Monitoring Water Level Trend Using River Cameras: Integrating Domain-Specific Models and Segment Anything Model (SAM) for Transferable Water Segmentation

Ze Wang[1], Heng Lyu[1*], Shunan Zhou[1], and Chi Zhang[1]

[1]School of Hydraulic Engineering, Dalian University of Technology, Dalian, Liaoning, China.

*Corresponding author: Heng Lyu (lyuheng@dlut.edu.cn)

Key points

- A deep learning framework for water segmentation in river camera images and water level trend monitoring is developed.

- The framework's transferability in new river scenarios is achieved by combining domain-specific models and General AI.

- The static observer flooding index calculated by the framework can accurately characterize water level variations.

Keywords:

River; Water level trend monitoring; Water segmentation; Segment Anything Model

Monitoring Water Level Trend Using River Cameras: Integrating Domain-Specific Models and Segment Anything Model (SAM) for Transferable Water Segmentation

Abstract

Water level variations influence the biochemical and hydrological processes within river networks. Through river cameras, obtaining reliable water segmentation from image data can practically support the monitoring of water level. However, limited annotated data and tedious local deployment restrict the applicability of current deep learning water segmentation models in new river scenarios. To pursue transferability, this study proposes a novel framework that combines domain-specific models with General AI for water segmentation. The framework utilizes a ResUnet model pre-trained on a non-local dataset to identify the pixel with the highest probability of being water from the image. The Segment Anything Model (SAM), a promptable foundational computer vision model developed by Meta AI, is then adopted to use the pixel as prompt for generating water masks. When prompted, different modes of SAM are used for comparison. We applied the framework to image sequences acquired from river cameras stationed at four locations in Tewkesbury, UK. The framework significantly improved segmentation performance, with an increase of over 15% in Intersection over Union (IoU) over the single ResUnet model. Meanwhile, the results substantiated *point prompt* as the optimal mode for feeding prior knowledge on water to SAM. The static observer flooding index (SOFI) time series calculated based on the framework's segmented masks under *point prompt* mode exhibit an average correlation of 0.90 with real water level fluctuations, significantly surpassing the correlation of 0.54 attained by ResUnet. Our study thus represents a step toward implementing river cameras for robust water level trend monitoring.

1. Introduction

River networks act as conduits between terrestrial and aquatic environments, mediating hydrological processes that regulate river flow status dynamics (Mosley, 2015; Whitworth et al., 2012). In both headwater streams and continental-scale basins, river flow is dominated by water level gradients. The river water level variation controls the biochemical transport and determines the inundated extent of floodplains (Richey et al., 2002). Thus, continuous monitoring of water level is key to unravelling the complex exchange fluxes among multiple hydrological components, as well as for water resources management and establishing flood early warning systems (Yamazaki et al., 2012).

Due to the high costs associated with the installation and long-term maintenance of gauging stations, water level observations currently are unattainable for many rivers (Fekete et al., 2012; Ruhi et al., 2018). Other observational approaches, such as remote sensing, can supplement the data from hydrometric networks (Tauro et al., 2018). Nevertheless, satellite and airborne optical techniques are limited to their daylight-only application, susceptibility to obstruction by clouds and vegetation, and relatively long revisit intervals (Grimaldi et al., 2016; Yan et al., 2015). Acquiring high spatial-temporal resolution water level data in real-time or long-term is, thus, still challenging.

As computer vision develops, river cameras provides a novel path to collect water level data (Spasiano et al., 2023). River cameras are generally consumer-grade field cameras powered by electricity grids or (backup) batteries, resulting in low costs on equipment, installation, and maintenance (Noto et al., 2022; Sabbatini et al., 2021). They are increasingly installed for hydrological monitoring, offering extensive coverage of the river network (Gupta et al., 2022; Lo et al., 2015; Perks et al., 2020). These cameras continuously transmit live images from rivers and can store images locally or upload them to the cloud in real-time. The accumulation of image data has laid the foundation for the automatic extraction of water level variation information.

Aided by deep learning models, interpreting river water levels in the continuous domain from images can already be realized (Vandaele et al., 2023). Deep learning models can transform water level interpretation into regression tasks by directly establishing the mapping relationship between images and water level values. However, the establishment of the training dataset necessitates sites equipped with both cameras (for input preparation) and gauging stations (for label preparation), a requirement that cannot be met in certain regions given their local monitoring conditions. Furthermore,

deep learning-based regression models are typically limited by their inability to extrapolate outputs, making it challenging to monitor extremely high or low water levels (Vanden Boomen et al., 2021).

An alternative approach to imaging-based water level monitoring involves segmenting water pixels within each image, which no longer requires additional deployment of gauging stations. The segmented water masks can be transformed using photogrammetric techniques and then overlaid onto the topography of river channels to derive scalar water level values (Sermet & Demir, 2023). If fine-scale terrain data cannot be obtained, each image's static observer flooding index (SOFI), which indicates the proportion of water pixels among the total pixels, can be adopted as surrogate index to monitor water level trends (Moy De Vitry et al., 2019). According to previous studies, a robust characterization of water level trends is already informative for hydrological model calibration, with the Spearman rank correlation coefficient serving as the optimization objective, even in the absence of scalar water level values (Etter et al., 2020; Seibert & Vis, 2016; Weeser et al., 2019). Therefore, imaging-based water level trend monitoring based on water segmentation can be a more flexible and robust technical path.

Currently, water segmentation mainly relies on deep learning semantic segmentation models for their high automation and scalability (Eltner et al., 2018). A series of classic model structures, including SegNet, Fully Convolutional Networks, Fully Convolutional DenseNets, and Conditional Adversarial Networks, have already been applied to water segmentation for river camera images (Akiyama et al., 2020; Erfani et al., 2022; Lopez-Fuentes et al., 2017). However, deep learning models are significantly influenced by the amount of training data. Large-scale annotated river image data is still unavailable, limiting the potential transfer of trained models to new river scenes (i.e., locations unseen in the training dataset). Transfer learning offers an available route to enhance model transferability by fine-tuning deep learning models on a small subset of annotated data from new monitoring sites (Akiyama et al., 2021; Eltner et al., 2021; Vandaele et al., 2021). Nevertheless, fine-tuned deep learning models often experience catastrophic forgetting (Kirkpatrick et al., 2017), wherein excessive adaption to new data results in a notable drop in performance on the original dataset, rendering the model to be overly localized. Meanwhile, one-size-fits-all transfer learning strategies that accommodate diverse model architectures and application scenarios are still lacking (Weiss et al., 2016). A transferable water segmentation framework that necessitates minimal local adjustments and can be readily deployed in new river scenarios is called for.

The emergence of General AI such as Segment Anything Model (SAM, Kirillov et al., 2023) for computer vision tasks are reshaping the application of deep learning, and also lay a groundwork for developing transferable water segmentation models. General AI enables the comprehension and preprocessing of images, significantly reducing the domain-specific knowledge (i.e., additional training data) required for downstream tasks. However, General AI cannot be independently used for handling specific tasks due to its generality. It requires prompts, either as additional inputs during the modeling process or for post-processing model results. In this study, we employ a combination of domain-specific models and General AI for water segmentation on images captured by river cameras, aiming to balance the strengths and weaknesses of each model to achieve transferability. This approach is intended to mitigate the necessity for localized adjustments, such as the manual annotation and model parameter fine-tuning, required in new river scenes.

Overall, the main objective of this study is to developing a novel transferable deep learning-based water segmentation framework for monitoring the water level trend using river cameras. The framework was applied to four spots in Tewkesbury, UK, under different prompt modes of SAM. Its performance was compared with a single domain-specific deep learning model, ResUnet. Finally, SOFI was used as a qualitative index for monitoring the water level trend.

The remainder of this paper is organized as follows: Section 2 elaborates on the framework for water segmentation, along with the dataset used in this study. The detailed model results, and the evaluation of the model performance are presented in Section 3. Section 4 further discusses the value of the extracted water level trend, the superiority of the methods, and the implications for future studies. Finally, the conclusions are given in Section 5.

2. Methods and materials

2.1 Water segmentation model structure

2.1.1 Brief introduction to ResUnet model

In this study, the ResUnet model is adopted as the representative domain-specific deep learning model and pre-trained for water segmentation. ResUnet is a conventional end-to-end convolutional neural network, employing the Unet structure as its backbone

(Figure 1a). As a fully convolutional network, Unet has shown efficacy in pixel-level tasks such as semantic segmentation (Ronneberger et al., 2015).

The Unet-based model comprises a contraction (encoding) and an expanding (decoding) path, creating a symmetric U-shaped architecture. During contraction, spatial information decreases while feature information increases. The expanding path decodes extracted features into spatial information. The model combines features and spatial information through skip connections, aiding in preserving spatial detail (Drozdzal et al., 2016). The change in feature size is opposite between the contraction and expanding paths. ReLU activation is applied to features in each block. The symmetric structure allows the decoding layers to match their encoding layers, transmitting initial context and texture information for accurate segmentation.

Each encoding or decoding block of ResUnet integrates ResNet-50, a validated model structure for water segmentation (Wagner et al., 2023). ResNet-50 is a well-known convolutional neural network architecture in computer vision tasks, exceling in image recognition and can capture intricate features of images. ResNet-50 addresses vanishing gradient problems through residual blocks, enabling the stable training of multiple-layer based model architecture. With 50 layers, shortcut connections efficiently learn residual functions (Figure 1 (b) and (c)).

The ResUnet can theoretically accomplish water segmentation tasks independently but may struggle with qualified precision, particularly given limited training data or when facing new scenes that are significantly divergent from its training dataset's contexts. The model may not accurately identify the edges of water bodies, thus affecting the precise delineation of water body contours.
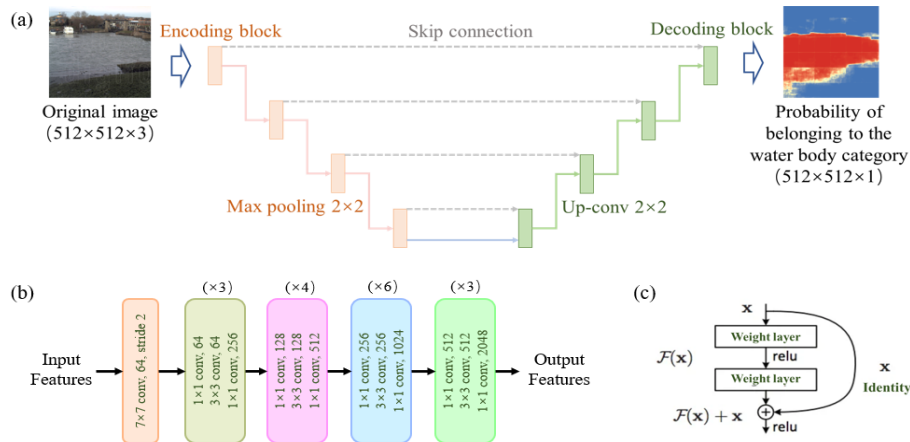

Figure 1. Diagram of (a) the structure of ResUnet model; (b) the structure of the ResNet-50; and (c) the residual operations in the ResNet-50.

2.1.2 Brief introduction to SAM model

SAM is an innovative foundational computer vision model developed by Meta AI for image segmentation. SAM departs from conventional segmentation frameworks by introducing a novel promptable segmentation task, which is facilitated by a prompting-enabled model architecture and a diverse pool of training data. In the model training phase, a *data engine* is used to establish a cyclic process that employs the model for data collection and then exploits the newly gathered data to enhance the model performance. Ultimately, SAM undergoes training on an extensive dataset consisting of over one billion masks extracted from 11 million images.

As shown in Figure 2, SAM comprises three components: an image encoder, a prompt encoder, and a mask decoder. The image encoder, built on the backbone of *ViT*, is pre-trained using the masked autoencoder technique (He et al., 2022). It takes a single image as input and generates image embeddings. The embeddings can be either combined with the prompt encoding output, generated by the prompt encoder that includes dense branches (for masks) and sparse branches (for points, boxes, and texts), or directly passed to the mask decoder for decoding the corresponding masks.

SAM offers support for both automatic *everything* and manual *prompt* modes. The fundamental distinction between the two modes lies in whether SAM uses guided prompts during its segmentation process and whether its resulting segmentation contains specific semantic information. For the former, SAM will automatically generate a series of semantically unknown masks for the image without manual priors. For the latter, users need to provide additional hints to SAM, including boxes, points, and texts. These hints serve to guide SAM in the segmentation process for the expected object.

Given no prior knowledge, SAM cannot accomplish water segmentation tasks independently under either mode. In *everything* mode, although SAM can fully exploit its segmentation capability to accurately segment the image, the semantic understanding of each individual object remains unknown. In *prompt* mode, SAM even requires prior knowledge on water to be provided in the input. Therefore, SAM needs to be used in combination with other domain-specific models either during the segmentation process or in the post-processing of segmentation results.
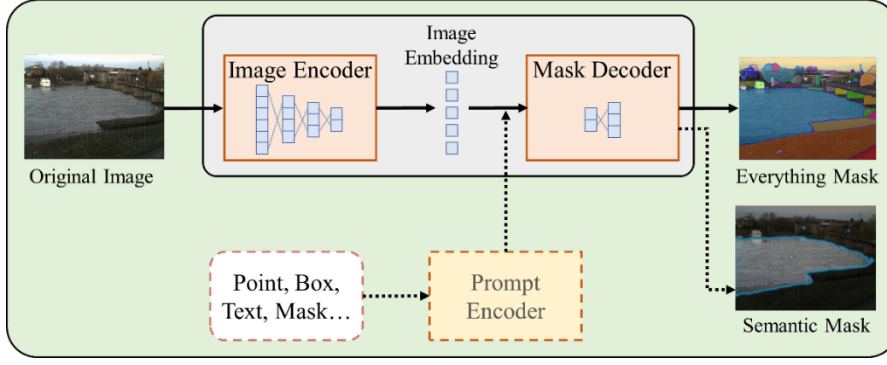
Figure 2. Diagram of the structure of the SAM model. The SAM supports both the *everything* and *prompt* modes for image segmentation.

2.1.3 The integration of ResUnet and SAM

As described in Section 2.1.1 and 2.1.2, in the context of image segmentation, domain-specific models and General AI manifest distinct characteristics. General AI models represented by SAM possess state-of-the-art segmentation capability at the pixel level, while mainly focus on localizing class-agnostic masks. Conversely, domain-specific models such as ResUnet model are tailored to exclusively extract visual semantics and their region-level variants, but limited in precision on edge detection. Integrating the two approaches can potentially lead to an enhancement on segmentation efficiency. Accordingly, this study integrates General AI with domain-specific models to formulate a novel deep learning framework for water segmentation in river images.

As depicted in Figure 3, within the framework, the ResUnet model is pre-trained on a limited dataset that does not include images taken at the river awaiting observation. It then outputs the water probability map and identifies the pixel with the highest water probability. Subsequently, we use SAM as the foundational model to generate water masks. Both SAM modes will be performed for comparison. In the *everything* mode, SAM automates the segmentation of river images into multiple discrete objects, even though their semantics are unknown ("*Everything Mask*"). The coordinates of the most water-like pixel identified by ResUnet are superimposed onto the "*Everything Mask*" segmented by SAM, facilitating the identification of the water object and thereby completing the water segmentation. In the *prompt* mode, the point prompt is adopted. The most-water-like pixel coordinate is fed to SAM along with the original image as inputs, and SAM will directly output the water mask. Finally, the SOFI is calculated based on the segmented water mask, enabling the monitoring of river water levels.
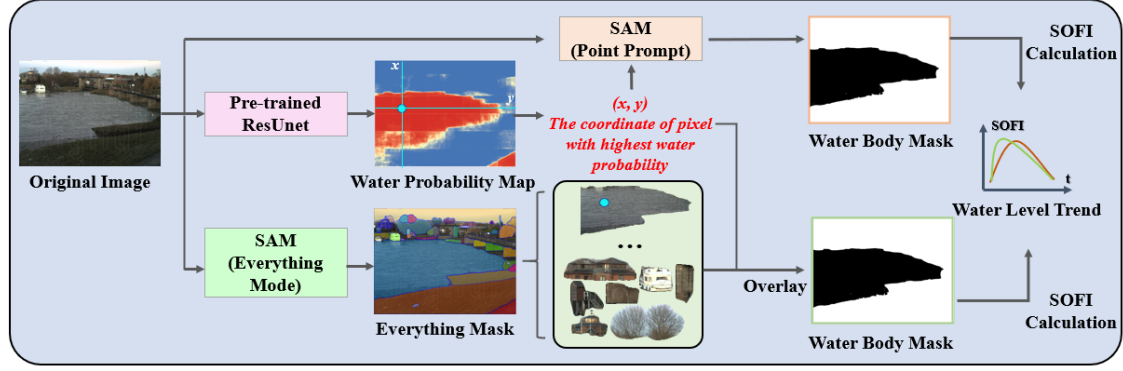
Figure 3. Diagram of the water segmentation framework. The framework is developed by coupling an SAM and a ResUnet model. The ResUnet outputs a water probability map for locating the coordinate of the pixel with highest water probability. The coordinate will then be adopted by SAM as water-related hints for water segmentation in either *everything* or *point prompt* modes.

2.2 Data

2.2.1 Dataset for pre-training ResUnet

The ResUnet model is pre-trained on the RIWA dataset (River Water Segmentation Dataset; Wagner et al., 2023). The dataset offers pixel-wise binary river water segmentation with resolutions of up to 1536×1536 pixels. Comprising a total of 789 training images, 228 validation images, and 111 testing images, RIWA is a compilation of fine-labeled images captured by smartphones, drones, and digital single lens reflex cameras, in addition to suitable images extracted from the Water Segmentation Dataset (Liang et al., 2020). As shown in Figure 4, the images in this dataset encompass various lighting conditions, weather scenarios, and perspectives.

Figure 4. Illustration of example images in the RIWA dataset. Images with widths or heights exceeding 512 are divided into 512×512 sub-patches for model input.

2.2.2 Dataset for the application of ResUnet+SAM framework

The framework was applied to a river camera image dataset collected in the Tewkesbury, UK (Vetra-Carvalho et al., 2020). The dataset comprises images and water level observations acquired from river cameras installed at four spots: Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina, situated along the rivers Avon and Severn in the UK (Figure 5). These observations cover the period between November 21st and December 5th, 2012, during a significant flooding event in the Tewkesbury region. The dataset offers daytime water level data for both River Avon and River Severn, encompassing both the rising and falling limbs of the flood.

The water level values for the four river camera images during this period are also extracted. High-accuracy field-of-view point measurements are utilized for each camera, employing Leica TS 12 (TS) and Leica CS10/CS15 & GS Sensor instruments (GNSS) and Total Station. The dataset includes a total of 141, 136, 144, and 138 images from the Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina, respectively. However, not every image in the dataset is labeled with a water level value, the number of images with associated valid water level records in the four locations is 50, 46, 114, and 138, respectively. The position of water and non-water pixels for each image is manually annotated to serve as the ground truth water masks.
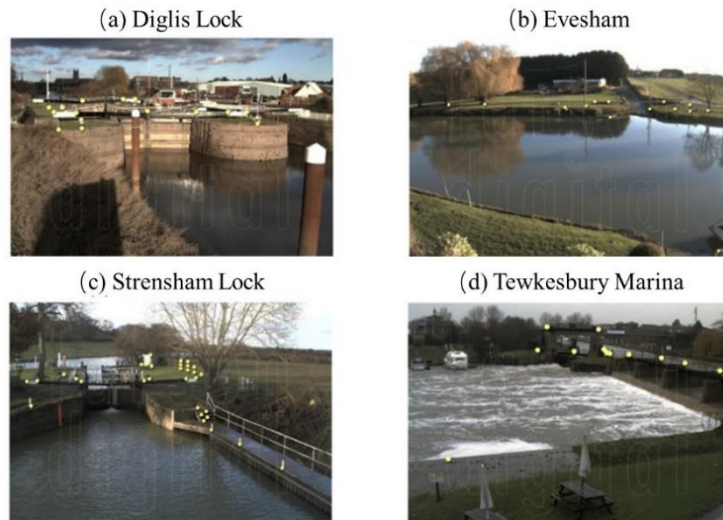

Figure 5. Camera perspectives from (a) Diglis Lock, (b) Evesham, (c) Strensham Lock, and (d) Tewkesbury Marina. The yellow dots depict a selection of measured points within the cameras' field of view in the construction of the original dataset, utilized for extracting water level from the images. The above example images are referenced from Vandaele et al (2021).

2.3 Experimental setup

2.3.1 Model setup

The ResUnet model was pre-trained on the RIWA dataset for 100 epochs using cross entropy (Lecun et al., 2015) as loss function, a learning rate of 0.001, and a batch size of 16. Parameter updates were performed using the training set, and the model parameters from the epoch with the optimal performance on the validation set were chosen as the final parameters.

The official GitHub repository of SAM offers three types of pre-trained models distinguished by varying backbone sizes: *ViT-B*, *ViT-L*, and *ViT-H*. These models' parameter sizes span from small to large. *ViT-H* notably outperforms *ViT-B*, though its increased complexity leads to multiplied testing time. For our research, we chose to adopt *ViT-H* as the encoder to achieve the optimal performance of SAM.

2.3.2 Image preprocessing

Both images and masks in the RIWA dataset have arbitrary sizes. To standardize image inputs for ResUnet, the dataset underwent automated preprocessing to generate squared input samples, all with dimensions of 512×512 pixels. For images larger than 512 pixels in width or height, they and their masks were divided into multiple 512×512 sub-patches (Figure 4). Images with dimensions smaller than 512 pixels in either width or height were resampled to 512 pixels in the corresponding dimension.

In the application of the ResUnet+SAM framework to the river image dataset in Tewkesbury, UK, the initial step also involved dividing the images into 512×512 sub-patches as described above. These sub-patches were then input into the pre-trained ResUnet to calculate the probability of each pixel belonging to the water class. The probability distributions of all sub-patches were combined to determine the position of the pixel with the highest probability in the complete image. Specifically, if different sub-patches had overlapping regions, the probability of the overlapping region was calculated by averaging the probabilities of each sub-patch. However, when SAM was applied to the image to perform segmentation under either *everything* or *point prompt* mode, the image was input to SAM in its entirety, eliminating the need to slice the image into sub-patches. Additionally, during the data quality control process, the size of the sub-patch for SOFI calculation was also configured as 512×512 pixels.

To assess the reliability of the pixel identified as being most water-like by ResUnet, accuracy was introduced as the metric to indicate the proportion of pixels that truly represent water among all the identified most water-like pixels for each of the four locations.

As for the water segmentation task, the intersection over union ratio (IoU), also known as the Jaccard index (Rezatofighi et al., 2019), was used for comparing the water segmentation result (S) to the manually annotation ($\hat{S}$). IoU is computed as:

$$IoU = \frac{1}{n}\left|\frac{S_i \cap \hat{S}_i}{S_i \cup \hat{S}_i}\right| \tag{1}$$

where S_i and $\hat{S}_i$ is the area covered by water in a segmented image and corresponding ground truth water mask, respectively. The index varies from 0% to 100% to represent complete misclassification to perfect classification.

Meanwhile, to further compare the model's ability to identify water bodies at different pixel coordinates within images, accuracy was used again but to indicate the proportion of times a pixel coordinate is correctly recognized as water body by models, out of all the times this pixel truly belongs to water body across different images taken in the same location.

Moreover, the Spearman correlation coefficient and Pearson correlation coefficient (de Winter et al., 2016), were applied to images affiliated with ground truth water level data in the four locations. These coefficients were used to evaluate and describe the extent of correlation between the estimated SOFI values and the scalar values of water level.

3. Results

3.1 The performance on most water-like pixel identification

Firstly, we investigates the distribution patterns of most water-like pixels identified by ResUnet, and whether the identified pixels correspond to actual water. This serves as the premise for ResUnet to provide effective water-related hints for SAM.

Figure 6 provides a visualization of the spatial distribution of pixels most resembling water, as identified by the ResUnet model in images captured at the four different locations by river cameras. The illustration highlights that, across different moments in

time, pixels identified as closely resembling water exhibited a clustering pattern, with their clustering centers shifting in response to fluctuations in river water levels. Moreover, these pixels mainly clustered within the central regions of the water bodies rather than at the interfaces between water and the non-water background. Therefore, the confidence of the single pre-trained ResUnet model in discerning water within the interior of river channels is higher, while its efficacy at the water body's periphery cannot be guaranteed.
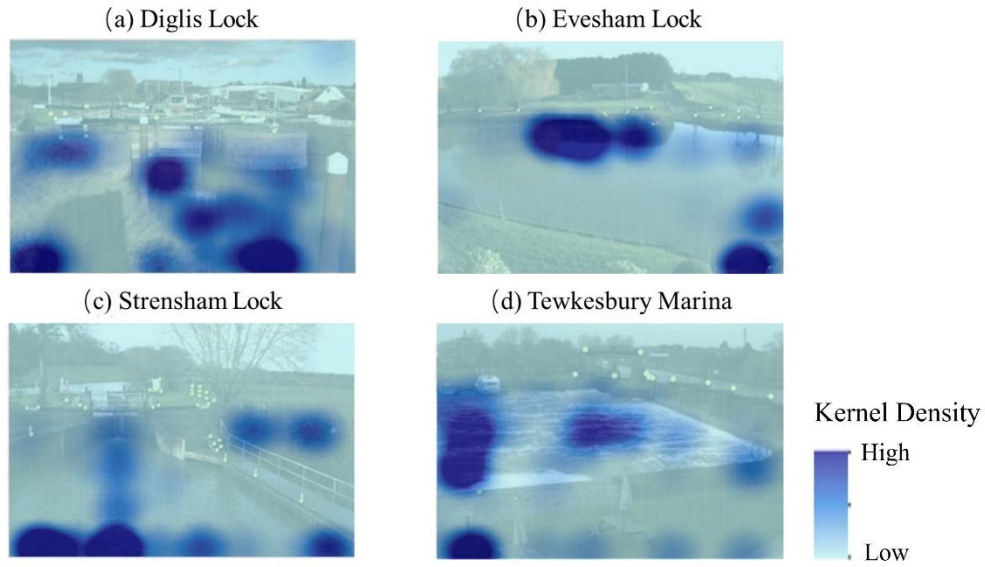

Figure 6. The kernel density of pixels with the highest probability of belonging to water identified by the ResUnet model in river camera images taken at different time points at the four locations.

Figure 7 further statistically analyzes the probability values associated with pixels identified by the pre-trained ResUnet model as having the highest likelihood of belonging to water. Across all four locations, the majority of the identified pixels exhibited a probability of being water exceeding 0.95, and the accuracy that the pixel identified as the most water-like pixel is truely a water pixel exceeded 90%. The identified pixels can effectively represent water bodies.
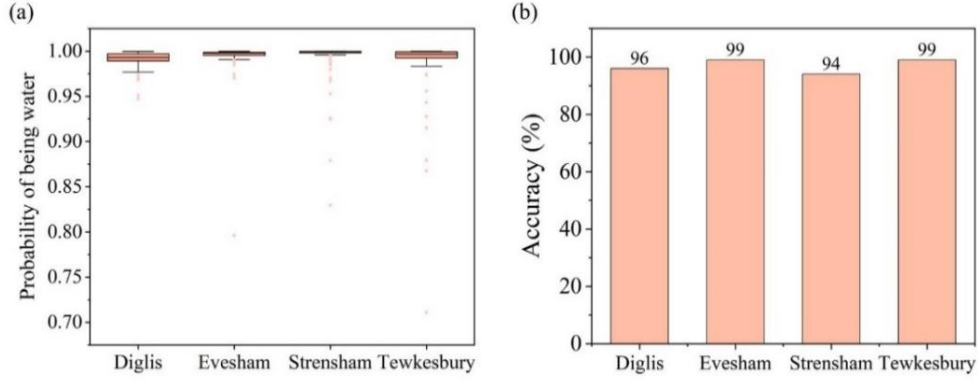
Figure 7. (a) The box plot depicting the probability values corresponding to the pixel with the highest probability of belonging to water among different images at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina. The upper and lower boundary of the box represent the upper (0.75) and lower quartile (0.25), the solid line represents the median, the whiskers extend to 1.5 times the interquartile range, and the dots are outliers. (b) The accuracy that the pixel with the highest probability of belonging to water is truly a water pixel.

3.2 The performance on water segmentation

Given the reliable coordinates of the pixel with the highest water probability provided by ResUnet, SAM can use it to filter the corresponding object from the *Everything Mask* under *everything* mode or take it as inputs for guiding subsequent segmentation under *point prompt* mode, and has produced the corresponding water masks.

Compared with the single pre-trained model, the framework's superior performance under either *everything* mode or *point prompt* mode on water segmentation is demonstrated. As depicted in Figure 8, at Diglis Lock, Evesham, and Strensham Lock, the ResUnet+SAM framework under both the two different modes consistently outperformed the single ResUnet model with statistically significant superiorities in terms of IoU values. The statistical significance was verified by the Analysis of Variance (ANOVA), with p-value less than 0.01 given the confidence level of 95%. At these three locations, the median IoU values for individual images all exceeded 0.95.

At Tewkesbury Marina, the advantage of the ResUnet+SAM framework under *everything* mode over the ResUnet model was less pronounced. This is attributed to SAM conflating water bodies and wet embarkment ground as a unified object in some images. However, the median IoU value of the framework for single images was still close to 1. Meanwhile, predictions with very low IoU values (<0.5) were also fewer than those

produced by the single ResUnet model. While under the *point prompt* mode, the framework continued to demonstrate considerable superiority over the ResUnet model. Unlike under *everything* mode, it did not exhibit an inclination to overestimate the water body area within the image, thereby facilitating a more precise water segmentation process.

The findings across the four locations collectively suggest that the integration of ResUnet and SAM can refine the water segmentation process and has generated more reliable water masks. Moreover, the segmentation results obtained under *point prompt* mode are slightly better than those achieved under *everything* mode. On average, there is a 15% and 16% improvement under *everything* mode and *point prompt* mode, respectively, compared to ResUnet.
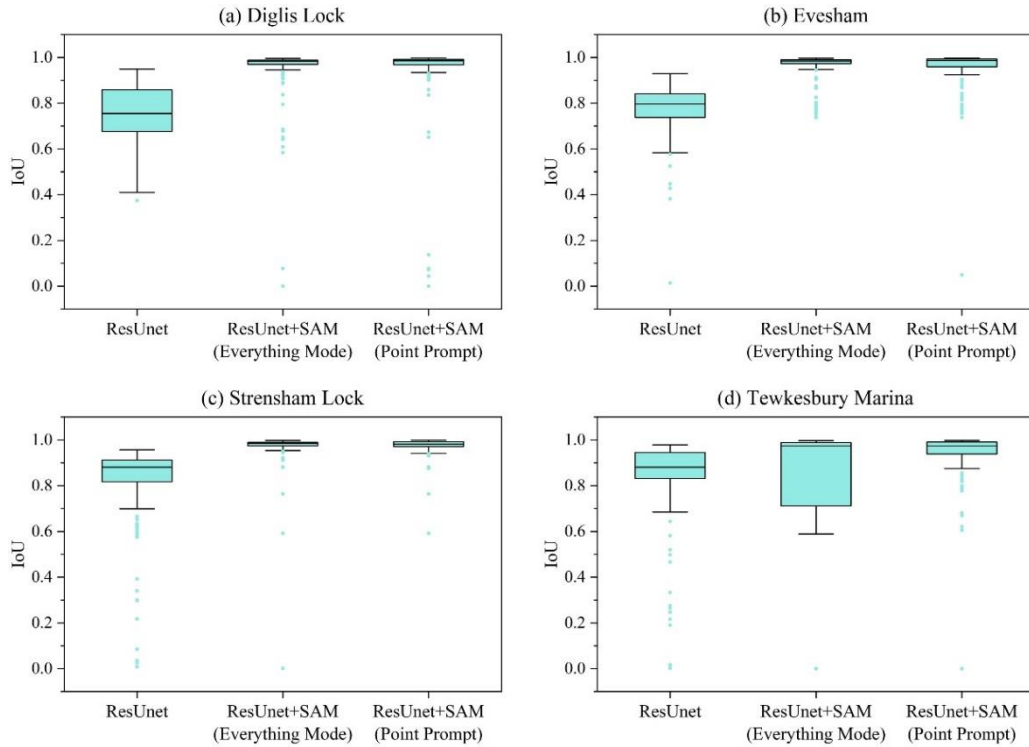

Figure 8. The IoU values for water body segmentation by the ResUnet+SAM framework under *everything* mode and *point prompt* mode, as well as the ResUnet model at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina.

Figure 9 further compares the accuracy achieved by the ResUnet model and the ResUnet+SAM framework under different modes across various pixel coordinates within the images. As for the single ResUnet model, it performed well on discerning water pixels within the central regions of the water bodies but encountered challenges in accurately

segmenting water pixels in transitional zones between water and non-water pixels, notably around the pillar at Diglis Lock, as well as the objects along the riverbank in the other three locations. Consequently, ResUnet falls short in delineating the contours of the water bodies with the same level of precision as the coupling framework. In contrast, no matter under *everything* mode or *point prompt* mode, the ResUnet+SAM framework consistently presented a high degree of accuracy at not only pixels residing within the water bodies, but also those positioned along the interfaces between the water and adjacent elements, such as river banks, trees, pillars, and other background features.
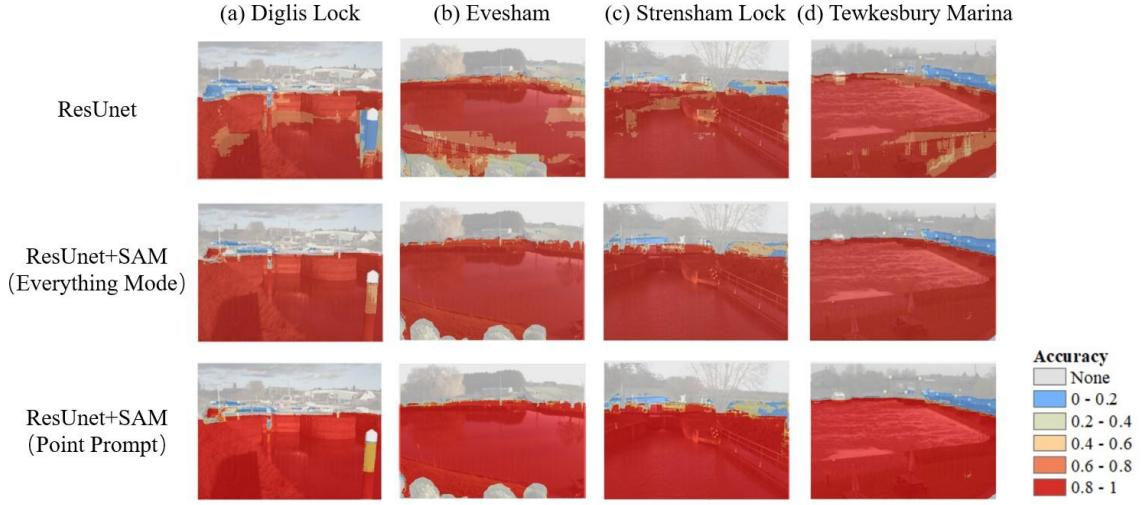

Figure 9. The accuracy achieved by the ResUnet+SAM framework and the ResUnet across various pixel coordinates within the river camera images for the four locations.

In Figure 10, three example images with varying water levels at each of the four locations, along with their corresponding water masks, as well as the water segmentation results achieved by different methods are visualized. These examples can be considered to demonstrate relatively high predictive performance within each method across all images. It can be observed that the single ResUnet model struggled to extract water pixels as accurately as the ResUnet+SAM framework. Similar to previous findings illustrated in Figure 10, in some images, the sky was misclassified as water due to its blue appearance, while in other cases, damp ground and the reflections of trees or pillars on the water surface led to misclassification by ResUnet, causing the water areas covered by reflections to be unrecognized. However, these interfering factors have not significantly impacted water segmentation when using the ResUnet+SAM framework. For each example image, the framework has effectively captured the water body outlines, achieving precise segmentation of water pixels.
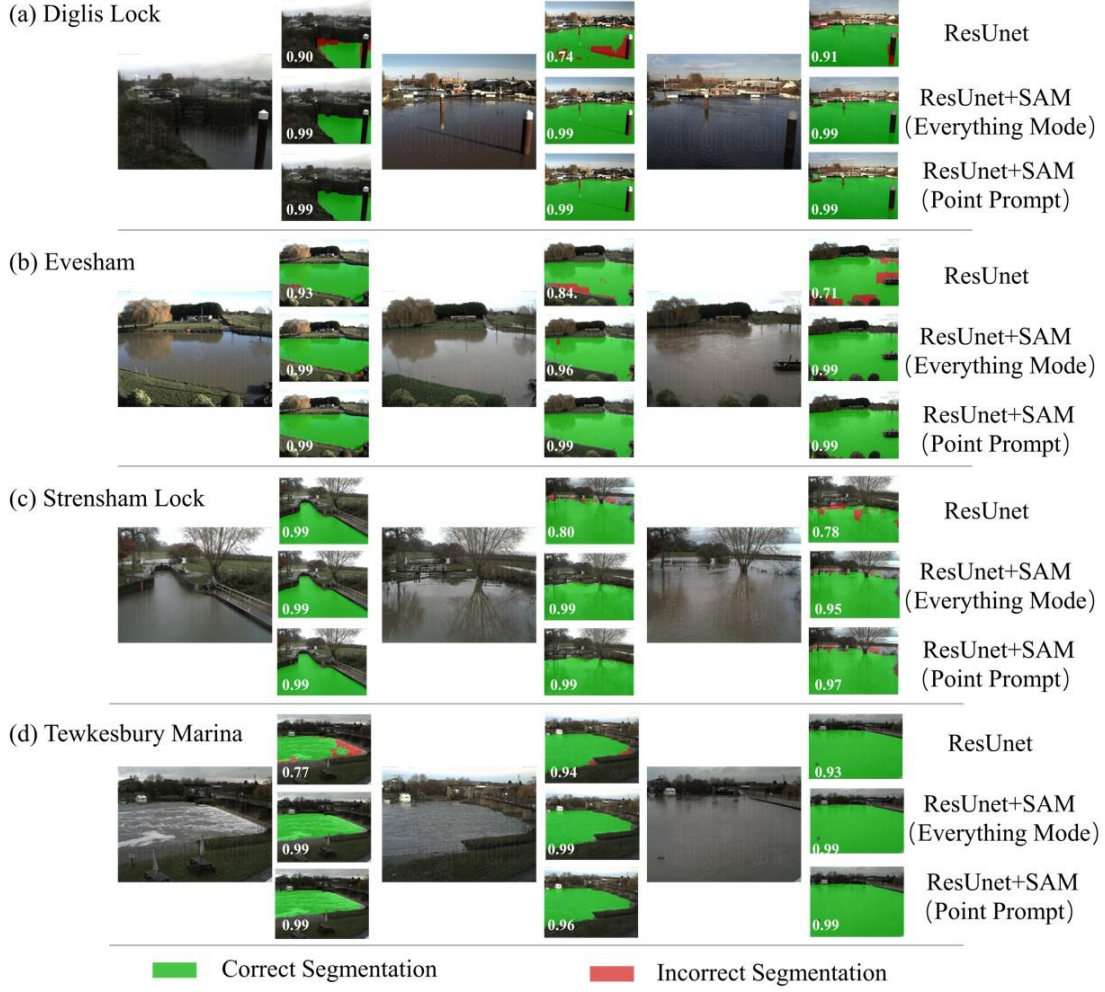
Figure 10. Illustrative examples of water segmentation results of ResUnet+SAM and ResUnet at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina. The IoU for each mask was attached.

However, as depicted above, significant mis-segmentation can occur under the *everything* mode, even worse than ResUnet, especially in Tewkesbury Marina. Figure 11 presents a specific instance of significant segmentation errors by the ResUnet+SAM framework under the *everything* mode. The image depicts a scenario where the riverbank, situated at the lower portion, is obscured by shadows, rendering it featureless and appearing entirely dark. Consequently, SAM, operating under the *everything* mode, erroneously identified this shadowed area as part of the water surface. However, the true delineation between the actual water surface and the shadowed region is actually distinct, with even single ResUnet model capable of accurately delineating its edges. While under the *point prompt* mode, SAM performed effectively in water segmentation for this scenario. Taking point as inputs has made a positive impact on guiding segmentation.
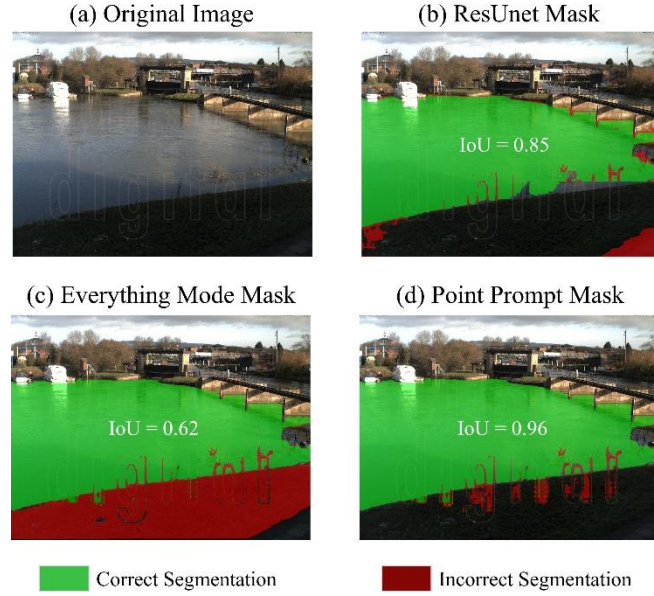
Figure 11. Examples of significant segmentation errors made by the ResUnet+SAM framework under the *everything* mode, along with segmentation results under *point prompt* mode or from ResUnet.

3.3 The performance on water level trend monitoring

The inception of this study is to employ river camera image sequences to monitor the trend in water levels. Figure 12 presents a comparative analysis of the trend in SOFI time series contrasted against the actual water level time series at four locations.

The figure illustrates that, across Diglis Lock, Evesham, and Strensham Lock, the SOFI variations obtained using the ResUnet+SAM framework under both *everything* mode and *point prompt* mode closely aligned with the actual water level fluctuations. The alignment remained consistent irrespective of high or low water levels, effectively capturing the water level dynamics. Notably at Strensham Lock, where water level changes were substantial, the SOFI derived from the framework still matched the water levels, accurately capturing even minor variations. In contrast, the SOFI values calculated based on the water segmentation results given by the ResUnet model exhibit significant variability, with a notable presence of extreme erroneous values. Many of these values tended to underestimate the water level due to insufficient segmentation of water bodies in the images, thereby compromising the accurate depiction of actual water level trends.

At Tewkesbury Marina, the framework still demonstrated its ability to accurately depict water level trends under *point prompt* mode, providing predictions that were relatively accurate in magnitude for extremely high and low water levels. However, under

the *everything* mode, the framework tended to overestimate water levels at specific intervals, thereby producing SOFI anomalies.
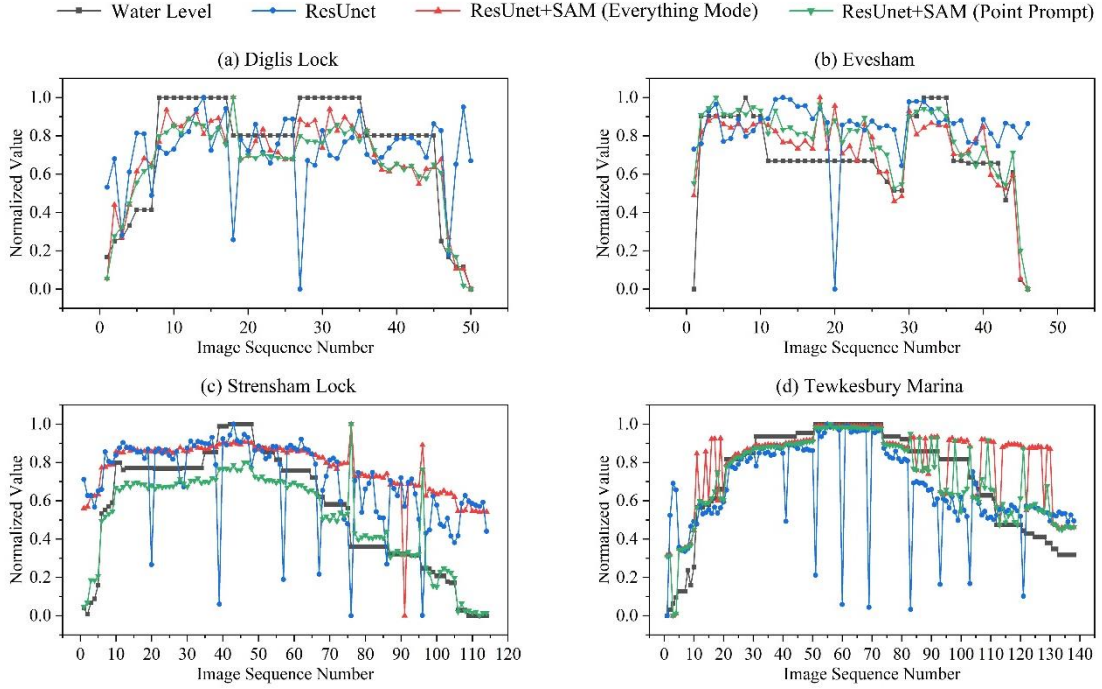

Figure 12. The trends of the actual water levels and the SOFI time series derived from both the ResUnet+SAM framework and the ResUnet model at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina. The images from different locations are sorted by time, but the time intervals between individual images are not uniform.

From quantitative perspective, the Spearman correlation coefficient and Pearson correlation coefficient between the SOFI time series obtained from the ResUnet+SAM framework and the actual water level time series under both modes were superior to those of the ResUnet model across all locations (Table 1). The superiority is more pronounced under *point prompt* mode: the Pearson correlation coefficient exceeded 0.90 in most locations, and the Spearman correlation coefficient even reached 0.94 at Tewkesbury Marina.

These results quantitatively demonstrate the strong correlation between the SOFI time series obtained from the ResUnet+SAM framework and the actual water level time series, confirming the earlier assertion from Figure 12. Hence, the framework based on the integration of ResUnet and SAM practically facilitates the monitoring of the temporal water level trend, especially under *point prompt* mode.

19

Table 1. The Spearman correlation coefficient and Pearson correlation coefficient between the SOFI time series obtained using the ResUnet+SAM framework under two different modes, the ResUnet model, and the actual water level time series at various locations

	Spearman correlation coefficient			Pearson correlation coefficient		
	ResUnet	Everything Mode	Point Prompt	ResUnet	Everything Mode	Point Prompt
Diglis Lock	0.28	0.84	0.87	0.26	0.88	0.90
Evesham	0.39	0.78	0.88	0.18	0.83	0.87
Strensham Lock	0.74	0.92	0.92	0.58	0.83	0.93
Tewkesbury Marina	0.76	0.78	0.94	0.59	0.78	0.92

4. Discussion

4.1 The value of the water segmentation-based river water level observations

If local topography situation is available, the water mask segmented by the framework can be transformed and overlaid with the riverbed terrain to derive scalar water level values, which can support real-time observations of river flow conditions. However, obtaining fine-scale terrain data is often challenging. Despite the limitation in underlying surface data, the water mask remains valuable for calibrating hydrological models through calculating SOFI based on it.

The calibration efficacy of hydrological models depends on the water segmentation algorithm's performance, with a higher correlation between SOFI and actual water level values enhancing hydrological model calibration. According to the scenarios analysis conducted by Moy de Vitry & Leitão (2020), with correlation coefficients no less than 0.6 between SOFI time series and real water level values, hydrological models could be calibrated to achieve significantly higher predictive level than uncalibrated benchmark. The SOFI sequences generated by the ResUnet+SAM framework under different modes can achieve correlation coefficients of over 0.8 for most regions in this study, thus are informative for hydrologic model calibration.

Especially under *point prompt* mode, the framework's SOFI time series display negligible systematic errors, avoiding overall overestimation or underestimation of water levels, with minimal and random errors. Based on the research by Moy de Vitry & Leitão (2020) and Ilja Van Meerveld et al. (2017), a greater degree of randomness in the error

distribution of water level class is more advantageous for model calibration when errors are minor. This phenomenon can be attributed to the compensating effect of the number of observations and their accuracy, as the random errors will average out when a sufficient number of observations are utilized. Therefore, SOFI under *point prompt* mode is more robust to practically support hydrological modeling.

4.2 The compromise between the Narrow AI and General AI in imaging-based hydrological monitoring

AI is gradually progressing from the era of Narrow AI towards the era of General AI (Bundy, 2017). During the Narrow AI stage, researchers aim to develop AI models with highly specialized intelligence in specific domains, while in the General AI stage, the potential for large models pre-trained on web-scale datasets to revolutionize computer vision with robust zero-shot and few-shot generalization capabilities has emerged. Taking SAM as an example, it possesses the capacity to segment distinct entities in any given image, as long as no less than two separate objects exist. However, similar to other General AI, SAM cannot provide insights into the identity of a segmented object without prior hints. Hence, for conducting downstream tasks such as water segmentation, prompt engineering becomes essential. Typically, the task of providing prompts is executed by humans, following predefined guidelines (Liu & Chilton, 2022). In our study, we automate the entire process by leveraging General AI (SAM) as the foundational component while employing Narrow AI (ResUnet) as a prompter.

Furthermore, as mentioned above, different prompt modes, i.e., the ways of integrating prior knowledge provided by domain-specific models, also influence the performance of General AI. Approaches such as those used in the *everything* mode for our study, which apply post-processing on General AI results, may not fully inspire General AI's image understanding capabilities. This conclusion aligns with findings in other fields such as medical image analysis (Huang et al., 2024), indicating that for General AI, or at least for SAM, it is not yet capable of flawlessly segmenting each object in an image independently. In contrast, incorporating prior knowledge directly into the segmentation process can guide the General AI to achieve conditional segmentation, resulting in more accurate outcomes, thus representing a more optimal prompt mode.

The framework proposed in this study also holds the potential for application to other

imaging-based hydrological monitoring tasks that are currently constrained by the availability of annotated data. Examples include water quality monitoring or the detection of floating debris on the water surface (Ramírez et al., 2023; Solé Gómez et al., 2022). The combination paradigm maximizes the utilization of costs already consumed in General AI development, while compressing the marginal costs associated with developing domain-specific models for downstream tasks. The developers of the General AI, often large corporations like *Meta AI* and *OpenAI* who can access substantial datasets and computational power resources, have shouldered the burden of training big foundation models for the public. The time and learning costs associated with the creation of standardized datasets and model selection by individual users in domain-specific model development can be substantially reduced. This facilitates a more convenient utilization of AI tools for non-computer science professionals, as exemplified by hydrologists in this study.

4.3 Future work

This study introduces a transferrable solution for monitoring river water level trends. However, adaptive strategies such as transfer learning, which were previously considered for adaptability to new locations, are not rendered completely obsolete. In fact, SAM can be enhanced through fine-tuning only two parameters via one-shot learning, facilitated by providing a single labeled image-mask pair (Zhang et al., 2023). This refinement may aid in overcoming challenges related to the confusion between water bodies and wet ground occurred in Tewkesbury Marina under *everything* mode. Furthermore, with a future consideration for deployment on mobile applications or real-time usage, the exploration of lightweight versions of SAM, such as EfficientSAM (Xiong et al., 2023), can be pursued to establish a more energy-efficient observational paradigm.

Besides *everything* mode and *point prompt* mode, SAM can also be prompted in other forms such as mask and text. Future work should continue to perform in-depth comparisons among more prompt forms and address the hydrodynamic conditions of various river types to determine optimal prompt configurations. Different surface flow characteristics may have varying requirements for the form and magnitude of prompts. The exploration of effective strategies for prompting General AI can contribute to more adaptive imaging-based hydrological monitoring across different rivers.

Moreover, in subsequent research, the framework proposed in this study will be extended to regions with more pronounced variations in seasonal conditions, lighting backgrounds, and other factors, for evaluation. Also, we will select appropriate locations to construct hydrological models like HBV (Seibert & Bergström, 2022) and utilize the SOFI sequences derived from our proposed framework for model calibration, aiming to elevate the local hydrological forecasting capabilities, facilitating improved understanding of the hydrological process. If fine-scale terrain data is available, the water segmentation results can also be used for overlaying analysis to infer scalar river water level, which can then be compared with the data recorded by gauging stations.

5. Conclusions

In this study, we propose a novel transferable deep learning framework that combines General AI (SAM) with a domain-specific model (ResUnet pre-trained on a non-local river image dataset) for water segmentation and water level trend monitoring. The framework was implemented in four different riverside locations in Tewkesbury, UK, and compared with the single ResUnet model. Moreover, different modes of SAM, including the *everything* mode and *point prompt* mode, were both adopted and contrasted.

Our results indicated the transferability of the ResUnet+SAM framework for water segmentation, with an average improvement of 15% and 16% in IoU under the *everything* mode and *point prompt* mode, respectively, compared to ResUnet. Requiring no local data annotation or model parameter fine-tuning, the proposed framework accurately identified water pixels in the images and delineated the water body's outline. Moreover, utilizing SAM under the *point prompt* mode represents a superior strategy for the fusion of prior knowledge provided by domain-specific models. For all four locations, the spearman correlations between the SOFI derived by the framework under *point prompt* mode and the actual water level exceeded 0.87, and the errors were randomly distributed. Regarding both the error magnitude and error distribution pattern, the segmentation results can practically support the monitoring of water level trend.

Overall, this study establishes a transferable imaging-based water level trend monitoring paradigm through the use of Narrow AI and General AI in tandem, substantially lowering the requirement for localized data annotation and model deployment. Future work is recommended to adopt one-shot learning or other forms of prompts to adapt the framework to more diverse and complex monitoring conditions.

Meanwhile, the data processed by the framework will be further integrated with hydrological models to evaluate its enhancement of hydrological forecasting performance or overlaid with riverbed terrain data to infer scalar water level values.

Code and data availability statement

1. The links to the river image datasets were given in the paper.

2. ResUnet and its pre-trained parameters on ImageNet dataset can be directly used within the *PyTorch* framework using the *segmentation_models_pytorch* module:

https://github.com/qubvel/segmentation_models.pytorch

3. The code of SAM and its pre-trained parameters can be accessed in the Github repository through the following link:

https://github.com/facebookresearch/segment-anything

Disclosure statement

The authors declare that they have no conflict of interest.

References

Akiyama, T. S., Junior, J. M., Gonçalves, W. N., Carvalho, M. de A., & Eltner, A. (2021). Evaluating Different Deep Learning Models for Automatic Water Segmentation. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4716–4719. https://doi.org/10.1109/IGARSS47720.2021.9553345

Akiyama, T. S., Marcato Junior, J., Gonçalves, W. N., Bressan, P. O., Eltner, A., Binder, F., & Singer, T. (2020). DEEP LEARNING APPLIED TO WATER SEGMENTATION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLIII-B2-2*, 1189–1193. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1189-2020

Bundy, A. (2017). Preparing for the future of Artificial Intelligence. *AI & SOCIETY*, *32*(2), 285–287. https://doi.org/10.1007/s00146-016-0685-0

de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, *21*(3), 273–290. https://doi.org/10.1037/met0000079

Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C. (2016). The importance of skip connections in biomedical image segmentation. In *Deep learning and data labeling for medical applications* (pp. 179–187). Springer.

Eltner, A., Bressan, P. O., Akiyama, T., Gonçalves, W. N., & Marcato Junior, J. (2021). Using Deep Learning for Automatic Water Stage Measurements. *Water Resources Research*, *57*(3). https://doi.org/10.1029/2020WR027608

Eltner, A., Elias, M., Sardemann, H., & Spieler, D. (2018). Automatic Image-Based Water Stage Measurement for Long-Term Observations in Ungauged Catchments. *Water Resources Research*, *54*(12), 10,362-10,371. https://doi.org/10.1029/2018WR023913

Erfani, S. M. H., Wu, Z., Wu, X., Wang, S., & Goharian, E. (2022). ATLANTIS: A benchmark for semantic segmentation of waterbody images. *Environmental Modelling & Software*, *149*, 105333. https://doi.org/https://doi.org/10.1016/j.envsoft.2022.105333

Etter, S., Strobl, B., Seibert, J., & van Meerveld, H. J. I. (2020). Value of Crowd-Based Water Level Class Observations for Hydrological Model Calibration. *Water Resources Research*, *56*(2), 1–17. https://doi.org/10.1029/2019WR026108

Fekete, B. M., Looser, U., Pietroniro, A., & Robarts, R. D. (2012). Rationale for monitoring discharge on the ground. *Journal of Hydrometeorology*, *13*(6), 1977–1986. https://doi.org/10.1175/JHM-D-11-0126.1

Grimaldi, S., Li, Y., Pauwels, V. R. N., & Walker, J. P. (2016). Remote Sensing-Derived Water Extent and Level to Constrain Hydraulic Flood Forecasting Models: Opportunities and Challenges. *Surveys in Geophysics*, *37*(5), 977–1034. https://doi.org/10.1007/s10712-016-9378-y

Gupta, A., Chang, T., Walker, J., & Letcher, B. (2022). Towards Continuous Streamflow Monitoring with Time-Lapse Cameras and Deep Learning. *Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, 353–363. https://doi.org/10.1145/3530190.3534805

He, K., Chen, X., Xie, S., Li, Y., Dollar, P., & Girshick, R. (2022). Masked Autoencoders Are Scalable Vision Learners. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2022-June*,

15979–15988. https://doi.org/10.1109/CVPR52688.2022.01553

Huang, Y., Yang, X., Liu, L., Zhou, H., Chang, A., Zhou, X., Chen, R., Yu, J., Chen, J., Chen, C., Liu, S., Chi, H., Hu, X., Yue, K., Li, L., Grau, V., Fan, D.-P., Dong, F., & Ni, D. (2024). Segment anything model for medical images? *Medical Image Analysis*, *92*, 103061. https://doi.org/https://doi.org/10.1016/j.media.2023.103061

Ilja Van Meerveld, H. J., Vis, M. J. P., & Seibert, J. (2017). Information content of stream level class data for hydrological model calibration. *Hydrology and Earth System Sciences*, *21*(9), 4895–4905. https://doi.org/10.5194/hess-21-4895-2017

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

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(13), 3521–3526. https://doi.org/10.1073/pnas.1611835114

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Liang, Y., Jafari, N., Luo, X., Chen, Q., Cao, Y., & Li, X. (2020). WaterNet: An adaptive matching pipeline for segmenting water with volatile appearance. *Computational Visual Media*, *6*(1), 65–78. https://doi.org/10.1007/s41095-020-0156-x

Liu, V., & Chilton, L. B. (2022). Design Guidelines for Prompt Engineering Text-to-Image Generative Models. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3491102.3501825

Lo, S. W., Wu, J. H., Lin, F. P., & Hsu, C. H. (2015). Visual sensing for urban flood monitoring. *Sensors (Switzerland)*, *15*(8), 20006–20029. https://doi.org/10.3390/s150820006

Lopez-Fuentes, L., Rossi, C., & Skinnemoen, H. (2017). River segmentation for flood monitoring. *2017 IEEE International Conference on Big Data (Big Data)*, 3746–3749. https://doi.org/10.1109/BigData.2017.8258373

Mosley, L. M. (2015). Drought impacts on the water quality of freshwater systems; review and integration. *Earth-Science Reviews*, *140*, 203–214. https://doi.org/https://doi.org/10.1016/j.earscirev.2014.11.010

Moy De Vitry, M., Kramer, S., Dirk Wegner, J., & Leitao, J. P. (2019). Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network. *Hydrology and Earth System Sciences*, *23*(11), 4621–4634. https://doi.org/10.5194/hess-23-4621-2019

Moy de Vitry, M., & Leitão, J. P. (2020). The potential of proxy water level measurements for calibrating urban pluvial flood models. *Water Research*, *175*. https://doi.org/10.1016/j.watres.2020.115669

Noto, S., Tauro, F., Petroselli, A., Apollonio, C., Botter, G., & Grimaldi, S. (2022). Low-cost stage-camera system for continuous water-level monitoring in ephemeral streams. *Hydrological Sciences Journal*, *67*(9), 1439–1448. https://doi.org/10.1080/02626667.2022.2079415

Perks, M. T., Dal Sasso, S. F., Hauet, A., Jamieson, E., Le Coz, J., Pearce, S., Peña-Haro, S., Pizarro, A., Strelnikova, D., Tauro, F., Bomhof, J., Grimaldi, S., Goulet, A., Hortobágyi, B., Jodeau, M., Käfer, S., Ljubičić, R., Maddock, I., Mayr, P., …

Manfreda, S. (2020). Towards harmonisation of image velocimetry techniques for river surface velocity observations. *Earth System Science Data*, *12*(3), 1545–1559. https://doi.org/10.5194/essd-12-1545-2020

Ramírez, S. B., van Meerveld, I., & Seibert, J. (2023). Citizen science approaches for water quality measurements. *Science of The Total Environment*, *897*, 165436. https://doi.org/https://doi.org/10.1016/j.scitotenv.2023.165436

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 658–666. https://doi.org/10.1109/CVPR.2019.00075

Richey, J. E., Melack, J. M., Aufdenkampe, A. K., Ballester, V. M., & Hess, L. L. (2002). Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric $CO2$. *Nature*, *416*(6881), 617–620. https://doi.org/10.1038/416617a

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR*, *abs/1505.0*. http://arxiv.org/abs/1505.04597

Ruhi, A., Messager, M. L., & Olden, J. D. (2018). Tracking the pulse of the Earth's fresh waters. *Nature Sustainability*, *1*(4), 198–203. https://doi.org/10.1038/s41893-018-0047-7

Sabbatini, L., Palma, L., Belli, A., Sini, F., & Pierleoni, P. (2021). A Computer Vision System for Staff Gauge in River Flood Monitoring. In *Inventions* (Vol. 6, Issue 4). https://doi.org/10.3390/inventions6040079

Seibert, J., & Bergström, S. (2022). A retrospective on hydrological catchment modelling based on half a century with the HBV model. *Hydrology and Earth System Sciences*, *26*(5), 1371–1388. https://doi.org/10.5194/hess-26-1371-2022

Seibert, J., & Vis, M. J. P. (2016). How informative are stream level observations in different geographic regions? *Hydrological Processes*, *30*(14), 2498–2508. https://doi.org/https://doi.org/10.1002/hyp.10887

Sermet, Y., & Demir, I. (2023). Camera-based intelligent stream stage sensing for decentralized environmental monitoring. *Journal of Hydroinformatics*, *25*(2), 163–173. https://doi.org/10.2166/hydro.2023.032

Solé Gómez, À., Scandolo, L., & Eisemann, E. (2022). A learning approach for river debris detection. *International Journal of Applied Earth Observation and Geoinformation*, *107*, 102682. https://doi.org/https://doi.org/10.1016/j.jag.2022.102682

Spasiano, A., Grimaldi, S., Nardi, F., Noto, S., & Braccini, A. M. (2023). Testing the theoretical principles of citizen science in monitoring stream water levels through photo-trap frames. *Frontiers in Water*, *5*. https://doi.org/10.3389/frwa.2023.1050378

Tauro, F., Selker, J., van de Giesen, N., Abrate, T., Uijlenhoet, R., Porfiri, M., Manfreda, S., Caylor, K., Moramarco, T., Benveniste, J., Ciraolo, G., Estes, L., Domeneghetti, A., Perks, M. T., Corbari, C., Rabiei, E., Ravazzani, G., Bogena, H., Harfouche, A., … Grimaldi, S. (2018). Measurements and Observations in the XXI century (MOXXI): innovation and multi-disciplinarity to sense the hydrological cycle. *Hydrological Sciences Journal*, *63*(2), 169–196. https://doi.org/10.1080/02626667.2017.1420191

Vandaele, R., Dance, S. L., & Ojha, V. (2023). Calibrated river-level estimation from

river cameras using convolutional neural networks. *Environmental Data Science*, 2, e11. https://doi.org/DOI: 10.1017/eds.2023.6

Vandaele, R., Dance, S., & Ojha, V. (2021). Deep learning for the estimation of water-levels using river cameras. *Hydrology and Earth System Sciences Discussions*, 1–29. https://doi.org/10.5194/hess-2021-20

Vanden Boomen, R. L., Yu, Z., & Liao, Q. (2021). Application of Deep Learning for Imaging-Based Stream Gaging. *Water Resources Research*, *57*(11), e2021WR029980. https://doi.org/https://doi.org/10.1029/2021WR029980

Vetra-Carvalho, S., Dance, S. L., Mason, D. C., Waller, J. A., Cooper, E. S., Smith, P. J., & Tabeart, J. M. (2020). Collection and extraction of water level information from a digital river camera image dataset. *Data in Brief*, *33*, 106338. https://doi.org/https://doi.org/10.1016/j.dib.2020.106338

Wagner, F., Eltner, A., & Maas, H.-G. (2023). River water segmentation in surveillance camera images: A comparative study of offline and online augmentation using 32 CNNs. *International Journal of Applied Earth Observation and Geoinformation*, *119*, 103305. https://doi.org/https://doi.org/10.1016/j.jag.2023.103305

Weeser, B., Jacobs, S., Kraft, P., Rufino, M. C., & Breuer, L. (2019). Rainfall-Runoff Modeling Using Crowdsourced Water Level Data. *Water Resources Research*, *55*(12), 10856–10871. https://doi.org/https://doi.org/10.1029/2019WR025248

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Whitworth, K. L., Baldwin, D. S., & Kerr, J. L. (2012). Drought, floods and water quality: Drivers of a severe hypoxic blackwater event in a major river system (the southern Murray–Darling Basin, Australia). *Journal of Hydrology*, *450–451*, 190–198. https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.04.057

Xiong, Y., Varadarajan, B., Wu, L., Xiang, X., Xiao, F., Zhu, C., Dai, X., Wang, D., Sun, F., Iandola, F., & others. (2023). EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything. *ArXiv Preprint ArXiv:2312.00863*.

Yamazaki, D., Lee, H., Alsdorf, D. E., Dutra, E., Kim, H., Kanae, S., & Oki, T. (2012). Analysis of the water level dynamics simulated by a global river model: A case study in the Amazon River. *Water Resources Research*, *48*(9). https://doi.org/https://doi.org/10.1029/2012WR011869

Yan, K., Di Baldassarre, G., Solomatine, D. P., & Schumann, G. J.-P. (2015). A review of low-cost space-borne data for flood modelling: topography, flood extent and water level. *Hydrological Processes*, *29*(15), 3368–3387. https://doi.org/https://doi.org/10.1002/hyp.10449

Zhang, R., Jiang, Z., Guo, Z., Yan, S., Pan, J., Dong, H., Gao, P., & Li, H. (2023). Personalize segment anything model with one shot. *ArXiv Preprint ArXiv:2305.03048*.