction comparison

To validate the effectiveness and stability of each model under challenging scenes, we conducted tests on seven models using nighttime strong illumination scenes and nighttime rainfall scenes (Wan et al., 2025). As shown in Fig. 5a presents the original scene where streetlights at night generate strong reflections and halos on the water surface. Additionally, the intense lighting affects the detailed features of the ground. By comparing the recognition results of each model, it is evident that the NWseg, ResNet50-FCN, and U-Net models accurately detected the flooding conditions in the scene. Notably, the NWseg model exhibited a more refined recognition ability in identifying water accumulation in road depressions. However, both ResNet50-FCN and U-Net showed certain false detections when recognizing the overall flooded areas. In contrast, the Mobilenetv2-DeepLabv3+, DeepLab, and LRASPP models could only sporadically identify small flooded regions and exhibited varying degrees of false detections. Although the ResNet101-DeepLabv3+ model recognized a larger flooded area, a comparison with the original image reveals a relatively high false detection rate, indicating deviations in prediction accuracy. Overall, the

Table 2. NWseg and DeepLabv3+ series model training results.

Models	*P* (%)	*R* (%)	*F*1 score (%)	MIoU (%)	Params (M)
Mobilenetv2-DeepLabv3+	67.46	50.64	57.85	46.15	5.81
ResNet101-DeepLabv3+	67.74	57.24	62.05	51.98	59.34
DeepLabv3+	53.34	50.61	51.94	46.07	54.70
NWseg	95.99	94.8	95.39	91.46	122.6

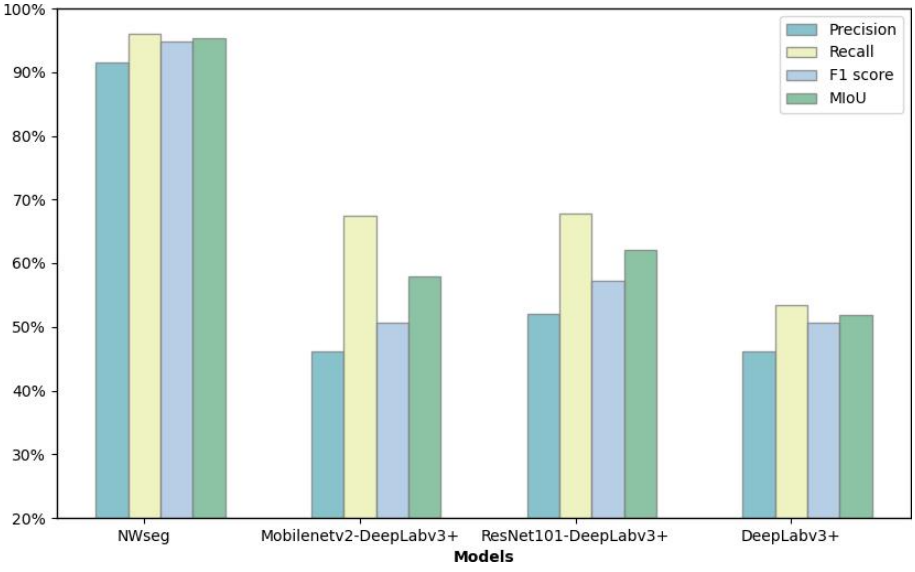

Figure 3. Comparison of experimental results between NWseg and DeepLabv3+ series of models.

Table 3. NWseg and other segmentation model training results.

Models	*P* (%)	*R* (%)	*F*1 score (%)	MIoU (%)	Params (M)
NWseg	95.99	94.8	95.39	91.46	122.6
ResNet50-FCN	85	77.23	80.93	82.7	35.31
LRASPP	80.17	25.39	38.57	59.21	3.22
U-Net	94.7	83.57	88.24	80.5	43.93

NWseg model outperformed the others in this scene recognition task, demonstrating superior capability in recognizing flooded areas under complex lighting conditions.

Furthermore, in the nighttime rainfall scene tests, we evaluated each model's performance to simulate urban flood recognition under real-world conditions. In such scenes, reflections from rainwater, slippery road surfaces, and interference from raindrops on the camera lens can adversely affect image clarity and the models' recognition accuracy (Zhao et al., 2025). As shown clearly in Fig. 6, the NWseg, ResNet50-FCN, and U-Net models were able to correctly identify the flooded areas in the images, with the NWseg model providing the most detailed performance by accurately capturing the edges of the flooded regions. While ResNet50-FCN and U-Net also identified the extent of flooding relatively well, they were somewhat insufficient in recognizing the flood boundaries and exhibited some false detections.

In contrast, the other four models performed relatively poorly. Specifically, the LRASPP and Mobilenetv2-DeepLabv3+ models were almost unable to detect the flooding, indicating weaker recognition capabilities in nighttime rainfall scenes. Although ResNet101-DeepLabv3+ and DeepLab could detect some flooded areas, comparison with the original images revealed that the regions identified did not accurately reflect the actual flooding conditions and had high false detection rates. Through comparative analysis, we further confirmed the challenges posed by nighttime rainfall environments for urban flood recognition and demonstrated the superior performance of the NWseg model in handling complex conditions such as nighttime rainfall.

5 Discussion

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

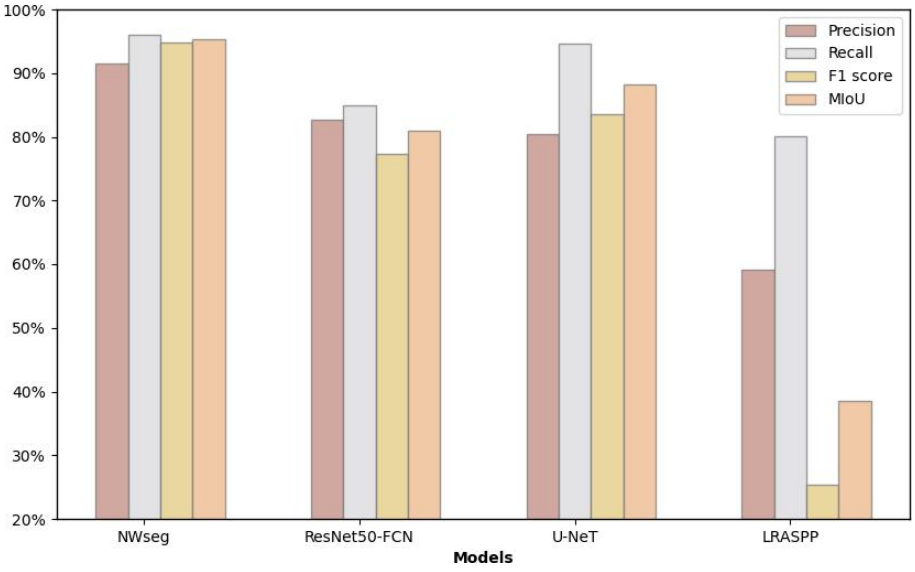

Figure 4. Comparison of experimental results between NWseg and other segmentation models.

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 ms 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, and its stability in complex scenarios needs to be further verified. Second, 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 addition, the current model is primarily designed for nighttime flood extent recognition and is not yet capable of sensing or estimating flood depth. It also lacks the ability to perform reliably under all-weather conditions. Furthermore, although the NWseg model can identify flooded areas more accurately, it is still difficult to achieve accurate modeling and area quantification of inundated areas. Finally, the dataset used in this study is mainly collected from some typical cities in China, and although it has covered diverse nighttime environments and lighting conditions, the model's generalization ability may be limited by the influence of geographic concentration and the dependence on surveillance cameras, and there is a certain identification bias when facing areas with different urban structures and lighting conditions.

In the future, the network structure will be further optimized to reduce the computational complexity of the model and improve the flexibility and efficiency of deployment. Meanwhile, the training dataset will be continuously expanded to enhance its diversity and representativeness in multiple dimensions, such as geographic distribution and urban structure. In addition, it is planned to introduce depth estimation technology to realize the accurate perception and quantification of the depth of the flood to meet more detailed flood monitoring needs. Finally, techniques such as infrared thermal imaging, microlight enhancement, and multimodal fusion will be combined to improve the robustness and adaptability of the model under extremely low light conditions. In subsequent practical applications, the NWseg model can be widely deployed in key scenarios such as urban emergency management, intelligent transportation monitoring, and disaster prevention and mitigation, especially for emergency response needs under extreme weather at night. By interfacing with existing traffic monitoring systems or urban sensing platforms, the model can automatically extract flooding information from the monitoring screen and realize rapid identification and early warning push for waterlogged areas. Combined with the city scheduling platform, NWseg can assist government departments in flood risk assessment, dynamic allocation of emergency resources, and trend analysis of disaster evolution, which significantly improves the efficiency of urban response and risk management capabilities in extreme weather events.

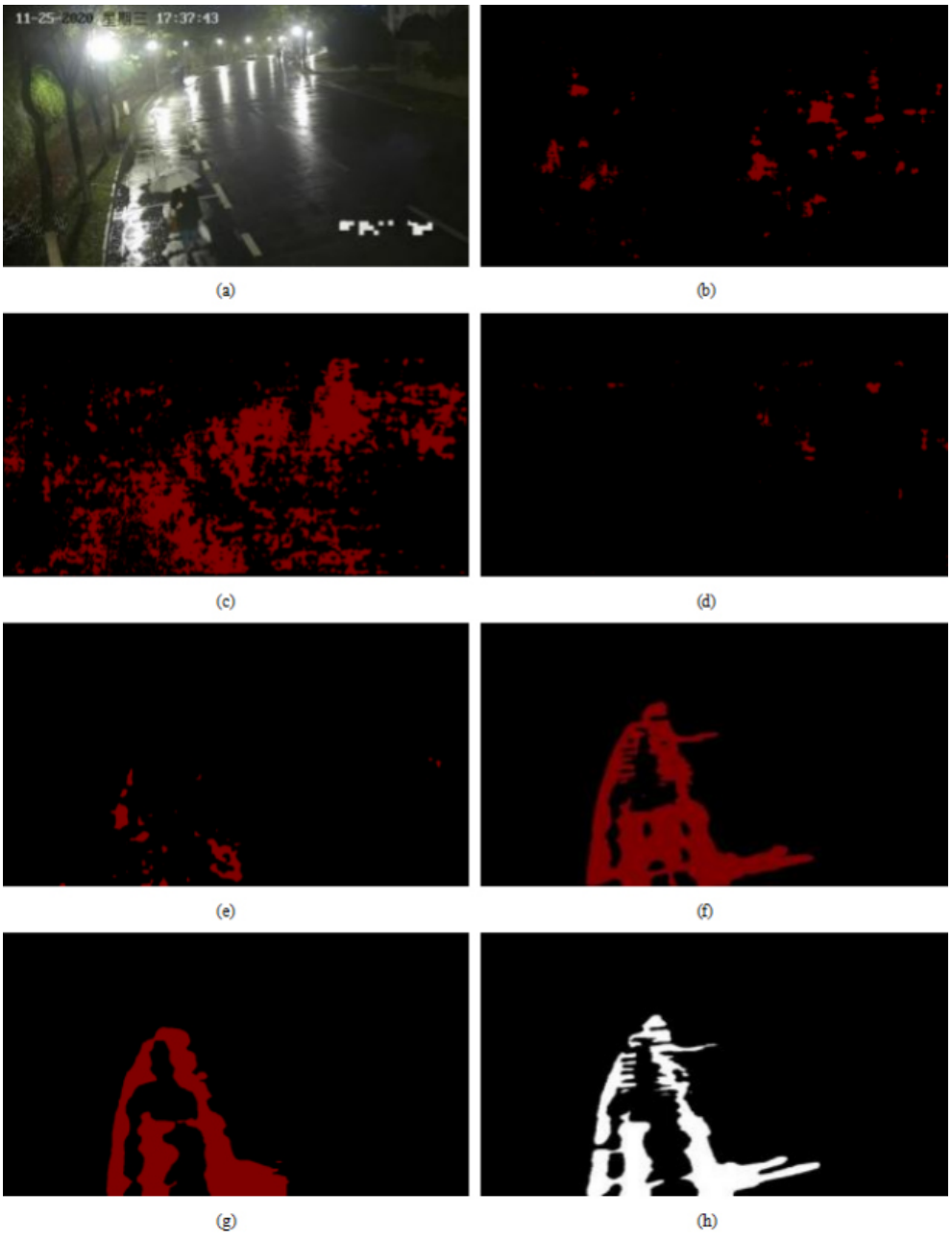

Figure 5. Scene with nighttime strong illumination: **(a)** the original scene; **(b)** the segmentation result of Mobilenetv2-DeepLabv3+; **(c)** the segmentation result of ResNet101-DeepLabv3+; **(d)** the segmentation result of DeepLabv3+; **(e)** the segmentation result of LRASPP; **(f)** the segmentation result of U-Net; **(g)** the segmentation result of ResNet50-FCN; **(h)** the segmentation result of NWseg.

6 Conclusions

This study successfully verified the excellent performance of the NWseg model in nighttime urban flood monitoring (Wan et al., 2024), which provides a new idea for multi-scene flood extent identification and helps to promote the flood monitoring system towards all-weather and all-scene intelligent identification. First, we constructed a representative dataset comprising 4000 images of nighttime urban flooding scenes, covering various nighttime environments and diverse urban backgrounds. Second, a model for nighttime waterlogging recognition, NWseg, is proposed to address the limitations in nighttime waterlogging recognition due to insufficient lighting and complex lighting conditions. Furthermore, we replaced the backbone networks of the DeepLabv3+ model with MobilenetV2 and ResNet101 and conducted ablation experiments to validate the performance of DeepLabv3+ with different backbones in nighttime flood

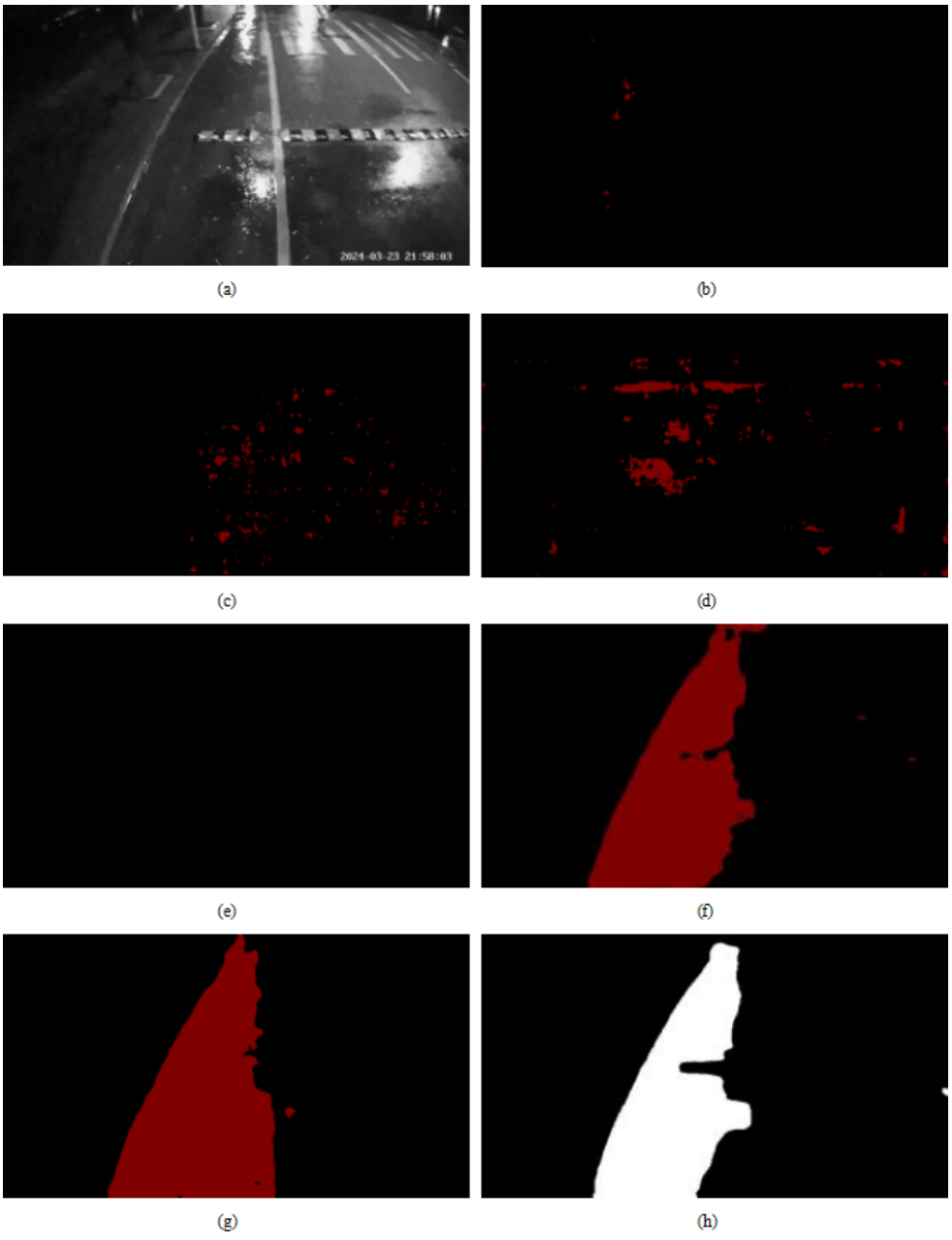

Figure 6. Scene with nighttime rainfall: **(a)** the original scene; **(b)** the segmentation result of Mobilenetv2-DeepLabv3+; **(c)** the segmentation result of ResNet101-DeepLabv3+; **(d)** the segmentation result of DeepLabv3+; **(e)** the segmentation result of LRASPP; **(f)** the segmentation result of U-Net; **(g)** the segmentation result of ResNet50-FCN; **(h)** the segmentation result of NWseg.

recognition. We also performed a comparative analysis between these DeepLabv3+ models and the NWseg model, as well as systematically analyzed the NWseg, ResNet50-FCN, U-Net, and LRASPP models. Based on this, we reached the following empirical findings:

1. Within the DeepLab series, the DeepLabv3+ model using ResNet101 as the backbone outperformed other variants in capturing water surface edges and shadow details. However, when compared to the NWseg model, there remains a considerable performance gap.

2. The NWseg, U-Net, and ResNet50-FCN models demonstrated excellent performance in recognizing large-scale flooded areas, effectively capturing the overall contours of flood zones and exhibiting strong generalization capabilities. Specifically, NWseg shows higher accuracy and robustness in complex scene tests, while

ResNet50-FCN and U-Net have some deficiencies and false detections in detecting edge details. In contrast, the lightweight LRASPP model showed limited ability to recognize flooded areas in nighttime scenes, resulting in relatively poor performance.

3. Through examining each model in complex scenes, we validated the NWseg model's effectiveness and stability in diverse environments and conditions.

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 (Wan et al., 2024). However, the model's decoupling and parsing process involves complex decomposition of lighting components and adaptive fusion, leading to high computational resource demands, which may impact its practical usability. Future work will focus on reducing the model's parameters and computational costs while maintaining accuracy. Additionally, further optimization of the dataset and model improvements will be pursued to enhance the overall performance of the NWseg model, broadening its potential applications.

Data availability. Data will be made available on request.

Author contributions. XW, JW, YC, CZ: Writing – original draft, Validation, Software, Methodology, Investigation. XW, JW, YC: Writing – review & editing, Validation. TY: Writing – review & editing, Supervision. FX: Formal analysis, Validation. FT: Data curation, Validation. QW: Data curation, Validation.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Also, please note that this paper has not received English language copy-editing. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Financial support. This research has been supported by the National Natural Science Foundation of China (NSFC) (grant nos. 42405140 and 52579003), the China Postdoctoral Foundation (grant no. 2024M761383), the Open Grants of China Meteorological Administration Radar Meteorology Key Laboratory (grant no. 2024LRMA02), and the Talent Startup project of NJIT (YKJ. 202315), and the Scientific Research Project of China Yangtze Power Co., Ltd. (grant no. Z242302064).

Review statement. This paper was edited by Lindsay Beevers and reviewed by Laurent pascal Dieme and one anonymous referee.

References

Bai, Y., Li, J., Shi, L., Jiang, Q., Yan, B., and Wang, Z.: DME-DeepLabV3+: a lightweight model for diabetic macular edema extraction based on DeepLabV3+ architecture, Frontiers in Medicine, 10, 11, https://doi.org/10.3389/fmed.2023.1150295, 2023.

Bofana, J., Zhang, M., Wu, B., Zeng, H., Nabil, M., Zhang, N., Elnashar, A., Tian, F., Da Silva, J. M., Botão, A., Atumane, A., Mushore, T. D., and Yan, N.: How long did crops survive from floods caused by Cyclone Idai in Mozambique detected with multi-satellite data, Remote Sens. Environ., 269, 112808, https://doi.org/10.1016/j.rse.2021.112808, 2022.

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

Choi, Y. H. and Yoo, S. J.: Quantized-state-based decentralized neural network control of a class of uncertain interconnected nonlinear systems with input and interaction time delays, Eng. Appl. Artif. Intel., 125, 106759, https://doi.org/10.1016/j.engappai.2023.106759, 2023.

Du, W., Qian, M., He, S., Xu, L., Zhang, X., Huang, M., and Chen, N.: An improved ResNet method for urban flooding water depth estimation from social media images, Measurement, 242, 12, https://doi.org/10.1016/j.measurement.2024.116114, 2025.

Fu, H., Meng, D., Li, W., and Wang, Y.: Bridge Crack Semantic Segmentation Based on Improved Deeplabv3+, Journal of Marine Science and Engineering, 9, 14, https://doi.org/10.3390/jmse9060671, 2021.

Ghosh, P., Sudarsan, J. S., and Nithiyanantham, S.: Nature-Based Disaster Risk Reduction of Floods in Urban Areas, Water Resources Management, 38, 1847–1866, https://doi.org/10.1007/s11269-024-03757-4, 2024.

Gu, Q., Chai, F., Zang, W., Zhang, H., Hao, X., and Xu, H.: A Two-Level Early Warning System on Urban Floods Caused by Rainstorm, Sustainability, 17, 16, https://doi.org/10.3390/su17052147, 2025.

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, 1799–1812, https://doi.org/10.1007/s11269-022-03107-2, 2022.

He, K., Zhang, X., Ren, S., and Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition, in: Proceedings of the 13th European Conference on Computer Vision (ECCV), 8691, 346–361, https://doi.org/10.1007/978-3-319-10578-9_23, 2014.

Jin, K., Zhang, J., Wang, Z., Zhang, J., Liu, N., Li, M., and Ma, Z.: Application of deep learning based on thermal im-

ages to identify the water stress in cotton under film-mulched drip irrigation, Agricultural Water Management, 299, 108901, https://doi.org/10.1016/j.agwat.2024.108901, 2024.

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

Kundzewicz, Z. W., Su, B., Wang, Y., Xia, J., Huang, J., and Jiang, T.: Flood risk and its reduction in China, Adv. Water Resour., 130, 37–45, https://doi.org/10.1016/j.advwatres.2019.05.020, 2019.

Land, E. H.: The retinex theory of color vision, Scientific American, 237, 108–128, https://doi.org/10.1038/scientificamerican1277-108, 1977.

Li, X., Wang, W., Feng, X., and Li, M.: Deep parametric Retinex decomposition model for low-light image enhancement, Computer Vision and Image Understanding, 241, 14, https://doi.org/10.1016/j.cviu.2024.103948, 2024

Liu, W., Feng, Q., Engel, B. A., Yu, T., Zhang, X., and Qian, Y.: A probabilistic assessment of urban flood risk and impacts of future climate change, Journal of Hydrology, 618, 11, https://doi.org/10.1016/j.jhydrol.2023.129267, 2023.

Liu, X., Song, L., Liu, S., and Zhang, Y.: A review of deep-learning-based medical image segmentation methods, Sustainability, 13, 1224, https://doi.org/10.3390/su13031224, 2020.

Mason, D. C., Davenport, I. J., Neal, J. C., Schumann, G. J. P., and Bates, P. D.: Near Real-Time Flood Detection in Urban and Rural Areas Using High-Resolution Synthetic Aperture Radar Images, IEEE Transactions on Geoscience and Remote Sensing, 50, 3041–3052, https://doi.org/10.1109/tgrs.2011.2178030, 2012.

Muhadi, N. A., Abdullah, A. F., Bejo, S. K., Mahadi, M. R., Mijic, A., and Vojinovic, Z.: Deep learning and LiDAR integration for surveillance camera-based river water level monitoring in flood applications, Natural Hazards, 120, 8367–8390, https://doi.org/10.1007/s11069-024-06503-6, 2024.

Munawar, H. S., Ullah, F., Qayyum, S., and Heravi, A.: Application of deep learning on UAV-based aerial images for flood detection, Smart Cities, 4, 1220–1242, https://doi.org/10.3390/smartcities4030065, 2021.

Peng, H., Zhong, J., Liu, H., Li, J., Yao, M., and Zhang, X.: ResDense-focal-DeepLabV3+ enabled litchi branch semantic segmentation for robotic harvesting, Computers and Electronics in Agriculture, 206, 12, https://doi.org/10.1016/j.compag.2023.107691, 2023.

Peng, H., Xiang, S., Chen, M., Li, H., and Su, Q.: DCN-Deeplabv3+: A Novel Road Segmentation Algorithm Based on Improved Deeplabv3+, Ieee Access, 12, 87397–87406, https://doi.org/10.1109/access.2024.3416468, 2024.

Sarp, S., Kuzlu, M., Cetin, M., Sazara, C., and Guler, O.: Detecting floodwater on roadways from image data using Mask-R-CNN, 2020 International Conference on INnovations in Intelligent SysTems and Applications, Novi Sad, Serbia, 24–26, https://doi.org/10.1109/INISTA49547.2020.9194655, 2020.

Sazara, C., Cetin, K., and Iftekharuddin, K.: Detecting floodwater on roadways from image data with handcrafted features and deep transfer learning, 2019 IEEE Intelligent Transportation Systems Conference, Auckland, New Zealand, 27–30, https://doi.org/10.1109/ITSC.2019.8917368, 2019.

Sengupta, S., Chyrmang, G., Bora, K., Das, H. S., Li, A., Lemos, B., and Mallik, S.: Assessment of different U-Net backbones in segmenting colorectal adenocarcinoma from H&E histopathology, Pathology Research and Practice, 266, 11, https://doi.org/10.1016/j.prp.2025.155820, 2025.

Siddique, N., Paheding, S., Elkin, C. P., and Devabhaktuni, V.: U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications, IEEE Access, 9, 82031–82057, https://doi.org/10.1109/access.2021.3086020, 2021.

Tang, Y., Tan, D, Li, H., Zhu, M., Li, X., Wang, X., Wang, J., Wang, Z., Gao, C., Wang, J., and Han, A.: RTC_TongueNet: An improved tongue image segmentation model based on DeepLabV3, Digital Health, 10, 20552076241242773, https://doi.org/10.1177/20552076241242773, 2024.

Visser, F.: Rapid mapping of urban development from historic Ordnance Survey maps: an application for pluvial flood risk in Worcester, J. Maps., 10, 276–288, https://doi.org/10.1080/17445647.2014.893847, 2014.

Wan, J., Qin, Y., Shen, Y., Yang, T., Yan, X., Zhang, S., Yang, G., Xue, F., and Wang, Q.: Automatic detection of urban flood level with YOLOv8 using flooded vehicle dataset, Journal of Hydrology, 639, 131625, https://doi.org/10.1016/j.jhydrol.2024.131625, 2024.

Wan, J., Xue, F., Shen, Y., Song, H., Shi, P., Qin, Y., Yang, T., and Wang, Q. J.: Automatic segmentation of urban flood extent in video image with DSS-YOLOv8n, Journal of Hydrology, 655, 12, https://doi.org/10.1016/j.jhydrol.2025.132974, 2025.

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, https://doi.org/10.1016/j.envsoft.2023.105939, 2024a.

Wang, Y., Yang, L., Liu, X., and Yan, P.: An improved semantic segmentation algorithm for high-resolution remote sensing images based on DeepLabv3+, Scientific Reports, 14, 15, https://doi.org/10.1038/s41598-024-60375-1, 2024b.

Wei, Z., Chen, L., Tu, T., Ling, P., Chen, H., and Jin, Y.: Disentangle then Parse: Night-time Semantic Segmentation with Illumination Disentanglement, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 21593–21603, https://doi.org/10.48550/arXiv.2307.09362, 2023.

Yadavendra and Chand, S.: Semantic segmentation of human cell nucleus using deep U-Net and other versions of U-Net models, Network-Computation in Neural Systems, 33, 167–186, https://doi.org/10.1080/0954898x.2022.2096938, 2022.

Yang, Y., Zhuang, Y., Bi, F., Shi, H., and Xie, Y.: M-FCN: Effective Fully Convolutional Network-Based Airplane Detection Framework, IEEE Geoscience and Remote Sensing Letters, 14, 1293–1297, https://doi.org/10.1109/lgrs.2017.2708722, 2017.

Yu, F., Koltun, V., and Funkhouser, T.: Dilated residual networks, in: Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 636–644, https://doi.org/10.1109/cvpr.2017.75, 2017.

Zeng, Y., Chang, M., and Lin, G.: A novel AI-based model for real-time flooding image recognition using super-resolution generative adversarial network, Journal of Hydrology, 638, 12, https://doi.org/10.1016/j.jhydrol.2024.131475, 2024.

Zhang, L., Han, G., Qiao, Y., Xu, L., Chen, L., and Tang, J.: Interactive Dairy Goat Image Segmentation for Precision Livestock Farming, Animals, 13, 17, https://doi.org/10.3390/ani13203250, 2023.

Zhao, J., Wang, X., Zhang, C., Hu, J., Wan, J., Cheng, L., Shi, S., and Zhu, X.: Urban Waterlogging Monitoring and Recognition in Low-Light Scenarios Using Surveillance Videos and Deep Learning, Water, 17, 19, https://doi.org/10.3390/w17050707, 2025.

Zhao, W., Zhang, H., Yan, Y., Fu, Y., and Wang, H.: A Semantic Segmentation Algorithm Using FCN with Combination of BSLIC, Applied Sciences-Basel, 8, 500, https://doi.org/10.3390/app8040500, 2018.

Zheng, F., Westra, S., Leonard, M., and Sisson, S. A.: Modeling dependence between extreme rainfall and storm surge to estimate coastal flooding risk, Water Resour. Res., 50, 2050–2071, https://doi.org/10.1002/2013WR014616, 2014.