Image Segmentation with Zero-Shot Generalization for Flood Prediction in Urban Environments

Conference Paper · October 2024

CITATIONS
0

READS
69

4 authors, including:

Caetano Mazzoni Ranieri
São Paulo State University
45 PUBLICATIONS **284** CITATIONS
SEE PROFILE

Saulo Neves Matos
University of São Paulo
23 PUBLICATIONS **22** CITATIONS
SEE PROFILE

Jó Ueyama
University of São Paulo
183 PUBLICATIONS **4,050** CITATIONS
SEE PROFILE

Image Segmentation with Zero-Shot Generalization for Flood Prediction in Urban Environments ⋆

Pedro A. S. Negrão * **Caetano M. Ranieri** ** **Saulo N. Matos** * **Jó Ueyama** *

* *Institute of Mathematical and Computer Sciences, University of São Paulo (USP), São Carlos, SP, Brazil (e-mail: pedron_1997@hotmail.com, saulo.matos@usp.br, joueyama@icmc.usp.br).*
** *Institute of Geosciences and Exact Sciences, São Paulo State University (UNESP), Rio Claro, SP, Brazil (e-mail: cm.ranieri@unesp.br).*

Abstract: Accurate and continuous monitoring of river levels is crucial for managing water resources and preventing natural disasters. This project aims to analyze the water levels of urban creeks using images and develop a prediction algorithm based on the extracted information. The study seeks to explore the feasibility and effectiveness of using image analysis to evaluate, monitor, and predict the water level of the Mineirinho Creek in São Carlos, Brazil. The methodology involves regularly capturing images over a specific period and processing them using various computer vision techniques to determine the water level. The Segment Anything (SAM) model was employed to segment the water body, while a sequence of steps was adopted to provide a flooding index. A comparative analysis confirmed the accuracy and reliability of the proposed method.

Keywords: Deep learning, Flood management, Image processing, Image segmentation, Water monitoring.

1. INTRODUCTION

Effective risk management in flood-prone areas is a critical concern, especially in regions of Brazil that are frequently affected by urban flooding. Implementing robust strategies to mitigate the impact on local communities is essential. These strategies are broadly categorized into structural measures, such as dams and water diversions, and non-structural measures, which include zoning and early warning systems Iqbal et al. (2021). Among these, early warning systems are particularly vital as they provide timely data that enable rapid responses.

Traditional flood management methods often depend on costly and maintenance-heavy techniques that utilize point data, such as precipitation and water levels (Lo et al., 2015b). The water level can be measured using either a contact or contactless method. Contact approaches, such as employing differential pressure transducers (Gold et al., 2023; Schenato et al., 2021), imply that the sensors might directly contact the liquid. Although these sensors may provide accurate measurements, they are at risk of damage or being covered by floods and the resulting debris and sand (Lo et al., 2015a).

Non-contact measurement techniques, such as ultrasonic sensors (Sahoo and Udgata, 2020; Yunita et al., 2020) and LiDAR (Ranieri et al., 2024a), offer increased durability by being less prone to physical damage, though they generally fall short of the accuracy achieved through direct contact methods. Alternatively, computer vision presents a cost-effective solution by utilizing images captured by monitoring cameras directed at water bodies. This method not only reduces the risk of equipment damage but also leverages readily available technology to monitor water levels.

Early attempts in this domain include the work of San Miguel and Ruiz Jr (2016), who applied image processing techniques to detect water bodies from diverse media sources. However, their methods were limited by varying image qualities and practical applicability. Furthermore, Lopez-Fuentes et al. (2017) demonstrated advanced algorithms for water segmentation on a limited dataset, underlining the difficulty of generalizing deep learning models without extensive annotated data.

To overcome these limitations, Huang et al. (2020) creatively employed readily accessible data from the internet and security cameras to train a Mask RCNN model, using vehicle tires as reference objects for assessing water levels. This approach still hinges on specific reference objects within the images.

⋆ This research was funded by the São Paulo Research Foundation (FAPESP), grants 2021/10921-2 and 2022/09644-7, and the National Council for Scientific and Technological Development (CNPq).

Building upon these foundations, our study aims to push the boundaries of this field using the E-Noé system's unique dataset, which has been developed through an extensive wireless sensor network (Ranieri et al., 2024b). Our project involved using the Segment Anything Model (SAM) (Kirillov et al., 2023) to automatically segment the water body in river images, eliminating the need for pre-annotated datasets. We analyzed the obtained masks based on their size in the image and defined an index that correlated the mask with the water level in the creek shown in the image. This index was then used with a classification criterion to generate a classification. This approach has the potential to broaden the model's applicability in various environments and enhance our understanding of effective flood management practices.

The remainder of this paper is organized as follows. Section 2 describes the dataset used in our experiments, detailing the image acquisition process and the labeling methodology. Section 3 outlines the proposed methodology, including data preprocessing, water segmentation using the Segment Anything Model (SAM), and the calculation of the water level index. Section 4 presents the experimental results, including quantitative evaluations and analysis of the classification performance. Finally, Section 5 concludes the paper with a summary of our findings and suggestions for future work.

2. DATASET

The dataset utilized in our experiments has been previously implemented, as discussed by Ranieri et al. (2024b). It consists of images taken from stationary cameras mounted on a pole, monitoring the water flow in urban creeks in the city of São Carlos, Brazil. Each image was manually labeled based on the observed water level and categorized into four discrete levels, numbered from 1 (low) to 4 (flooded). Special labels were provided for images where the water stream was not visible due to poor lighting conditions or camera failure. These images were discarded from the dataset before our experiments.

Figure 1 illustrates the images obtained at the three locations contained in the dataset. The locations were named SESC2, SHOP, and SHOP2.

We used only the images at the SHOP and SHOP2 locations because they followed a standardized and objective procedure for annotation, as discussed by Ranieri et al. (2024b). In any case, the dataset exhibits severe imbalance, as illustrated by the distribution of levels in Figure 2, which concatenates data from the SHOP and SHOP2 locations. The graph was built using logarithmic scale because the number of images in low water level is so small compared to the other classes.

Although existing techniques to handle data imbalance could help (Thabtah et al., 2020), they would not suffice in training models as accurate as pressure or distance sensors' measurements. For this reason, our methodology assessed the water level based on the proportion of water detected in the image through a segmentation model with zero-shot generalization - the Segment Anything Model (SAM) model by Meta AI (Kirillov et al., 2023).

3. METHODOLOGY

The methodology involved preprocessing the data and employing a water segmentation approach based on the SAM model. It then extracts an index from the segmentation mask proportional to the water level in the stream and provides a classification based on this index. The following subsections provide details on each step performed.

3.1 Data preprocessing

Before using the segmentation model, we conducted a preprocessing step to address any issues that could impact water segmentation in the images. The data in our dataset was obtained from fixed cameras placed in a real-world environment. The image-capturing process was influenced by various factors such as light incidence, light reflectance on the water surface, and low illumination at night when public lighting was down due to extreme weather conditions. Figure 3 showcases examples of these conditions.

Illumination is one of the most crucial factors in image capture (Shih, 2010). Proper illumination is essential for highlighting scene details and creating a pleasant appearance. Inadequate lighting can lead to problems during capture, such as incorrect exposure, excessive brightness, or low luminosity. To tackle these issues, we utilized morphological analysis and filter convolution to address the effects of illumination problems.

We used a cleaning procedure to automatically discard images where the water stream was barely visible due to lighting conditions. The process involved converting the images to grayscale, applying a Gaussian filter to remove high-frequency noise, and using a threshold to select only intensity values above 200. Finally, we calculated the number of white pixels in the final image and the number of black pixels in the image blurred by the filter.

3.2 Water segmentation

When applied to the water streams, segmentation models can allow for a better understanding of the hydrological conditions of watercourses (Eltner et al., 2021). While training a segmentation model based on collected data could provide a specialized solution (Fernandes Junior et al., 2021), which could require retraining each time a new camera location is provided, the SAM model is a general-purpose solution that identifies image regions after receiving one or more points representing the object to be identified.

No training procedure was necessary since we used a generalized zero-shot learning approach (Pourpanah et al., 2022). For making inferences with the SAM model, the coordinates of a set of points are required as part of the input prompt. Each point might be labeled as foreground or background. Multiple points can be provided to select a single object. If a mask from a previous iteration is available, it can also be provided to assist the model in prediction.

By default, SAM generates three masks and scores to estimate the quality of these masks. This setup is designed to handle unclear input prompts and enable the model to

(a) (b) (c)

Figure 1. Three camera locations present in the dataset, named (a) SESC2, (b) SHOP, and (c) SHOP2.

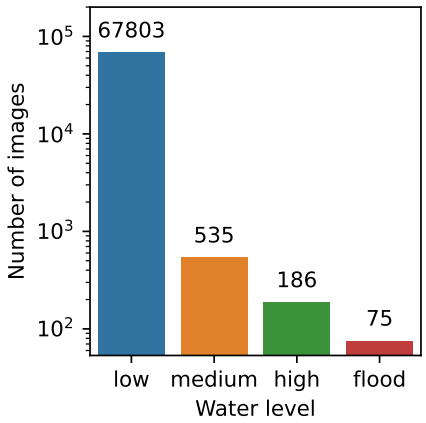
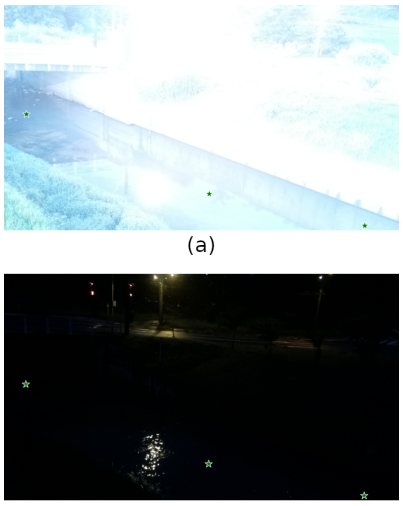


Figure 2. Number of images for each water level, in log scale, for the SHOP and SHOP2 locations.

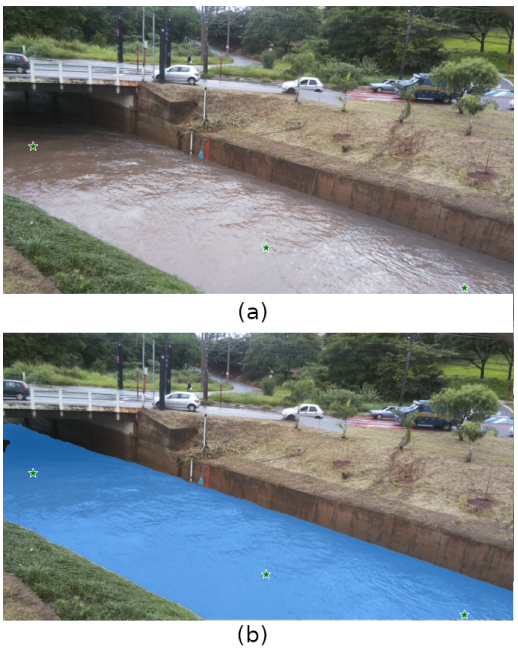


(a) (b)

Figure 3. Examples of illumination problems: (a) excessive brightness and (b) low luminosity.

differentiate between different objects based on the input. The model produces masks, scores, and logits for low-resolution masks, which can be used for the next inference. An example of this ambiguity is shown in Figure **??**, in which a single point was used, and the model returned multiple objects in the same mask.

Since we considered images from two camera positions (i.e., SHOP and SHOP2), the same point coordinates could not be used for all images. Therefore, we created a dictionary with three to five points for each place, year, and month in which the image was captured. A file compiling the ideal points for each of the 26 possible combinations was created. A sample outcome of this approach can be observed in Figure 4. If one of the points provided is poorly positioned, the model might infer a mask that does not correspond to the object intended for segmentation.



(a) (b)

Figure 4. Sample result from the second inference using the previous masks and adding two points. (a) original image with the selected points; (b) mask identified by the model.

If any of the provided points select an object other than the river, the mask area can become much wider than the limits of the water stream in the image. To address this, we created a contour of the river mask based on the highest water level assessed for each month and location. Then, we defined a bounding box to encompass the entire segmented area. If parts of the mask are outside this box, those pixels are considered out of bounds, indicating an error in the water segmentation. An example of it is shown in Figure 5.

3.3 Water level index

We hypothesized that the ratio between the areas of the segmented region and the bounding box was proportional to the water level. This assertion could be validated further based on the annotations present in the dataset. Formally, for an input image x, let $s_b(x)$ be the total number of pixels in the bounding box, and $s_s(x)$, the total number of pixels in the segmented water area inscribed to it. We assess the water level by computing the ratio $r(x)$ as in Equation 1.

Figure 5. Bounding box encompassing the segmentation mask.

$$r(x) = \frac{s_s(x)}{s_b(x)} \tag{1}$$

We performed inference and post-processing on all images in the dataset and validated our hypothesis by comparing the values of the index $r(x)$ to the labeled data in the dataset.

3.4 Classification criteria

To provide a classification of an image x based on the index $r(x)$, a straightforward criterion was employed, introducing a threshold based on the value of a quantile of the distribution of the index for each class. Hence, for each category $y_i \in \{\texttt{low}, \texttt{medium}, \texttt{high}, \texttt{flood}\}$, a quantile q was established for the distribution of $r(x)$ and set these values as threshold. If $z(y_i)$ is the threshold computed for category y_i, then the predicted label $y'(x)$ is computed as in Equation 2.

$$\begin{cases} r(x) < z(\texttt{low}) \implies y'(x) = \texttt{low} \\ z(\texttt{low}) \leq r(x) < z(\texttt{medium}) \implies y'(x) = \texttt{medium} \\ z(\texttt{medium}) \leq r(x) < z(\texttt{high}) \implies y'(x) = \texttt{high} \\ r(x) \geq z(\texttt{high}) \implies y'(x) = \texttt{flood} \end{cases} \tag{2}$$

In our experiments, we analyzed the values $q_i \in \{0.6, 0.75, 0.9\}$. We took into account the confusion matrix normalized by the number of samples from each class and the balanced accuracy, which is the average of recall (Brodersen et al., 2010). We did this because these metrics can provide a useful evaluation despite the severe class imbalance in the dataset, making traditional metrics like overall accuracy unsuitable (Lakshmanan et al., 2020).

4. RESULTS

We assessed the proportion of the pixels identified for each annotated level to evaluate the segmentation results. A quantitative evaluation was performed by analyzing a boxplot graph, shown in Figure 6. Comparing the median of each level, we can observe a gradual increase according to the manually annotated classes, while there was little superposition between the boxplots for each water level, apart from outliers. This result implies that the index proposed is indeed related to the water level and can be used to provide classifications.

Figure 7 illustrates the confusion matrix for the three quantiles considered in our evaluations for the classification procedure described in Section 3.4. The results indicate that for $q = 0.6$ and $q = 0.75$, the proposed criteria lead to high accuracy in classifying flood images, despite a significant number of false positives originating from the `high` class (40% for $q = 0.6$ and 25% for $q = 0.75$). Increasing the threshold to $q = 0.9$ enhances the accuracy for instances with low water levels but reduces the accuracy for detecting other classes. Given that identifying high water levels and floods are the most critical conditions the model aims to recognize, the results indicate that, for this dataset, using a threshold around 0.75 is the best choice.

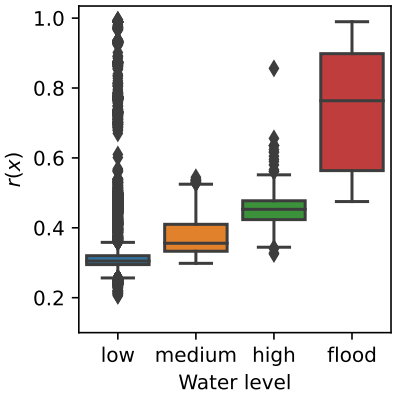


Figure 6. Pixels' ratio across all images.

Table 1 displays the balanced accuracy for different threshold values. The best result, with a balanced accuracy of 74%, was achieved when the threshold value was set to $q = 0.75$. It's important to note that increasing the threshold value to $q = 0.9$ resulted in poorer performance (i.e., 43%), despite an improvement in the majority class (i.e., `low`). This is because the metric is designed to account for class imbalance.

Table 1. Balanced accuracy for each quantile considered as threshold.

Threshold (quantile)	**Balanced accuracy**
0.6	0.67
0.75	0.74
0.9	0.43

5. CONCLUSION

This study has demonstrated the effectiveness of using image segmentation techniques for monitoring and predicting flood levels in urban environments. By employing the Segment Anything Model (SAM), we successfully segmented water bodies in images captured by stationary cameras and developed an index to classify water levels.

Our approach offers a cost-effective and durable alternative to traditional flood monitoring methods, which often rely on contact or non-contact sensors that can be prone to damage or require extensive maintenance. The image-based method reduces these risks while leveraging readily available technology.

Future work could focus on refining the segmentation model to further improve accuracy and explore the inte-

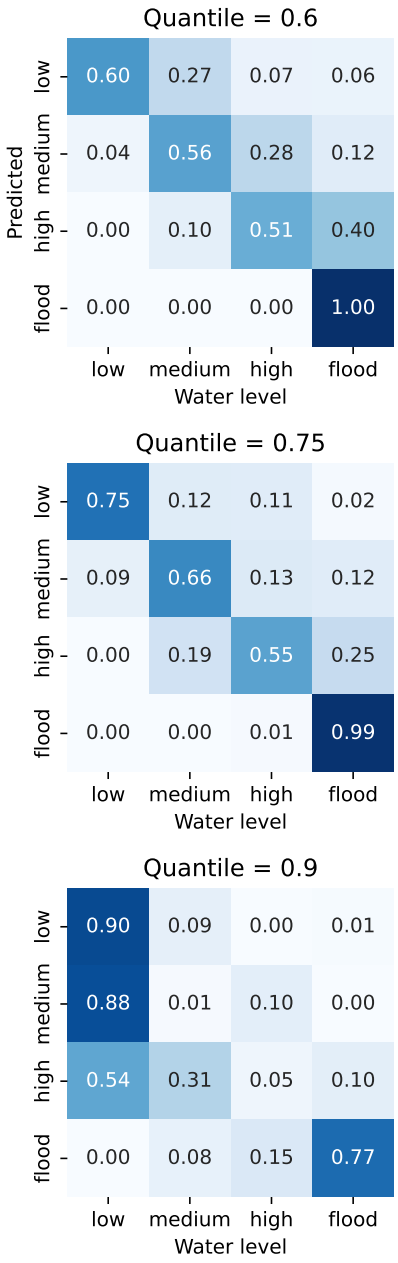


Figure 7. Confusion matrix for classifying the water level based on the ratio $r(x)$, for each quantile considered for thresholding.

gration of additional environmental data to enhance predictive capabilities. Additionally, expanding the dataset to include more diverse urban settings could help validate the generalizability of the proposed method.

ACKNOWLEDGMENTS

This research was funded by the São Paulo Research Foundation (FAPESP), grants 2021/10921-2 and 2022/09644-7, and the National Council for Scientific and Technological Development (CNPq).

REFERENCES

Brodersen, K.H., Ong, C.S., Stephan, K.E., and Buhmann, J.M. (2010). The balanced accuracy and its posterior distribution. In *2010 20th international conference on pattern recognition*, 3121–3124. IEEE.

Eltner, A., Bressan, P.O., Akiyama, T., Gonçalves, W.N., and Marcato Junior, J. (2021). Using deep learning for automatic water stage measurements. *Water Resources Research*, 57(3), e2020WR027608.

Fernandes Junior, F.E., Nonato, L.G., Ranieri, C.M., and Ueyama, J. (2021). Memory-based pruning of deep neural networks for iot devices applied to flood detection. *Sensors*, 21(22), 7506.

Gold, A., Anarde, K., Grimley, L., Neve, R., Srebnik, E.R., Thelen, T., Whipple, A., and Hino, M. (2023). Data from the drain: a sensor framework that captures multiple drivers of chronic coastal floods. *Water Resources Research*, 59(4), e2022WR032392.

Huang, J., Kang, J., Wang, H., Wang, Z., and Qiu, T. (2020). A novel approach to measuring urban waterlogging depth from images based on mask region-based convolutional neural network. *Sustainability*, 12(5), 2149. doi:10.3390/su12052149. URL `http://dx.doi.org/10.3390/su12052149`.

Iqbal, U., Perez, P., Li, W., and Barthelemy, J. (2021). How computer vision can facilitate flood management: A systematic review. *International Journal of Disaster Risk Reduction*, 53, 102030. doi:10.1016/j.ijdrr.2020.102030. URL `http://dx.doi.org/10.1016/j.ijdrr.2020.102030`.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., and Girshick, R. (2023). Segment anything.

Lakshmanan, V., Robinson, S., and Munn, M. (2020). *Machine learning design patterns*. O'Reilly Media.

Lo, S.W., Wu, J.H., Lin, F.P., and Hsu, C.H. (2015a). Cyber Surveillance for Flood Disasters. *Sensors*, 15(2). doi:10.3390/s150202369.

Lo, S.W., Wu, J.H., Lin, F.P., and Hsu, C.H. (2015b). Visual sensing for urban flood monitoring. *Sensors*, 15(8), 20006–20029. doi:10.3390/s150820006. URL `http://dx.doi.org/10.3390/s150820006`.

Lopez-Fuentes, L., Rossi, C., and Skinnemoen, H. (2017). River segmentation for flood monitoring. In *2017 IEEE International Conference on Big Data (Big Data)*. IEEE. doi:10.1109/bigdata.2017.8258373. URL `http://dx.doi.org/10.1109/BigData.2017.8258373`.

Pourpanah, F., Abdar, M., Luo, Y., Zhou, X., Wang, R., Lim, C.P., Wang, X.Z., and Wu, Q.J. (2022). A review of generalized zero-shot learning methods. *IEEE transactions on pattern analysis and machine intelligence*.

Ranieri, C.M., Foletto, A.V., Garcia, R.D., Matos, S.N., Medina, M.M., Marcolino, L.S., and Ueyama, J. (2024a). Water level identification with laser sensors, inertial units, and machine learning. *Engineering Applications of Artificial Intelligence*, 127, 107235.

Ranieri, C.M., Souza, T.L.D.e., Nishijima, M., Krishnamachari, B., and Ueyama, J. (2024b). A deep learning workflow enhanced with optical flow fields for flood risk estimation. *Applied Intelligence*, 1–22.

Sahoo, A.K. and Udgata, S.K. (2020). A novel ANN-based adaptive ultrasonic measurement system for accurate water level monitoring. *IEEE Transactions on Instrumentation and Measurement*, 69(6), 3359–3369. doi:10.1109/tim.2019.2939932.

San Miguel, M.J.P. and Ruiz Jr, C.R. (2016). Automatic flood detection using the video of static cameras. In *DLSU Research Congress*. De La Salle University Manila, Philippines.

Schenato, L., Aguilar-Lopez, J.P., Galtarossa, A., Pasuto, A., Bogaard, T., and Palmieri, L. (2021). A rugged FBG-based pressure sensor for water level monitoring in dikes. *IEEE Sensors Journal*, 21(12), 13263–13271. doi:10.1109/JSEN.2021.3067516.

Shih, F.Y. (2010). *Image processing and pattern recognition: fundamentals and techniques*. John Wiley & Sons.

Thabtah, F., Hammoud, S., Kamalov, F., and Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, 513, 429–441.

Yunita, A.M., Wardah, N.N., Sugiarto, A., Susanti, E., Sujai, L., and Rizky, R. (2020). Water level measurements at the cikupa pandeglang bantendam using fuzzy sugenowith microcontroler-based ultrasonik sensor. In *Journal of Physics: Conference Series*, volume 1477, 052048. IOP Publishing. doi:10.1088/1742-6596/1477/5/052048.