Toward Transferable Water Level Trend Monitoring Using River Cameras: Integrating Domain-Specific Models and Segment Anything Model (SAM) for 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 General AI and domain-specific models.

- An unsupervised clustering method is proposed for quality control on qualitative water level data.

Keywords:

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

Abstract

Water level data is critical for hydrologic model calibration. The extensive river camera networks, in conjunction with advanced deep learning techniques, form the foundation for the imaging-based monitoring of water level trends. However, limited annotated data and tedious local deployment restricts the applicability of deep learning models in new river scenarios. This study proposes a novel transferable deep learning framework by combining General AI with domain-specific models for water segmentation, and uses the static observer flooding index (SOFI) as the proxy for water level variations. The framework uses the Segment Anything Model (SAM), a generic computer vision model by Meta AI, for segmenting images into discrete while semantically unknown objects. A ResUnet model pre-trained on a non-local dataset simultaneously identifies pixels with the highest probability of being water, which are then overlaid onto the segmented images to specify the water object. The framework was applied to image sequences acquired from river cameras stationed at four locations in Tewkesbury, UK, for water segmentation and water level trend monitoring. The SOFI time series were calculated based on the segmented masks and underwent data quality control using an unsupervised clustering method. The obtained SOFI signal showed an average correlation of 0.83 with real water level fluctuations, significantly surpassing the single ResUnet model's correlation of 0.54. The data provided by the framework was qualified for hydrologic model calibration referring to both error magnitudes and distribution patterns. Our study has thus moved toward an ease-of-use implementation of river cameras for transferable water level trend monitoring.

1. Introduction

Hydrologic models commonly used in water resource management need at least some data to be adapted or calibrated for the target catchment (Silberstein, 2006). Observations (i.e., data) are, thus, required before these models can be used to predict floods and droughts, or the impacts of climate change (Wang et al., 2023). Among observations of different hydrological variables, observing water level is relatively more feasible, given its greater resilience to the complex instantaneous water flow movement. Moreover, the effectiveness of water level trends in calibrating hydrological models, even in the absence of specific scalar values for water levels, has already been validated in previous studies (Etter et al., 2020; Seibert & Vis, 2016).

However, for many rivers, water level observations are unattainable due to the high costs associated with the installation and long-term maintenance of gauging stations (Fekete et al., 2012; Ruhi et al., 2018). Other observational approaches, such as remote sensing, can supplement the data from hydrometric networks and help to understand hydrological processes better and improve water resource management (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 or characterize its trend (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 at ungauged locations 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 extraction of water level variation information in space and time.

However, the challenges associated with obtaining high-resolution terrain data and detailed camera parameters make it difficult to directly deduce water level values from river camera images. An alternative approach to imaging-based water level monitoring involves segmenting water body within each image and calculating each image's static

observer flooding index (SOFI), which indicates the proportion of water pixels among the total pixels (Moy De Vitry et al., 2019). The temporal changes in SOFI can approximately represent the water level trend, therefore, the core of the imaging-based monitoring within this pathway lies in accurate water segmentation. While some studies attempted to identify water pixels in images by frame-crossing pixel analyses, they often suffer from reliance on staff gauges or local environments (Bentir et al., 2018; W.-C. Liu et al., 2023). In comparison, deep learning offers a high level of automation and strong scalability, thus is increasingly applied (Eltner et al., 2018). For instance, Akiyama et al (2020) adopted a classic deep learning model, SegNet, for water segmentation and achieved favorable outcomes on a medium-sized river in Germany. Lopez-Fuentes et al (2017) used three deep learning model structures, including Fully Convolutional Networks, Fully Convolutional DenseNets, and Conditional Adversarial Networks, for water segmentation, and systematically compared their performances. Moreover, Erfani et al (2022) further designed a novel deep learning model using distinct paths to handle aquatic and non-aquatic regions for improved accuracy in water segmentation.

Deep learning models are significantly influenced by the amount of training data. However, 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). Data augmentation involving rotation, cropping, and brightness adjustments can expand the dataset and enhance the model's adaptability to local variations in lighting and camera movement (Wagner et al., 2023). Nonetheless, achieving transferability in entirely new scenes remains challenging. 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). However, during the fine-tuning process, deep learning models often experience catastrophic forgetting (Kirkpatrick et al., 2017), wherein excessive focus on 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). Furthermore, manual data annotation for new scenes will consume extra manpower and time. A transferable deep learning-based 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., training data) required for downstream tasks. However, General AI cannot be independently used for handling specific tasks due to its generality. In this study, we employ a combination of General AI and domain-specific models for water segmentation on images captured by river cameras, aiming to balance the strengths and weaknesses of each model. 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. Its performance was compared with a single domain-specific deep learning model. Furthermore, an unsupervised clustering method, Density-Based Spatial Clustering of Applications with Noise (DBSCAN, Ester et al., 1996) was applied for data quality control on the extracted water mask sequences for each spot. 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 and data quality control, 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 superiority of the methods, the value of the extracted water level trend and the implications for future studies. Finally, the conclusions are given in Section 5.

2. Methods and materials

2.1 Water segmentation model structure

This study integrates General AI with domain-specific models to formulate a transferable deep learning framework for water segmentation in river images. As shown in Figure 1, the framework employs SAM, an innovative image segmentation model developed by Meta AI, as the foundational computer vision model. SAM automates the segmentation of original river images into multiple discrete objects, even though their

semantics are unknown ("*Everything Mask*"). Simultaneously, a conventional deep learning semantic segmentation model (ResUnet, Ronneberger et al., 2015), pre-trained on a limited river image dataset that does not include images taken at the river awaiting observation, is employed as the domain-specific model to identify the pixel with the highest water probability in the image. The coordinates of this pixel are then superimposed onto the "*Everything Mask*" segmented by SAM, facilitating the identification of the water object and thereby completing the water segmentation process. The specific introductions to model principles and structures for both SAM and ResUnet model within the framework are provided in the following section.
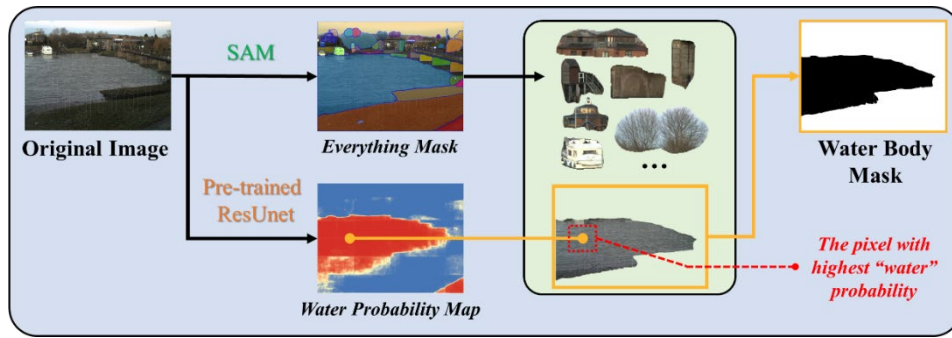

Figure 1. Diagram of the water segmentation framework. The framework is developed by coupling an SAM and a ResUnet model. The SAM and ResUnet outputs an everything mask and a water probability map, respectively. Consequently, the pixel with the highest water probability is overlain with the everything mask to specify the water body and generate the corresponding water body mask.

2.1.1 Brief introduction to SAM model

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. Following the above training process, SAM can serve as a foundational model in this study for segmenting images into a series of non-semantic masks, obviating the need for local fine-tuning processes.

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. Moreover, even if an object is separated into two or more sub-objects of the same semantics by other objects, *everything* mode SAM can output these sub-objects as a whole. For the latter, users need to manually provide additional hints to SAM, including boxes, points, and texts. These hints serve to guide SAM in the segmentation process for the expected object

In this study, we opt to leverage *everything* mode to fully exploit the segmentation capability of SAM itself. In the *everything* mode, SAM is used as a domain-agnostic General AI. This implies that, although the image is accurately segmented, the semantic understanding of each individual object remains unknown. Therefore, SAM needs to be used in combination with other domain-specific models for obtaining water-related hints.
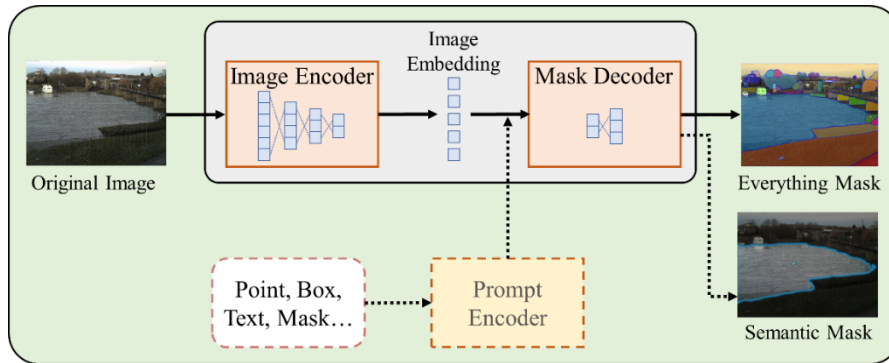

Figure 2. Diagram of the structure of the SAM model. Though the SAM supports both the automatic everything and manual prompt modes, this study only adopted the former mode to fully exploit the segmentation capability of the SAM. The prompt encoder was, thus, not used.

2.1.2 Brief introduction to ResUnet model

In this study, the task of providing water-related hints to filter objects segmented by

SAM is conducted by the ResUnet model. ResUnet is a conventional end-to-end convolutional neural network, employing the Unet structure as its backbone (Figure 3a). 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 3 (b) and (c)).

The ResUnet will struggle with qualified precision given limited training data or when facing new scenes that are significantly divergent from its training dataset's contexts. Therefore, this study avoids using the ResUnet model directly for mask generation. Instead, we rely on the water probability distribution output by it to identify the most water-like pixel in each image. By aligning the pixel with SAM's segmentations, the water body object is pinpointed. In essence, ResUnet50 can be viewed as an automated prompter, handling the object filter task that typically necessitates manual visual recognition in the post-processing of SAM's outputs.
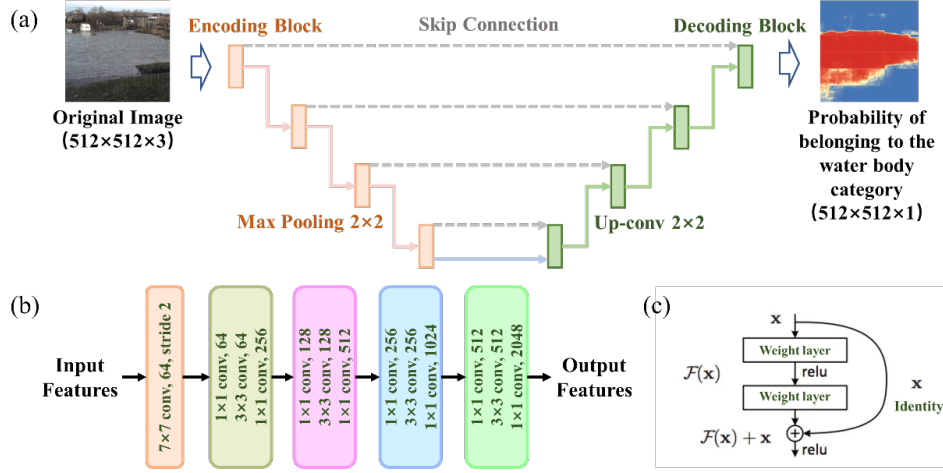
Figure 3. Diagram of (a) the structure of ResUnet model; (b) the structure of the ResNet-50; and (c) the residual operations in the ResNet-50. While the ResUnet model can generate probabilities for all the pixels in images, only the pixel with the highest probability will be used in subsequent applications.

2.2 Data quality control method

The above procedure aided by the ResUnet+SAM framework can produce water masks for each image along with its corresponding SOFI value, capturing temporal fluctuations in water levels. However, in instances where there is a considerable discrepancy in identifying the most water-like pixel by the domain-specific model (ResUnet), the framework may yield water segmentation results with significant errors. For these exceptional images, the calculated SOFI values are erratic and lack informativeness for model calibration, thus should be excluded.

This study uses an unsupervised clustering algorithm as the data quality control method for the removal of anomalous water masks. Images captured by river cameras are typically taken at small time intervals, resulting in a temporally evolving water mask sequence derived from the image sequence. Specifically, a water mask at a given time is expected to exhibit morphological similarity to water masks captured at adjacent moments. In datasets covering a lengthy time span with a substantial accumulation of images, a given water mask should have similar counterparts not only in adjacent time frames but also in moments with similar hydrological conditions. Building upon this assumption, the absence of morphologically similar water masks for a specific mask within the temporal sequence indicates an outlier, leading to its removal from the dataset.

As depicted in Figure 4, each mask generated by the framework undergoes uniform

division into multiple sub-regions of identical size, with subsequent calculation of SOFI values within each sub-region. These SOFI values are then assembled into a one-dimensional vector, serving to approximately characterize the morphological pattern of the mask. Subsequently, the DBSCAN method is applied to cluster these vectors.

DBSCAN is a density-based clustering algorithm designed to cluster data with arbitrary shapes in the presence of noise within high-dimensional data (Khan et al., 2014). The fundamental concept behind DBSCAN is that each data point within a cluster must have a neighborhood of a defined radius (*Eps*) containing at least a specified minimum number of data points (*MinPts*). In other words, the number of data points within this neighborhood must surpass a certain threshold. The *Eps*-neighborhood of a given data point '*p*', N_{Eps}, is defined as follows:

$$N_{Eps} = \{q \in D/dist(p,q) < Eps\} \tag{1}$$

Here, *D* represents the database of data points (SOFI vectors). The function $dist(\cdot)$ is used to calculate the Euclidean distance between two data points. If the *Eps*-neighborhoods of a data point '*p*' contain at least the required minimum number of points, the point is termed a core point. The core point is defined by the condition:

$$N_{Eps}(P) > MinPts \tag{2}$$

DBSCAN identifies clusters by examining the *Eps*-neighborhood of each data point in the dataset. If the *Eps*-neighborhood of a point '*p*' contains more than *MinPts*, a new cluster is formed with '*p*' as the core point. The algorithm then iteratively collects points density-reachable from these core points, potentially leading to the creation of a new density-reachable cluster. In our study, *Eps* and *MinPts* were configured as 0.5 and 5, respectively.

The aforementioned process continues until no further points can be added to any cluster, signifying the algorithm's termination. Ultimately, mask images corresponding to outlier data points, which do not belong to any cluster, are identified as anomalies and subsequently removed. The remaining data is retained for characterizing the trend of water levels over time.
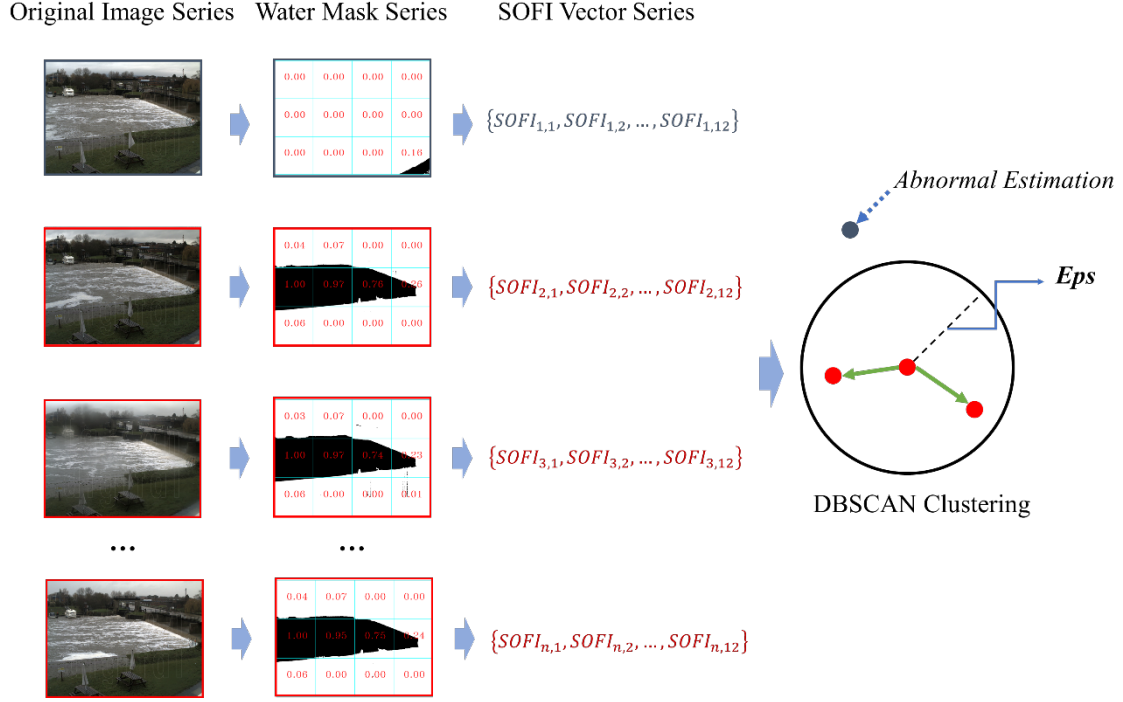
Figure 4. Diagram of data quality control on water masks. In this study, each water mask was divided into 3×4 patches. SOFI was calculated separately in each patch, and then synthesized into a twelve-dimension SOFI vector. The SOFI vector series were used for subsequent abnormal water segmentation identification based on DBSCAN.

2.3 Data

2.3.1 Dataset for pretraining ResUnet

The ResUnet model is pre-trained on the RIWA dataset (River Water Segmentation Dataset; Wagner et al., 2023). The dataset represents the first version of pixel-wise binary river water segmentation that offers 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 5, the images in this dataset encompass various lighting conditions, weather scenarios, and perspectives.

Figure 5. 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.

2.3.1 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 6). 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.
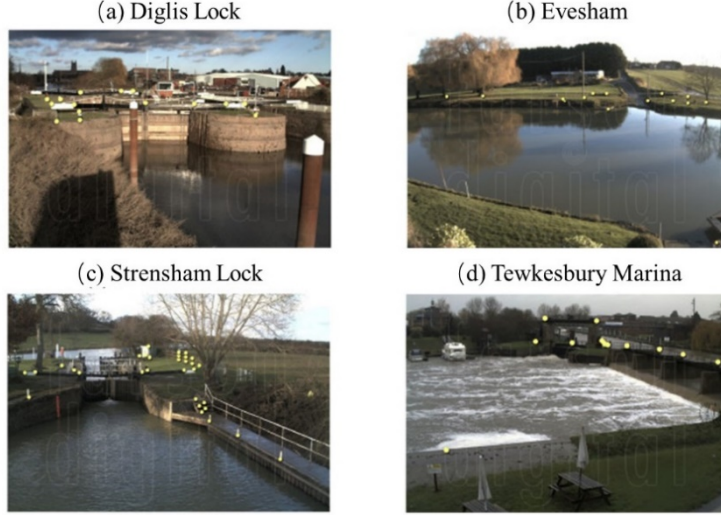
12

Figure 6. 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).

2.4 Experimental setup

2.4.1 Model setup

The ResUnet model was pretrained 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 the everything modes of SAM.

2.4.2 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 6). 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 the everything mode segmentation, 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.

2.4.3 Evaluation metrics

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:

$$IoU = \frac{1}{n}\left|\frac{S_i \cap \hat{S}_i}{S_i \cup \hat{S}_i}\right| \tag{3}$$

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 water-related hints for SAM.

Figure 7 provides a visualization of the spatial distribution of pixels most resembling water, as identified by the ResUnet model in images captured at various 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.
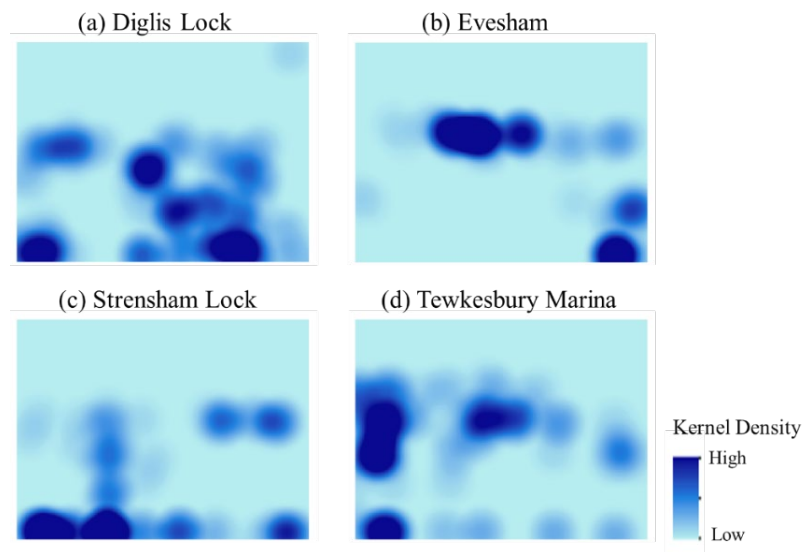

Figure 7. The kernel density of pixels with the highest probability of belonging to water identified by the ResUnet model in river camera images taken at different time points at the four locations.

Figure 8 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.
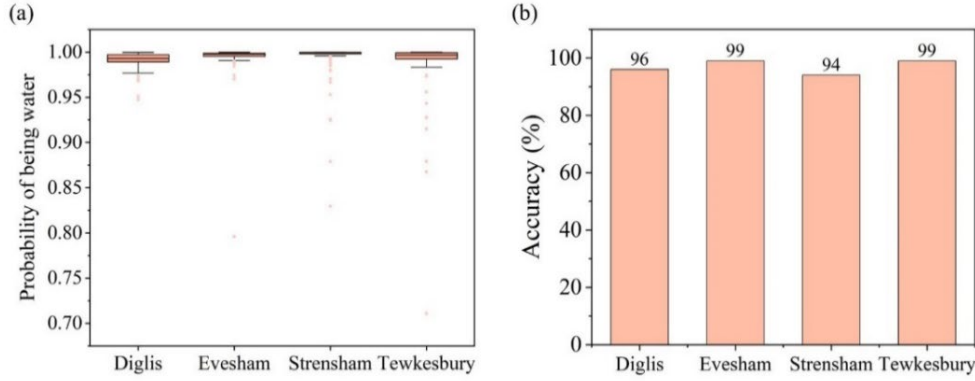

Figure 8. (a) The box plot depicting the probability values corresponding to the pixel with the highest probability of belonging to water among different images at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina. The upper and lower boundary of the box represent the upper (0.75) and lower 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

Aided by the robust most-water-like pixel identification by ResUnet and the class-agnostic segmentation by SAM, the framework has produced the corresponding water masks. Compared with the single pre-trained model, the framework's superior performance on water segmentation is demonstrated. As depicted in Figure 9, at Diglis Lock, Evesham, and Strensham Lock, the ResUnet+SAM framework 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.

16

At Tewkesbury Marina, the advantage of the ResUnet+SAM framework 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) by the ResUnet+SAM framework were also fewer than those produced by the ResUnet model. 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.
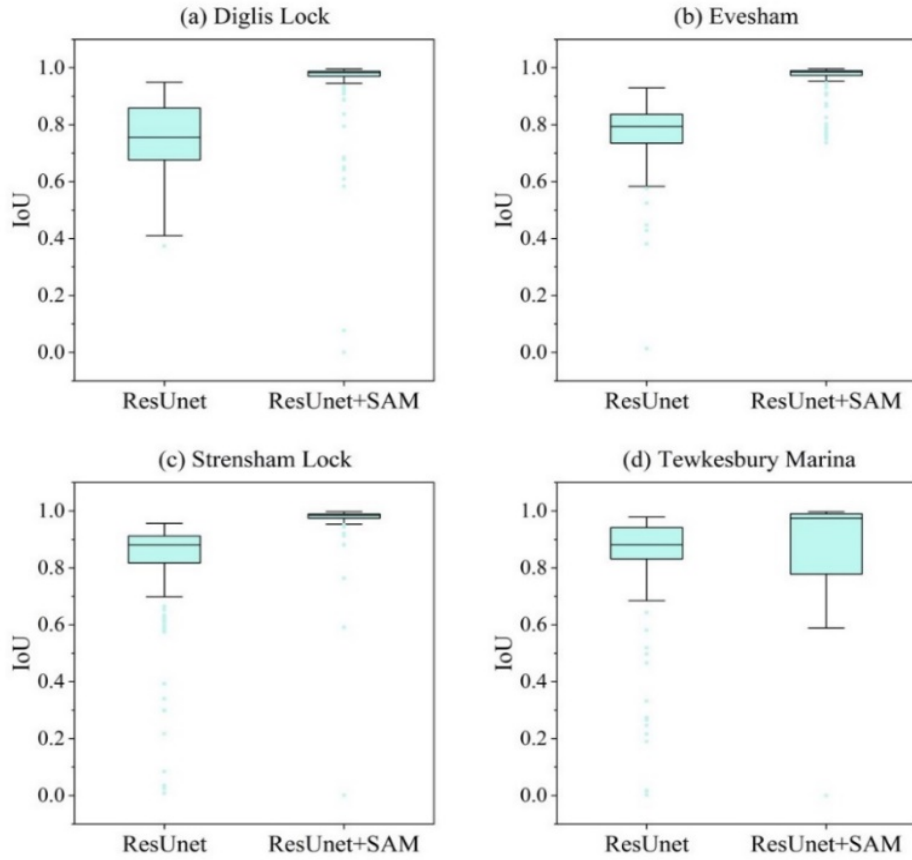

Figure 9. The IoU values for water body segmentation by the ResUnet+SAM framework and the ResUnet model at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina.

Figure 10 further compares the accuracy achieved by the ResUnet+SAM framework and the ResUnet model across various pixel coordinates within the images. the single ResUnet model 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,

the single ResUnet falls short in delineating the contours of the water bodies with the same level of precision as the coupling framework. In contrast, 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.
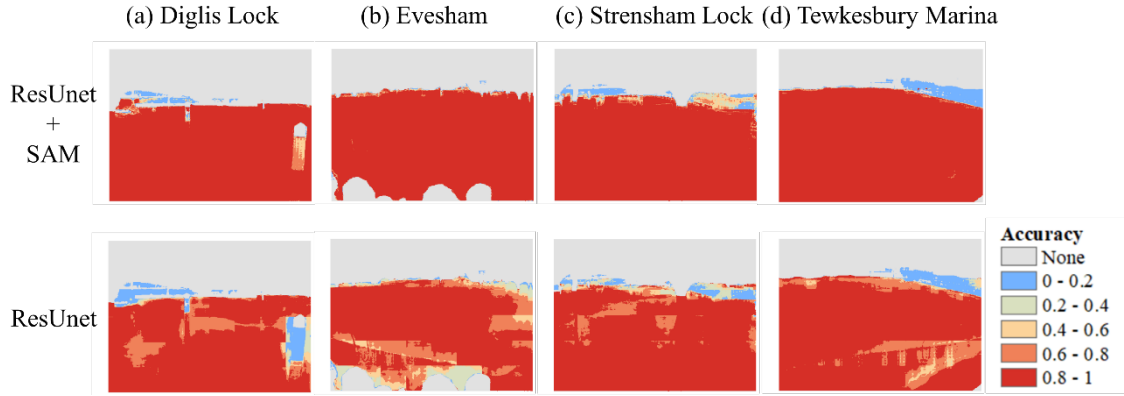

Figure 10. The accuracy achieved by the ResUnet+SAM framework and the ResUnet across various pixel coordinates within the river camera images for the four locations.

In Figure 11, 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. It can be observed that the single ResUnet model struggled to accurately extract water pixels. 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.
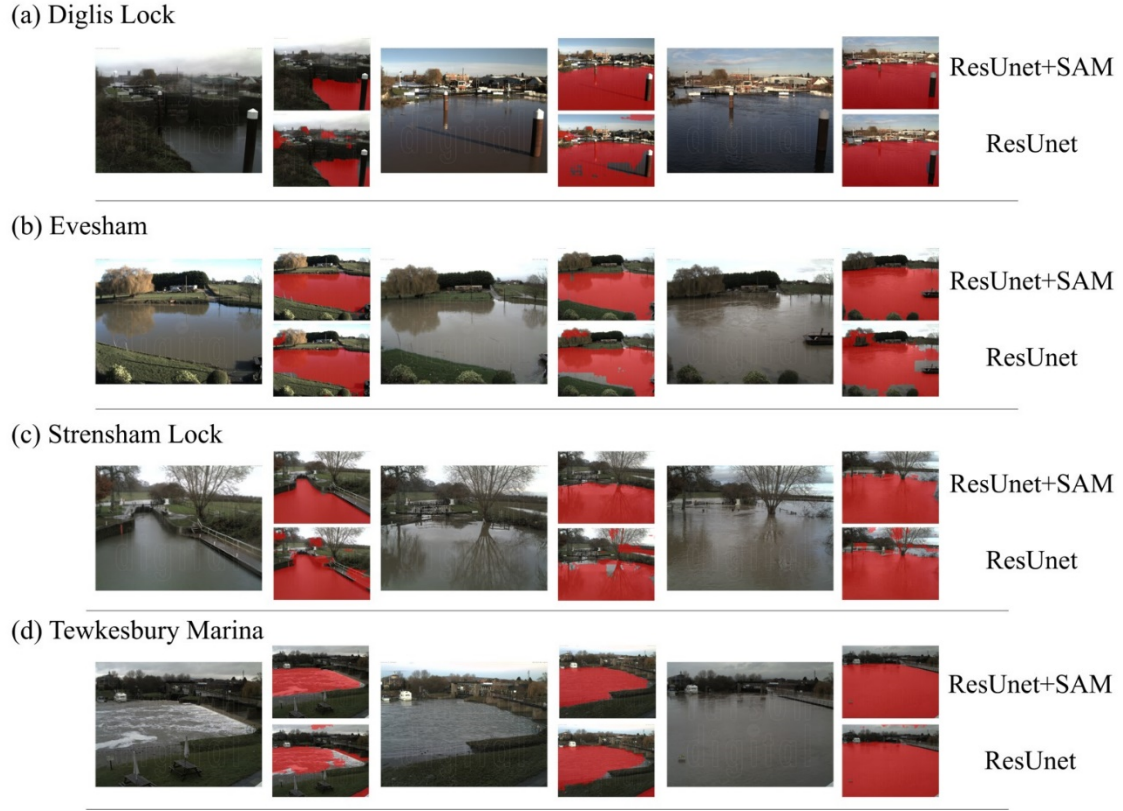
Figure 11. Illustrative examples of water segmentation results of ResUnet+SAM and ResUnet at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina.

3.3 The effectiveness of data quality control

Based on the water segmentation results, SOFI values representing the proportion of water pixels for each image can be computed. Preceding the comparison between the SOFI sequences and actual variations in water levels, the DBSCAN unsupervised clustering algorithm was employed to identify anomalous prediction values for data quality control.

As shown in Figure 12, at Diglis Lock, Strensham Lock, and Tewkesbury Marina, water mask sequences underwent the anomaly detection process, resulting in the removal of 3, 3, and 4 images with significant prediction deviations (IoU value<0.5), respectively. Consequently, the minimum IoU values for individual images in these sequences were elevated to above 0.6, notably reaching 0.88 for Tewkesbury Marina after data quality control. For the Evesham, ResUnet+SAM exhibited reasonable predictions across all images, with a minimum IoU value of 0.74, thus leading to no exclusion of images at this

site though processed by data quality control method. The substantial improvement in minimum IoU values for most locations signifies the efficacy of the DBSCAN method.

Regarding the single ResUnet model, the errors within its generated water masks were categorized as systematic errors (overall overestimation or underestimation), and its results with significant errors could not be eliminated through the outlier detection methods.
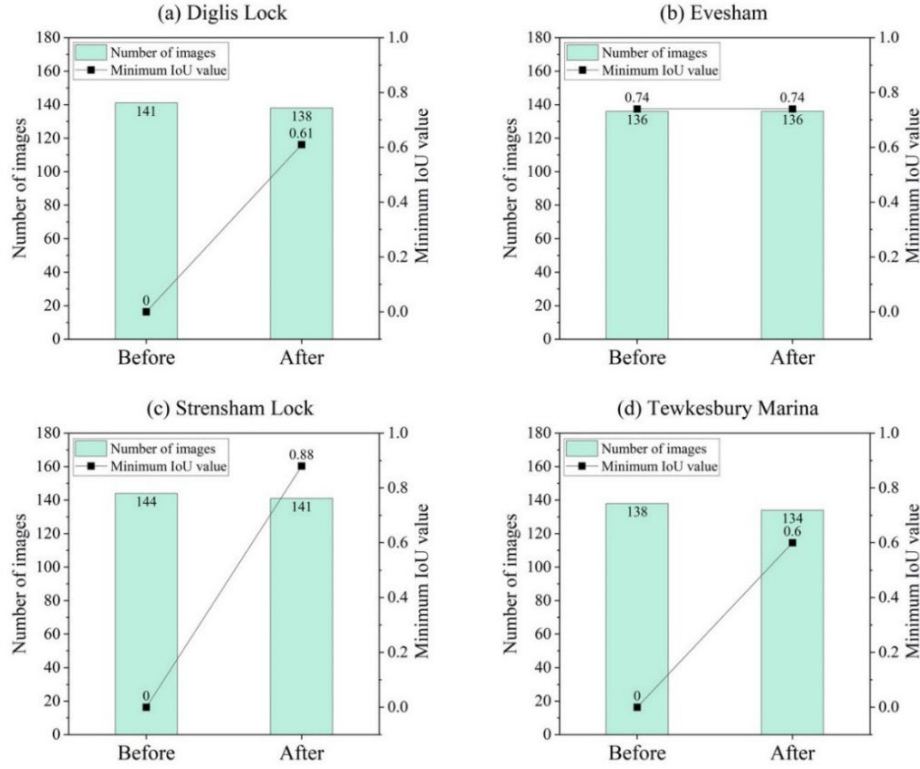

Figure 12. Number of images and corresponding minimum IoU values before and after data quality control at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina.

3.4 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 13 presents a comparative analysis of the trend in SOFI time series contrasted against the actual water level time series at four locations. It should be noted that the images removed through data quality control were not considered for both the single ResUnet and the ResUnet+SAM framework to focus on the inherent performance differences between them.

The figure illustrates that, across Diglis Lock, Evesham, and Strensham Lock, the SOFI variations obtained using the ResUnet+SAM framework 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.

At Tewkesbury Marina, the framework tended to overestimate water levels at specific intervals. Nevertheless, it still provided predictions that were relatively accurate in magnitude for extremely high and low water levels. In contrast, the SOFI values calculated based on the water segmentation results of ResUnet presented larger errors, especially at Diglis Lock, Evesham, and Strensham Lock, with a greater number of anomalies, thereby compromising the accurate depiction of actual water level trends. Moreover, most of these errors were systematic errors that are difficult to be eliminated through data quality control.
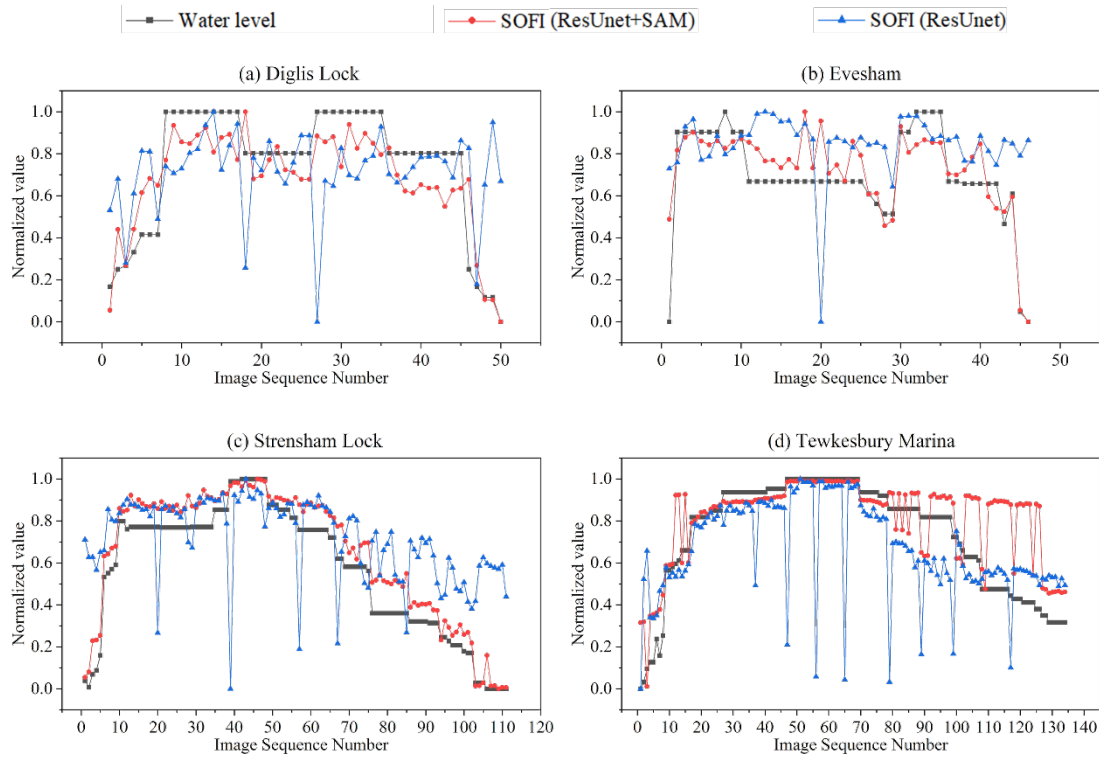

Figure 13. The trends of the actual water levels and the SOFI time series derived from both the ResUnet+SAM framework and the ResUnet model at Diglis Lock, Evesham, Strensham Lock, and Tewkesbury Marina. It should be noted that the images from different locations are sorted by time, but the time intervals 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 were also superior to those of the ResUnet model across all locations (Table 1). The Pearson correlation coefficient exceeded 0.8 in most locations, and the Spearman correlation coefficient even reached 0.92 at Strensham Lock. 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 13. Hence, the framework practically facilitates the monitoring of the temporal water level trend.

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

	Spearman correlation coefficient		Pearson correlation coefficient	
	ResUnet	ResUnet+SAM	ResUnet	ResUnet+SAM
Diglis Lock	0.28	0.84	0.26	0.88
Evesham	0.39	0.78	0.18	0.83
Strensham Lock	0.74	0.92	0.58	0.83
Tewkesbury Marina	0.76	0.78	0.59	0.78

Table 2 further explores the error distribution patterns of the normalized values of the SOFI time series obtained from the ResUnet+SAM framework and the ResUnet model. The correlation between the errors of the ResUnet+SAM framework and the normalized actual water levels was lower compared to that of the ResUnet model. The error distribution of the ResUnet+SAM framework appeared to be more random, while the errors of the ResUnet model exhibited a higher correlation with the water level values. Particularly, errors generated under low-flow conditions tended to be more pronounced.

Table 2. The Pearson correlation coefficient between the error of SOFI time series obtained using the ResUnet+SAM method, the ResUnet model, and the actual water level time series at various locations.

	Diglis Lock	Evesham	Strensham Lock	Tewkesbury Marina
ResUnet	-0.81	-0.81	-0.78	-0.60
ResUnet+SAM	-0.64	-0.53	-0.08	-0.64

4. Discussion

4.1 Why monitor water level trends rather than scalar values?

This study monitors the water level trend based on the SOFI sequences derived from water segmentation results, as opposed to fetching the scalar river water level value. Using deep learning models, directly predicting river water levels in the continuous domain from images can also be realized (Vandaele et al., 2023). However, training a regression model that can establish a general mapping relationship between images and water level values poses a greater challenge compared to developing a transferable water segmentation model. The challenge is, for one thing, related to the construction complexity and inherent defect of the deep learning-based regression model. The imaging-based water level prediction demands more specific knowledge and manual parameter tuning for specific sites. Meanwhile, 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, another limitation of the deep learning-based regression model is its inability to extrapolate outputs, making it challenging to monitor extremely high or low water levels (Vanden Boomen et al., 2021).

Necessity and flexibility are two additional factors considered in not using regression models to directly deduce scalar water level from images. According to previous studies, a robust characterization of water level trends is already valuable enough for hydrologic model calibration, with the Spearman rank correlation coefficient serving as the optimization objective (Weeser et al., 2019). Therefore, the acquisition of scalar water level values becomes less necessary, as it will not yield significant information increment. Moreover, when there is a solid requirement arises for the acquisition of river stage data, this objective can be accomplished through the application of photogrammetric techniques to transform water segmentation results, then overlaying the transformed water masks onto the topography of channel geometry to derive scalar water level values (Sermet & Demir, 2023). The framework's ability to accurately characterize water body edges, as demonstrated in the results, can support this potential application. Therefore, compared to directly predicting scalar water level values from images, monitoring water level trends based on water segmentation can be a more flexible technical path.

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. 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 (V. 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. This approach represents a fusion of the capabilities from the two models, thus achieving a balance between the two AI paradigms.

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 own 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 Are the SOFI time series derived by ResUnet+SAM informative for hydrologic model calibration?

The higher the correlation between SOFI and actual water level values, the better the calibration efficacy for the hydrologic model. This is why there is a continuous pursuit of higher-performance water segmentation algorithms. 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, hydrologic models could be calibrated to achieve significantly higher predictive level than uncalibrated benchmark. The SOFI sequences generated by the ResUnet+SAM framework in this study can achieve correlation coefficients of over 0.8 for most regions, thus are informative for hydrologic model calibration.

Despite the relative high accuracy level of ResUnet+SAM framework's predictions, certain degrees of forecasting bias still exist (Section 3.2). To handle data points with large errors, this study employs the DBSCAN unsupervised clustering algorithm for removal. Within the remaining data, the uncertainty of the framework has been stabilized and its error distribution pattern exhibits more randomness compared to the single ResUnet (Section 3.4). 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. Accordingly, the existing errors in predictions by the coupling framework will not impact their effectiveness for model calibration. Therefore, the river camera-based water level trend observations are robust enough to practically support the hydrologic modeling.

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 using the Static Observer Flooding Index (SOFI) as the proxy. The framework was implemented in four different riverside locations in Tewkesbury, UK, supported by an unsupervised

clustering method, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), for data quality control, and compared with the single ResUnet model.

Our results indicated the transferability of the ResUnet+SAM framework for water segmentation, with an average improvement of 13% in Intersection over Union (IoU) 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. Meanwhile, the average increase of 0.7 in minimum IoU across image sequences after the data quality control substantiated the effectiveness of the DBSCAN method. For all four locations, the spearman correlations between the SOFI and the actual water level exceeded 0.75, and the errors were randomly distributed. Regarding both the error magnitude and the error distribution pattern, the SOFI sequences obtained by the framework were informative for hydrological model calibration.

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 different 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 hydrologic models to evaluate its enhancement of hydrological forecasting performance.

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

Bentir, S. A. P., Balan, A. K. D., Ballado, A. H., & Lazaro, J. B. (2018). Optimization of object — based image segmentation in classifying water region. *2018 3rd International Conference on Control and Robotics Engineering (ICCRE)*, 249–253. https://doi.org/10.1109/ICCRE.2018.8376474

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

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–231.

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

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

Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S. (2014). DBSCAN: Past, present and future. *The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014)*, 232–238. https://doi.org/10.1109/ICADIWT.2014.6814687

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

Liu, W.-C., Chung, C.-K., & Huang, W.-C. (2023). Image-based recognition and processing system for monitoring water levels in an irrigation and drainage channel. *Paddy and Water Environment*. https://doi.org/10.1007/s10333-023-00935-9

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

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

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., & 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

Silberstein, R. P. (2006). Hydrological models are so good, do we still need data? *Environmental Modelling & Software*, *21*(9), 1340–1352. https://doi.org/https://doi.org/10.1016/j.envsoft.2005.04.019

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

Wang, Z., Seibert, J., van Meerveld, I., Lyu, H., & Zhang, C. (2023). Automatic water-level class estimation from repeated crowd-based photos of streams. *Hydrological Sciences Journal*, 1–15. https://doi.org/10.1080/02626667.2023.2240312

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

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