Journal of Internet Services and Applications, 2026, 17:1, doi: 10.5753/jisa.2026.6670

Improving Image Segmentation in Adverse Conditions for Coastal Infrastructure Monitoring

Pedro V. Rubinstein [**Universidade Federal do Rio de Janeiro** | ***prubinstein@gta.ufrj.br***]
Ricardo L. Kauer [**Universidade Federal do Rio de Janeiro** | ***rickauer@gta.ufrj.br***]
Alexandre G. Cardeman [**Centro de Operações e Resiliência (COR)** | ***acardeman@gmail.com***]
Marcelo Abelheira [**Centro de Operações e Resiliência (COR)** | ***marceloabelheira@gmail.com***]
Pedro Cruz [**Universidade Federal do Rio de Janeiro** | ***cruz@gta.ufrj.br***]
Rodrigo S. Couto [**Universidade Federal do Rio de Janeiro** | ***rodrigo@gta.ufrj.br***]
Luís Henrique M. K. Costa [**Universidade Federal do Rio de Janeiro** | ***luish@gta.ufrj.br***]

Centro de Tecnologia - 972, Av. Horácio Macedo, 2030 - Sala H-301 - Cidade Universitária da Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ, 21941-598, Brazil.

Received: 10 October 2025 • **Accepted:** 23 December 2025 • **Published:** 29 May 2026

Abstract

Smart cities increasingly rely on AI-driven solutions to improve citizen safety. In coastal regions such as Rio de Janeiro, Brazil, maritime risks present significant challenges, as exemplified by the collapse of the Tim Maia Bike lane. This paper proposes a method for monitoring the coastal infrastructure using a custom segmentation model based on YOLO, aiming to reduce the need for continuous human supervision. However, the external placement of cameras introduces lighting and weather-related challenges, which complicate accurate segmentation. To address these issues, we investigate how domain-specific data augmentation techniques affect model performance under adverse visual conditions. As a case study, we apply this method to develop a system for the Tim Maia Bike lane, with the improved models achieving 97.6% mAP50-95 throughout the day. Furthermore, we analyze the correlation between our system's outputs and environmental measurements obtained from an ocean buoy. Our findings highlight the potential for integrating AI-based monitoring into broader urban risk management frameworks to provide real-time protection for coastal infrastructure.

Keywords: Smart cities, urban risk management, computer vision, segmentation, data augmentation.

1 Introduction

The growing urbanization and challenges related to public safety have driven the development of different technologies for smart cities. In particular, the use of computer vision systems for monitoring critical infrastructure has proven essential for disaster prevention and ensuring rapid responses in emergency situations. A striking example of this demand occurred in the city of Rio de Janeiro, in April 2016, when strong waves caused the collapse of the Tim Maia Bike lane. This event highlighted the vulnerability of coastal urban environments and the urgency for more precise and continuous monitoring systems (Rio [2016]). In response, the city administration installed a camera strategically positioned toward the rocky shore beside the bike path, which enabled real-time monitoring of maritime conditions.

Despite the installation of such systems, it is not possible to rely solely on human operators to monitor camera feeds around the clock. The continuous observation of multiple video streams is both costly and prone to human oversight. In this context, computer vision algorithms can play a key role by automatically analyzing visual data and identifying potentially hazardous situations. These computer vision algorithms can be the core of early alert systems independent of human direct supervision.

This work focuses on developing a method for the automatic detection of maritime risks for urban infrastructure,

Figure 1. Raw image from monitoring camera number 4118 of the *Centro de Operações e Resiliência* (COR), used to monitor the coastal area near the site of the 2016 accident.

using the Tim Maia Bike lane (Figure 1) as a case study based on images captured by the monitoring cameras. To tackle this challenge, we design a computer vision system in partnership with the *Centro de Operações e Resiliência* (COR), a municipal center responsible for real-time urban integration. The system applies image segmentation techniques to estimate the wave height along the shoreline and generate risk alerts. Nonetheless, variations in lighting conditions, both daytime and nighttime, create significant challenges for image segmentation models. These same factors also make manual annotation harder, since noise, glare, and low visibility can obscure the contours of waves and other relevant features. On the other hand, the accuracy of these models plays a crucial

role in generating reliable warning signals and, consequently, improving the safety of bike path users and the population in general.

Our investigation shows that the application of data augmentation techniques, directly related to the studied domain, such as brightness adjustment, blur, and rain simulation, can significantly improve the generalization of segmentation models, enabling better adaptation between daytime and nighttime domains. To this end, we conduct an experimental analysis using datasets specific to each lighting condition, as well as a combined dataset. We validate the results through cross-validation. Our main contributions include four key areas. We develop a new monitoring method and system, analyzing how environmental variations affect segmentation model performance for risk detection on the bike path. We also demonstrate that augmentation techniques reduce performance disparities between daytime and nighttime imagery. Finally, we build an experimental approach which integrates multiple augmentation strategies to build more reliable models.

Unlike other studies that have applied YOLO and other convolutional neural networks (CNNs) to various tasks, such as vehicle Alam *et al*. [2025], fish Wang and Zhao [2024], and coastal phenomena detection Scardino *et al*. [2022], this work takes a different approach. We specifically investigate how lighting and weather conditions affect segmentation model performance in a limited-data context. We use datasets specific to daytime, nighttime, and combined images, and we apply data augmentation techniques such as brightness, blur, and rain simulation to deepen the understanding of how to adapt models for shoreline monitoring. This comparative approach evaluates model robustness across distinct domains, improves the understanding of model generalization throughout the day, and guides more effective efforts to enhance monitoring systems. As a result, we obtain several models with high mAP50-95, which is a comprehensive evaluation of object detection models, measuring precision by averaging the mean average precision (mAP) scores across ten different Intersection over Union (IoU) thresholds, ranging from 0.5 to 0.95 in increments of 0.05, reaching 97.6% accuracy (mAP50-95) throughout the day.

We summarize the contributions of this paper as follows:

- We develop a complete computer-vision-based monitoring method and system tailored for wave-induced risk detection on coastal infrastructure, validated in a real operational scenario.
- We show that it is possible to achieve high accuracy under adverse environmental conditions by combining a limited but carefully selected dataset with domain-specific data augmentation techniques, thereby accelerating learning without relying on large-scale annotation efforts.
- We provide a comparative analysis across daytime, nighttime, and mixed domains to characterize model robustness and generalization throughout the day, establishing a foundation for future studies that will contrast our representative-data strategy with approaches trained on larger datasets.

The paper is organized as follows. Section 2 reviews related work. Section 3 describes the application scenario of the models and the Tim Maia Bike lane case study. Section 4 explains the process of obtaining the dataset and details the experimental methodology used in this work. Section 5 presents and discusses the model results. Section 6 presents and discusses correlations found with other monitoring methods. Finally, Section 7 summarizes the conclusions and outlines directions for future work[1].

2 Related Work

Researchers have widely used real-time object detection and segmentation in the context of smart cities (Midigudla *et al*. [2025]; Herath and Mittal [2022]; Ullah *et al*. [2020]), successfully identifying and tracking elements in images and videos. In this scenario, approaches based on CNNs stand out, offering high performance in visual analysis. The literature uses several models, such as YOLO, MobileNet, and Detectron (Midigudla *et al*. [2025]; Yuldashev *et al*. [2023]; Talaat and ZainEldin [2023]). In this work, we adopt YOLO due to its ability to process images with high precision and efficiency, in addition to its ease of implementation. Applications of these methods include identifying vehicles, detecting suspicious actions, and monitoring natural phenomena (Liu [2024]; Kaur *et al*. [2023]; Talaat and ZainEldin [2023]).

Scardino *et al*. [2022] proposed a video monitoring mechanism using CNNs combined with optical flow techniques, which allows for estimating the apparent velocity of objects in an image sequence, to extract parameters related to waves and tides. Their study highlights the growing use of computer vision techniques to automate this analysis, reducing the need for human intervention. The approach employs cameras to capture images of the coastal environment, facilitating the detection of wave behavior patterns. Unlike our work, which focuses on the protection and monitoring of urban infrastructure, their study concentrates on obtaining oceanic meteorological parameters.

Muhadi *et al*. [2021] implemented a system based on CNNs to detect the water level in rivers using cameras. Their methodology uses two types of CNNs: *DeepLabV3+* and *SegNet*. When using semantic segmentation, the system uses two classes: water and background. The system classifies risk levels into different categories, allowing for the anticipation of critical events and assisting in the mitigation of environmental disasters. The authors evaluate the models using accuracy, intersection over union, and F1-contour score metrics. For our work, we opt to use YOLOv11, which stands out for its ease of implementation due to the ecosystem that Ultralytics provides (Jocher *et al*. [2023]). A key distinction is our focus on environmental challenges like lighting and weather, which we address by exploring the model's accuracy with various data augmentation techniques.

Wang and Zhao [2024] employ the YOLOv8-MSS algorithm, a modified version of YOLOv8, for detecting objects on the water surface. Their model features adjustments in data

[1] Part of this work is based on our paper published in Portuguese in the Proceedings of the IX Workshop de Computação Urbana (CoUrb) available at`https://sol.sbc.org.br/index.php/courb/article/view/35263`. This utilization is permitted by the Brazilian publisher, as seen in `https://sol.sbc.org.br/index.php/indice/conduta`.

processing and the loss function to refine detection accuracy. They evaluate its performance using mAP in comparison with the original YOLOv8 and other CNNs architectures. Martinho *et al*. [2023] use YOLO in conjunction with image restoration techniques to detect fish in marine environments. Their study addresses challenges in a low-visibility environment, using image restoration techniques to improve detection accuracy.

Although investigating similar aspects, the work of Alam *et al*. [2025] has a different application context. They explore the use of YOLOv5 for real-time vehicle detection at urban intersections. The training methodology they adopt involves separating the dataset into three categories: daytime, nighttime, and mixed images, which allows the model to adapt more effectively to different lighting conditions. Similarly, our work also employs the division of datasets into daytime and nighttime categories to evaluate the impact of lighting variations on the performance of segmentation models we apply to risk detection on the bike lane. This approach allows us to directly evaluate the performance of our models under different environmental conditions.

Table 1 summarizes the main characteristics of the related works and highlights how they differ from our solution. The comparison shows that, while previous studies address oceanographic analysis, object detection, or water-level and wave estimation, none of them explore the specific problem of coastal infrastructure protection under strong day–night variation. Our contribution lies in the combination of a segmentation-based estimation method, a domain-specific augmentation strategy designed for adverse visual conditions, and its validation in a real operational environment.

Overall, the literature demonstrates the use of CNN-based techniques across different scenarios, from underwater monitoring to urban video analysis. However, we found no studies targeting wave-induced risks on coastal slopes or proposing a methodology that integrates segmentation, environmental robustness, and practical deployment. Our approach infers risk from wave height relative to the hillside and focuses on challenges observed during the Tim Maia Bike Lane case study, such as lighting transitions, glare, rain, and blur. To address these issues, we employ data augmentation techniques, including Gaussian blur, brightness adjustment, rain simulation, and car headlight simulation, and evaluate their effect across three datasets: daytime, nighttime, and a combined set. This setup enables us to study the impact of lighting on model performance, identify effective pre-processing strategies for each scenario, and analyze generalization when models are tested outside their original domain.

3 Scenario

Coastal infrastructure close to the sea often faces risks, especially during periods of heavy swells, when waves can reach parts of the urban infrastructure. In this context, we present a computer vision–based monitoring system that detects and assesses the risk level associated with waves near public infrastructure. The system's main goal is to generate alerts about the risk level based on the height of the waves that reach the coastline.

3.1 General Description

The system uses video images of a structure on the hillside captured by a strategically positioned camera, as shown in Figure 1. Starting from these images, a YOLO-based segmentation model crops the region that contains the sea. Using manually or automatically annotated reference lines, the system is able to estimate how much the sea level rises and, as a result, whether the waves pose an imminent risk to the urban infrastructure.

Figure 2 shows a simplified diagram of the system, which we implemented in Python. The Video Capture module captures real-time video from a URL provided by the city hall and sends it to the Object Segmentation module, which receives the video at a rate of 15 frames per second. The Object Segmentation module is the core of the system, where a machine learning model performs a segmentation task to identify areas of interest within the image and return a mask containing these areas. The segmentation is used to separate the sea from the rest of the scene. The Rule Application module receives the mask and estimates the distance between the sea and the bike lane. The Rule Application module then persists the obtained data and, if the sea is considered too close to the bike lane, triggers a warning message which is sent by the Messaging module.

We train the YOLOv11n (i.e., "nano" version) segmentation model on images collected from city hall cameras to recognize visual patterns that characterize the sea as a whole. We annotate the data and generate the datasets using the open-source tool CVAT CVAT.ai Corporation [2023], chosen because of the ease of adjusting image parameters dynamically during annotation. The polygon representing the segmentation mask provides the basis for height estimates.

To estimate the wave height, the system analyzes the segmentation mask within an area of interest (in our case, the hillside and the bike path). It defines three main contours: a baseline curve that represents the sea level, an upper curve that marks the bike path, and a curve that defines the natural slope of the hill. Using the YOLO mask results, the system dynamically fits a line between the sea level and bike path curves, parallel to the slope curve, as illustrated in Figure 3. It expresses the wave height as a percentage of the hillside covered by the waves. This percentage is obtained by comparing the segment from sea level to a wave point with the segment from sea level to the bike path, and selecting the maximum value observed. The resulting value determines the risk level. Based on this level, the system applies rules to trigger alerts. The message-sending module uses the Telegram public API through the Aiogram library Onufriichuk [2025]. The system also stores all wave height data and alerts in a database for log-keeping through the Database Persistence module.

3.2 Challenges

Segmenting water bodies poses significant challenges, also because the cameras operate in an outdoor environment. These challenges motivate the data augmentation strategies we investigate throughout this work. To address these issues

Table 1. Comparison of key aspects among related works and the proposed solution.

Work	Adverse Condition Prevention	Day/Night Analysis	Wave/Water Level Estimation	Coastal / Shoreline Shoreline Focus
Scardino et al. (2022)	–	–	✓	✓
Muhadi et al. (2021)	–	–	✓	–
Wang & Zhao (2024)	–	–	✓	–
Martinho et al. (2023)	✓	–	–	–
Alam et al. (2025)	–	✓	–	–
This work	✓	✓	✓	✓

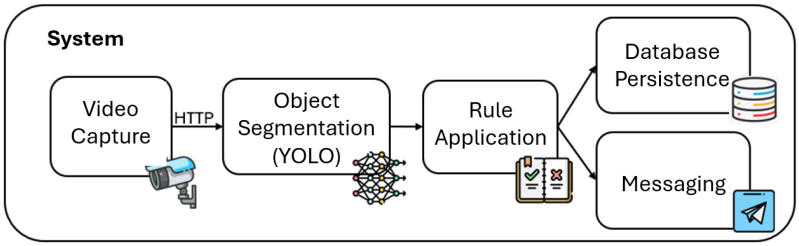


Figure 2. Diagram showing the operation of the developed system.

Figure 3. Example of lines used to compute the wave height based on the defined curves. The orange line represents the fitted segment, while the red line represents auxiliary lines.

and enhance the model's ability to handle diverse conditions, we will employ various data augmentation techniques. The main factors that we identified as a cause of model accuracy reduction include:

- **Lighting**: Buildings, vehicles, and urban infrastructure cause abrupt lighting changes, which may hinder the model's ability to generalize. In addition, during nighttime it becomes challenging to distinguish the water boundary.
- **Lens obstruction**: Dust particles and raindrops can impair visibility, smudging specific regions of the image. This produces distortions on familiar sea shapes and can confuse the model.
- **Fogging**: In coastal areas, high humidity and frequent condensation on the camera lens cause fogging, which effectively reduces the image's resolution across the whole image.

3.3 The Tim Maia Bike Lane

The Tim Maia Bike lane is particularly vulnerable to accidents due to its proximity to the sea, especially during big swells. This makes it an interesting case for deploying a risk detection system, while also highlighting the challenges involved in image capture. Among the most significant difficulties are the complexity of annotation and the system's reduced performance during the night, when identifying the sea contour becomes harder. Figure 4 illustrates real examples of these challenges.

We develop the system presented in this work, which has been in operation since May 2024, to analyze the percentage of the hillside covered by waves. The system classifies waves into distinct risk levels according to the percentage of the slope the sea covers. Currently, the system uses the following thresholds, which we chose empirically from our discussions with the technical staff of Rio's City Hall COR (*Centro de Operações e Resiliência*):

- **Very High**: The wave reached part of the bike path;
- **High**: The wave covered at least 80% of the slope;
- **Medium**: The wave covered from 60% up to 80% of the slope;
- **Low**: The wave covered 60% of the slope or less.

During operation, the system records videos corresponding to the triggered alerts and also stores false positives and the historical series of wave heights. Whenever the system identifies an alert, it sends a notification to the city hall through the Telegram messaging application Telegram [2025], as Figure 5 illustrates. Since its first prototypes, the system has sent alerts when it considers the risk level "high" or "very high". In addition to real-time alerts, the application generates periodic reports: control messages and summaries every hour, as well as a daily report containing the series of recorded heights and their respective risk levels throughout the monitoring period. Figure 6 shows an example of a daily chart produced by the system.

4 Methodology

Building and annotating datasets stands out as one of the greatest challenges for accurate segmentation in the computer vision system we develop for wave monitoring. The difficulties described in Section 3 make it hard for even a human observer to precisely identify the sea's contour in those situations. As a consequence, producing high-quality masks during annotation requires painstaking effort, which makes the use of auxiliary tools indispensable. To address this challenge, we apply temporary manual adjustments to

(a) Camera image with car headlights.

(b) Camera image under low light.

(c) Camera image affected by rain.

(d) Camera image with smudged lens.

Figure 4. Images of the challenges encountered in the Tim Maia Bike lane camera case study. All images are part of the datasets used in this work.

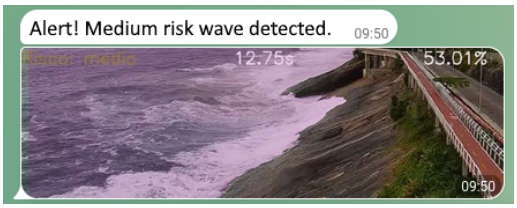


Figure 5. Example of an alert from the developed system, generated by a medium-risk wave. The sent image consists of a screenshot of the system's telegram group. The model's segmentation mask is overlaid in purple. In the top-center of the image, is the duration of the risk event in seconds. The number on the right indicates the wave height as a percentage of the monitored slope.

image parameters, such as brightness and contrast throughout the annotation process. While we annotate some images captured under less demanding conditions quickly, most of them require considerable time to complete.

The datasets we use consist of images that the camera captured pointing towards the Niemeyer Avenue coast in Rio de Janeiro. In the present work, we use images which we have stored between January and August 2024. We construct two main datasets, covering diverse weather conditions, such as rain, wind, sun, and cloudy skies, as well as variations in sea coloration and the amount of foam. During the initial testing phase, we annotate 500 daytime images to validate the method, on which we train the production model the system currently uses. For this work, we construct a second dataset of an equivalent size using the same process, but with nighttime images. We divide each dataset into two splits: 80% for training and 20% for testing. Due to the modest data availability, we use the "nano" version of YOLOv11 to avoid overfitting problems in the real system.

To improve the performance of the model, we apply data augmentation techniques, which we believe are fit to address the problems listed in Section 3.3. We choose the following techniques, which are also illustrated in Figure 7:

- **Random Gaussian blur:** Blurs the image by up to 2.5 pixels.
- **Brightness adjustment:** Randomly alters the brightness, increasing or decreasing it by up to 25%.
- **Rain simulation:** Randomly adds one to five blurred ellipses to the image to simulate the effect of raindrops on the camera lens.
- **Glow:** Inserts a point of light with a bloom effect to simulate the brightness of car headlights.

We test several data augmentation strategies, which involve applying each of the four techniques in isolation and in different combinations where multiple were applied at once. For each augmentation strategy, we modify only the training set. Applying these strategies resulted in generating multiple new images for each original image, which significantly expanded the size and diversity of our training datasets.

We separate the datasets into daytime and nighttime categories to investigate the impact of domain conditions on the models' performance given a limited data context. For comparison purposes, we implement an additional approach that involves constructing a complete dataset, comprising 250 images from the daytime dataset and another 250 from the nighttime dataset. We keep the total number of images constant to be consistent in the comparison between the trained models. We use *k-fold* cross-validation, a technique that divides the dataset into multiple *folds* (subsets), where, in each iteration, the process uses k−1 subsets for training and the remainder for validation. This process allows us to train and evaluate the model on different combinations of these subsets, reducing the risk of overfitting and providing a more reliable estimate of its performance. For this experiment, we adopt k equal to 5, which results in each subset containing 70 images.

We select the images to compose the complete dataset, which includes both daytime and nighttime images, based on the validation results we obtained for each *fold*. We evaluate the quality of the masks the models obtained by using mAP50-95. The mAP measures the average precision across different intersection over union (IoU) thresholds, ranging from 0.50 to 0.95, and provides an indicator of how accurately the model delineates the regions of interest. A higher mAP value indicates better segmentation performance.

Based on our cross-validation results, we select the valida-

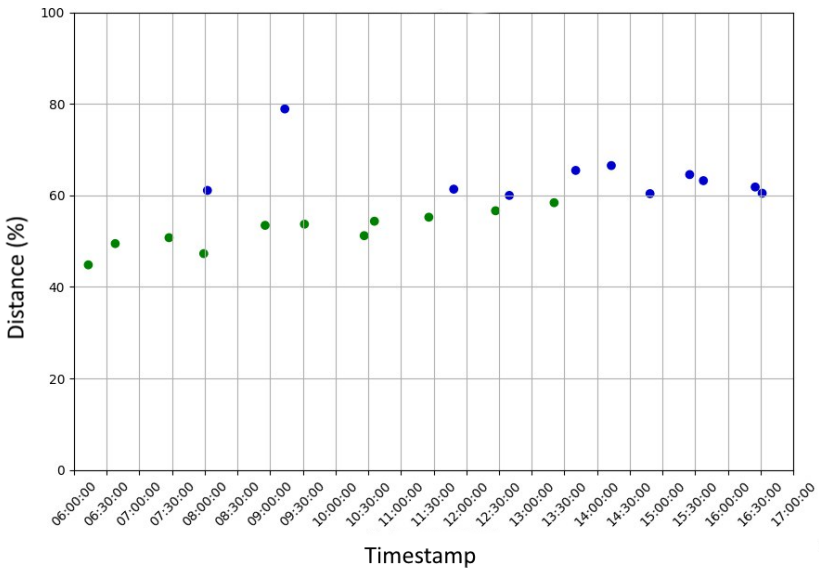


Figure 6. Example of a daily chart from the system, collected on 11/13/24. Each point represents the highest detected height as a percentage of the coast within 30-minute windows. The colors indicate the alert level generated by the wave, with green and blue representing low and medium, respectively.

Table 2. mAP50-95 per fold (Day). In bold, the folds selected to compose the dataset.

Fold	mAP50-95 (%)
Fold 1	**97.84**
Fold 2	98.87
Fold 3	**97.21**
Fold 4	**96.63**
Fold 5	99.32

Table 3. mAP50-95 per fold (Night). In bold, the folds selected to compose the dataset.

Fold	mAP50-95 (%)
Fold 1	**74.491**
Fold 2	87.503
Fold 3	90.437
Fold 4	**80.446**
Fold 5	**72.578**

tion sets from the folds corresponding to the two models with the worst performance, which we observe in Tables 2 and 3 (day and nighttime, respectively). We base this choice on the observation that these images are essential for training. They represent patterns that are difficult for the models to learn, and including them in the training set leads to superior performance in other folds. Additionally, we choose the third model with inferior performance and, from it, we use half of the validation set. The union of these selected images (from the two worst models plus half of the third) will constitute 50% of the training set for the complete dataset.

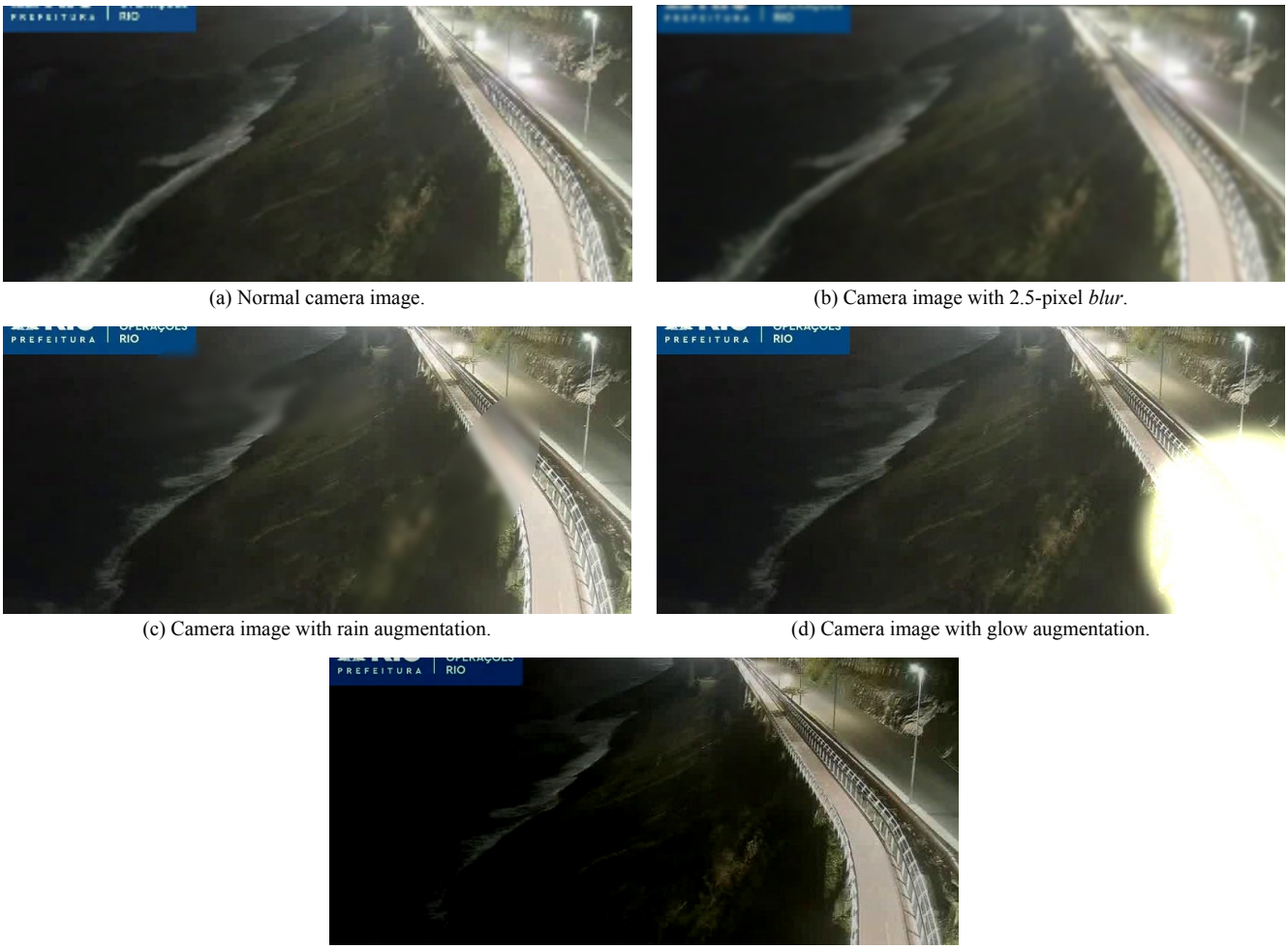
In addition to selecting images based on validation sets, we also apply cross-validation during training on the different datasets, which is essential to identify the variability in the models' performance and, therefore, their generalization capability. We use the standard deviation between the results of each model we obtain through cross-validation to evaluate the robustness of the generalization.

5 Results

The experiments aim to evaluate the impact of an expanded set of data augmentation techniques. We train 48 models, divided into three groups (day, night, and complete), and subject them to four data augmentation strategies and their combinations: Gaussian blur, brightness adjustment, rain simulation, and glow simulation. We evaluate the performance of the system based on two primary metrics: precision, measured by mAP50-95, and predictive stability, measured by the standard deviation from cross-validation. Figures 8 and 9 show the average mAP50-95 scores. Figure 8 shows the results for daytime images, while Figure 9 shows the results for nighttime images. We evaluate the models on the test sets for each dataset using images not included in the cross-validation process. Table 4 presents the detailed standard deviations for each condition (Day, Night, and Complete).

The results indicate that models trained exclusively on images from a specific domain show high performance in that same domain, but experience a sharp drop in performance when evaluated on the opposite domain. For example, Figure 8 shows the Day `Blur+Brightness+Rain+Glow` model achieved an mAP50−95 of 97.2% on the day test set, but its mAP dropped to 45.1% on the night set. Figure 9 shows that models trained on the day dataset, which correspond to the orange points, drop significantly when tested on night images. Similarly, night models also show reduced performance in day conditions, as seen in Figure 8 with the green points. This difference demonstrates that even a robust set of data augmentations is not sufficient to overcome the domain gap when training is restricted to a single period.

Applying data augmentations to domain-specific models drastically improves performance within other domains; nonetheless, it does not fully solve the core challenge of generalizing between day and night conditions, as the final mAP values remain relatively low. Still, clear improvements are observed: the best augmented night model achieved 82.5% mAP on day compared to 69.3% without augmentation, while the



(a) Normal camera image.

(b) Camera image with 2.5-pixel *blur*.

(c) Camera image with rain augmentation.

(d) Camera image with glow augmentation.

(e) Camera image with brightness augmentation.

Figure 7. Images captured by the camera with and without the data augmentation techniques.

best day-trained model reached 45.7% mAP on night images, surpassing its non-augmented counterpart with 19.6%. These results also help to explain why the overall performance increases when evaluating on the complete dataset (Figure 10): models trained with augmentations become more versatile and better equipped to handle the diversity of visual conditions present in the full set. This suggests that simulating adverse conditions is more effective when the model has already been exposed to an initial variety of lighting conditions, rather than attempting to teach a specialized model to recognize visual patterns drastically different from those seen during its training.

The group of models trained with the complete dataset shows more balanced and superior results in all scenarios, as seen in Figures 8 and 9 with the blue points. The best-performing model, Complete `Blur+Bright+Rain+Glow`, trained with the combination of all four techniques, obtained a stable mAP50−95 of 97.6% on the complete set, maintaining high performance on both the day (97.5%) and night (97.9%) subsets. Thus, in operational scenarios with variable lighting, using a combined set, along with data augmentation strategies, optimizes system performance with a single model.

Table 4 reveals distinct patterns of predictive stability. Models trained with the day dataset show low variability in day tests but high instability in night tests. The standard deviation of the Day `Blur+Bright+Rain+Glow` model, for example, increased from a low value of 0.41% in the day test to a volatile 18.78% in the night test. On the other hand, the models trained on the complete set proved to be highly robust, with significantly lower standard deviations in all evaluated domains. The same Complete `Blur+Bright+Rain+Glow` model registered low and consistent deviations: 0.43% (Day), 0.39% (Night), and 0.24% (Complete).

In general, the results indicate that training on a combined dataset, along with a comprehensive range of augmentation techniques, is the most effective strategy to reduce prediction variability and improve model stability. This approach allows for compensating for the limitations imposed by the low volume of available data, expanding the diversity of the training set and resulting in a final model that is not only accurate but also reliable in different operational conditions.

6 Correlation with Other Sources

To validate the effectiveness of the proposed monitoring system, we compare its outputs with independent oceanographic measurements obtained from buoy-based sensors installed on the Rio de Janeiro coast. Rather than relying on classical error metrics such as RMSE or MSE, which assume the estimation of absolute values, we focus on correlation. This choice reflects the fact that our system does not directly

Table 4. Standard deviation of models trained with cross-validation on the different test sets.

Model	Day			Night			Complete		
Augmentation	**Day Set**	**Night Set**	**Complete Set**	**Day Set**	**Night Set**	**Complete Set**	**Day Set**	**Night Set**	**Complete Set**
None	0.9820	0.7883	0.7450	15.6290	0.8981	8.0972	0.7022	0.2122	0.3391
Blur	0.4073	2.9137	2.0203	5.5444	2.7524	1.9282	0.6150	0.2326	0.3091
Brightness	0.5511	12.6162	5.8599	3.6164	2.0663	1.8241	0.9176	0.2197	0.5647
Rain	0.3401	1.5463	0.9068	4.3634	3.0346	1.4884	0.4214	0.4286	0.4850
Glow	0.6542	0.7583	1.3260	2.7732	1.2920	1.6250	0.9180	0.7147	0.8662
Blur + Brightness	0.4956	15.6084	7.6230	2.9394	3.0037	1.7969	0.5876	0.1599	0.3465
Blur + Rain	0.3670	0.8420	0.8930	8.3918	3.8376	3.7719	1.0026	0.5502	0.7615
Blur + Glow	**0.3077**	2.1586	1.7031	1.6161	0.9899	1.0574	1.5265	0.3365	0.8057
Brightness + Rain	0.5664	10.2071	3.4862	2.7746	0.8272	1.6616	1.2830	0.1876	0.8079
Brightness + Glow	0.3606	13.6807	4.4471	2.3024	0.9926	1.1375	0.7175	0.4505	0.5080
Rain + Glow	1.0442	0.6068	0.8346	1.8895	5.7738	3.1052	0.8202	0.4076	0.5666
Blur + Brightness + Rain	0.5005	11.7651	4.6060	0.7754	0.9069	0.7427	1.5484	0.1627	0.8818
Blur + Brightness + Glow	0.6371	16.4206	6.7956	2.8498	1.0524	1.8022	0.5946	**0.0944**	0.2630
Blur + Rain + Glow	0.8267	0.7546	1.2981	3.3960	1.1310	1.4177	0.8249	0.2425	0.5639
Brightness + Rain + Glow	0.5040	11.3346	6.1415	2.4818	1.8224	1.5530	0.4045	0.4330	0.3945
Blur + Brightness + Rain + Glow	0.4072	18.7834	7.9870	1.3822	0.9919	1.0314	0.4254	0.3941	**0.2380**

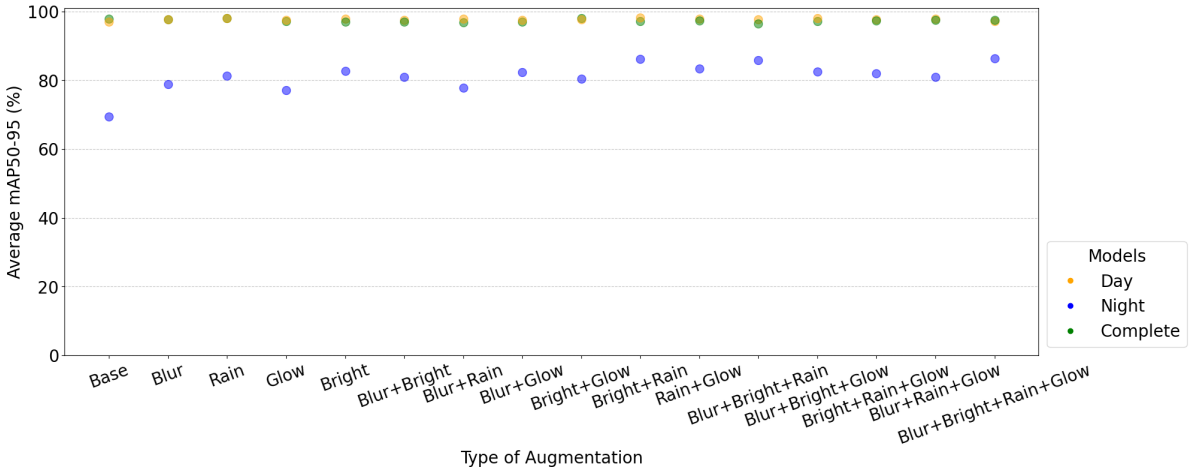


Figure 8. Mask segmentation mAP50-95, tested on day dataset.

estimate absolute wave height but instead captures relative variations. Correlation therefore provides a fairer assessment of how well the vision-based measurements track buoy observations. In particular, we focus on comparing five key wave height indicators:

- **Hsig (Significant Wave Height)**: the average height of the highest one-third of the waves observed during the sampling period, widely used as a standard descriptor of sea state;
- **Hmax (Maximum Wave Height)**: the largest individual wave measured by the buoy over a 30-minute window;
- **Hmo (Spectral Significant Wave Height)**: an estimate of significant wave height derived from the wave energy spectrum;
- **height_max (Camera Max Height)**: maximum wave height estimated from our segmentation system;
- **height_mean (Camera Mean Height)**: mean wave height estimated from our segmentation system.

The buoy data used for validation is publicly available through the SIMCosta platform (https://www.simcosta.furg.br/), provided with a 30-minute resolution. In this study, we select the closest available buoy, RJ-4, which has been consistently operational with 93.1% uptime since January 2024. The camera data we use in this comparison covers the period from April to August 2025.

The camera system generates estimates at 15 frames per second, requiring aggregation strategies to enable a fair comparison between data sources. For this purpose, we aggregate the results into 30-minute intervals, starting at 06:00 and ending at 17:00 (inclusive), to compute the camera mean and maximum signals. Although the system operates continuously, segmentation quality is reduced under low-light conditions such as early morning and nighttime. Restricting the analysis to daytime hours ensures consistency in the validation process, while recognizing that ongoing improvements have already reduced nighttime false positives from firing every 30 seconds to only one or two per day. Achieving full reliability during low-light periods remains a direction for future work, where additional data can further enhance system performance.

To be fair with each buoy variable, we compare maximum values (height_max) with Hmax and mean values (height_mean) with Hsig and Hmo. Missing values in the camera data, typically caused by temporary connection issues, are handled by replacing zero entries with the mean value for

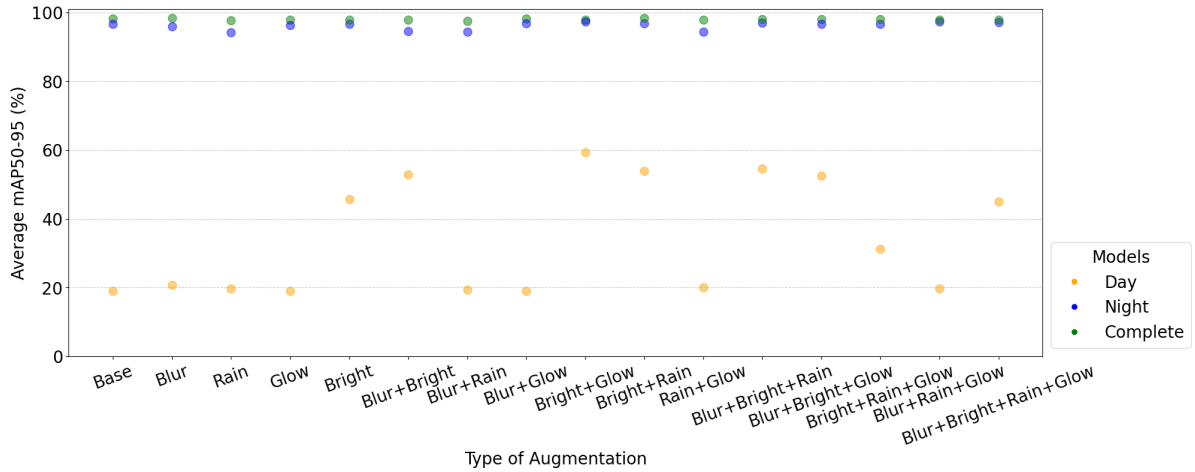


Figure 9. Mask segmentation mAP50-95, tested on night dataset.

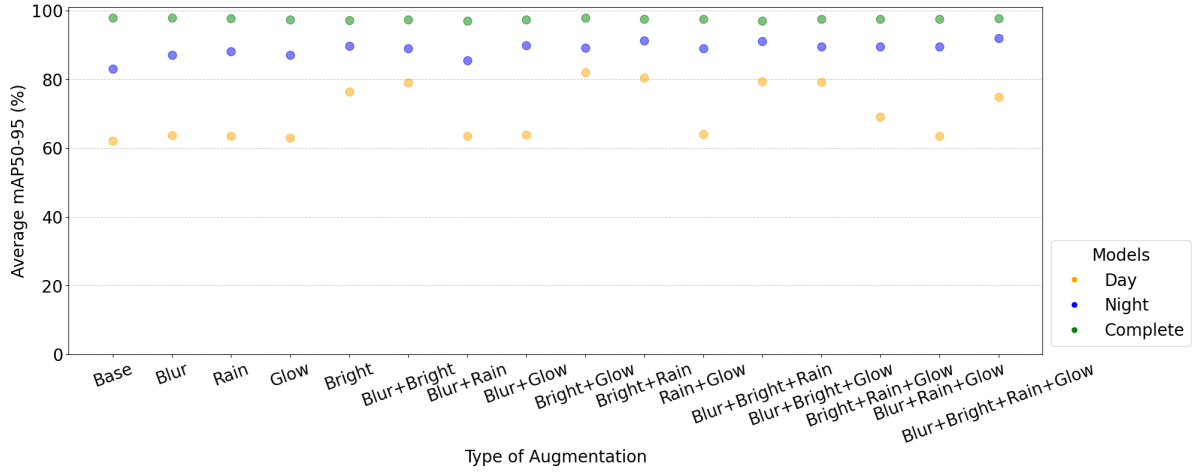


Figure 10. Mask segmentation mAP50-95, tested on complete dataset.

the corresponding hour. We exclude from the analysis days with system instabilities due to segmentation model performance, or camera inaccessibility, to ensure consistency and reliability. Overall, 65% of the data was considered valid.

Figure 11 presents the Spearman correlation between the camera-based height estimates and the buoy measurements for different wave height parameters during daytime hours. The correlation coefficient between Hmax and camera_max is 0.537, while the correlations between camera_mean and Hsig and Hmo are 0.463 and 0.459, respectively. These results demonstrate a modest but positive relationship between the buoy and camera measurements, confirming that the vision-based estimates effectively capture relative wave height variations despite differences in measurement principles and spatial coverage.

7 Conclusion

This work presented a computer vision–based monitoring method to detect risks associated with waves impacting coastal infrastructure, focusing on the Tim Maia Bike Lane, in Rio de Janeiro. The study showed that domain-specific data augmentation techniques, such as brightness adjustment, blur, and rain or headlight simulation, improve the robustness of segmentation models under adverse conditions.

Experiments revealed that models trained only on specialized datasets (daytime or nighttime) perform poorly outside their original domain, whereas training on a combined dataset together with augmentation techniques achieved better generalization and consistency, reaching 97.6% mAP50-95 across different lighting conditions. These results indicate that, given the current data volume, there is no advantage in maintaining separate models for distinct periods, since a single model trained with diverse data proves to be more effective and stable. The correlation analysis with buoy data further reinforces the potential of the proposed system to complement existing ocean monitoring networks and support urban coastal risk management frameworks.

Future work should focus on expanding the dataset to better comprehend how data volume impacts model performance, on exploring new domain adaptation strategies, and investigating alternative architectures beyond YOLO to achieve additional gains in precision and efficiency.

Declarations

Authors' contributions

P.V.R. and R.L.K. collected the datasets, developed the experimental code, and trained the models. P.V.R. also de-

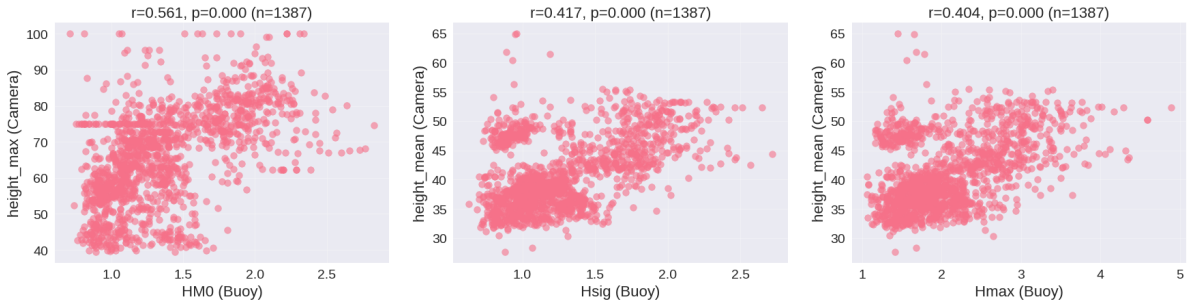


Figure 11. Scatter plots comparing wave height estimates from the camera system with buoy measurements. Spearman correlation coefficients (r) with their p-values are shown above each plot.

veloped the live system. P.C., R.S.C., and L.H.M.K.C. defined the methodology and reviewed the experimental results. A.G.C. and M.A. provided support for accessing COR's systems and continuously tested the application. All authors contributed to writing and reviewing the manuscript.

Availability of data and materials

All data used in this study were collected by the authors and are original. The datasets and models will be made publicly available on the GTA-UFRJ GitHub repository.

Competing interests

The authors declare no competing interests.

Funding

This work was partially sponsored by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Funding Code 001, CNPq (408255/2023-4, 405940/2022-0 and 308917/2025-1), FAPERJ (E-26/204.122/2024 e E-26/204.562/2024), and FAPESP (2023/00673-7 and 2023/00811-0).

References

Alam, M. J., Hasan, M. Z., Rahman, M. M., Rahman, M. A., Sarker, N. H., Azad, S., Islam, T. N., Paul, B., Anjum, T., Halder, B., and Fattah, S. A. (2025). An optimized yolov5 based approach for real-time vehicle detection at road intersections using fisheye cameras. DOI: 10.48550/arxiv.2502.04566.

CVAT.ai Corporation (2023). Computer Vision Annotation Tool (CVAT). DOI: 10.5281/zenodo.18750037.

Herath, H. and Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, 2(1):100076. DOI: 10.1016/j.jjimei.2022.100076.

Jocher, G., Qiu, J., and Chaurasia, A. (2023). Ultralytics YOLO. Available at:`https://github.com/ultralytics/ultralytics`.

Kaur, R., Roul, R. K., and Batra, S. (2023). A hybrid deep learning CNN-ELM approach for parking space detection in smart cities. *Neural Computing and Applications*, 35(18):13665–13683. DOI: 10.1007/s00521-023-08426-y.

Liu, X. (2024). The application of CNN in smart city management and supervision under the background of smart city. *Measurement: Sensors*, 33:101227. DOI: 10.1016/j.measen.2024.101227.

Martinho, L., Neto, O., Calvalcanti, J., Pio, J., and Oliveira, F. (2023). An approach for fish detection in underwater images. In *Anais do XVIII Workshop de Visão Computacional*, pages 6–11, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/wvc.2023.27524.

Midigudla, R. S., Dichpally, T., Vallabhaneni, U., Wutla, Y., Sundaram, D. M., and Jayachandran, S. (2025). A comparative analysis of deep learning models for waste segregation: YOLOv8, EfficientDet, and Detectron 2. *Multimedia Tools and Applications*, 84:35941–35964. DOI: 10.1007/s11042-025-20647-y.

Muhadi, N. A., Abdullah, A. F., Bejo, S. K., Mahadi, M. R., and Mijic, A. (2021). Deep learning semantic segmentation for water level estimation using surveillance camera. *Applied Sciences*, 11(20). DOI: 10.3390/app11209691.

Onufriichuk, O. (2025). aiogram. Available at:`https://github.com/aiogram/aiogram`.

Rio, G. (2016). Parte da ciclovia desaba em são conrado, zona sul do rio. Available at:`https://g1.globo.com/rio-de-janeiro/noticia/2016/04/parte-da-ciclovia-desaba-em-sao-conrado-zona-sul-do-rio.html`.

Scardino, G., Scicchitano, G., Chirivì, M., Costa, P. J. M., Luparelli, A., and Mastronuzzi, G. (2022). Convolutional neural network and optical flow for the assessment of wave and tide parameters from video analysis (leucotea): An innovative tool for coastal monitoring. *Remote Sensing*, 14(13). DOI: 10.3390/rs14132994.

Talaat, F. M. and ZainEldin, H. (2023). An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Computing and Applications*, 35(28):20939–20954. DOI: 10.1007/s00521-023-08809-1.

Telegram (2025). Telegram. Available at:`https://telegram.org`. Version: Android: 11.8.0.

Ullah, Z., Al-Turjman, F., Mostarda, L., and Gagliardi, R. (2020). Applications of artificial intelligence and machine learning in smart cities. *Computer communications*, 154:313–323. DOI: 10.1016/j.comcom.2020.02.069.

Wang, J. and Zhao, H. (2024). Improved YOLOv8 algorithm for water surface object detection. *Sensors*, 24(15). DOI: 10.3390/s24155059.

Yuldashev, Y., Mukhiddinov, M., Abdusalomov, A. B., Nasi-

mov, R., and Cho, J. (2023). Parking lot occupancy detection with improved MobileNetV3. *Sensors*, 23(17). DOI: 10.3390/s23177642.