

Water Resources Research

RESEARCH ARTICLE

10.1029/2020WR027608

Key Points:

- Development of a camera gauge to measure the water stage from single images using convolutional neural network (CNN)
- Combining automatic/robust water area detection with high-resolution 3D data for stage retrieval in challenging environmental conditions
- CNN based water stage measurements reveal accuracy potential at cm-range and up to 93% agreement with traditional reference gauge

Supporting Information:

- Supporting Information S1

Correspondence to:

A. Eltner,
anette.eltner@tu-dresden.de

Citation:

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, e2020WR027608. <https://doi.org/10.1029/2020WR027608>

Received 7 APR 2020
Accepted 11 FEB 2021

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by-nc/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Using Deep Learning for Automatic Water Stage Measurements

Anette Eltner¹ , Patrik Olá Bressan^{2,3} , Thales Akiyama⁴ , Wesley Nunes Gonçalves^{3,4}, and José Marcato Junior⁴ 

¹Institute of Photogrammetry and Remote Sensing, Technische Universität Dresden, Dresden, Germany, ²Federal Institute of Mato Grosso do Sul, Jardim, Brazil, ³Faculty of Computer Science, Federal University of Mato Grosso do Sul, Campo Grande, Brazil, ⁴Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil

Abstract Image-based gauging stations can allow for significant densification of monitoring networks of river water stages. However, thus far, most camera gauges do not provide the robustness of accurate measurements due to the varying appearance of water in the stream throughout the year. We introduce an approach that allows for automatic and reliable water stage measurement combining deep learning and photogrammetric techniques. First, a convolutional neural network (CNN), a class of deep learning, is applied to the segmentation (i.e., pixel classification) of water in images. The CNNs SegNet and fully convolutional network (FCN) are associated with a transfer learning strategy to segment water on images acquired by a Raspberry Pi camera. Errors of water segmentation with the two CNNs are lower than 3%. Second, the image information is transformed into metric water stage values by intersecting the extracted water contour, generated using the segmentation results, with a 3D model reconstructed with structure-from-motion (SfM) photogrammetry. The highest correlations between a reference gauge and the image-based approaches reached 0.93, and average deviations were lower than 4 cm. Our approach allows for the densification of river monitoring networks based on camera gauges, providing accurate water stage measurements.

1. Introduction

Spatiotemporally dense hydrological observation networks are required to provide a suitable database for modeling and planning of water resources. One important hydrological observation is the water stage, which can be retrieved using several methods. Floating and pressure gauges are very common (Morgenschweis, 2010); however, they require the installation in the water and therefore entail the risk of losing them during severe flooding. Thus, remote observation techniques, such as those based on radar or ultrasonic devices (Hersch, 2008), might be preferred in such situations because they have the advantage that they can be installed remotely.

Another remote option is the application of image-based methods. In general, they can be low-cost in terms of the device compared to conventional water stage retrieval approaches if, for instance, basic camera circuit boards are used in combination with single board computers or microcontrollers. These setups also have the advantage that data processing can be performed on the spot, and only small data amounts have to be transmitted if needed, for example, via IoT infrastructures (Da Xu et al., 2014). Furthermore, camera gauges provide the benefit of potentially measuring flow velocities simultaneously by capturing short videos and tracking particles at the water surface (e.g., Eltner et al., 2020), which allows quantifying discharge with the same device eventually and thereby avoiding the installation of additional equipment. The application of camera gauges is suitable at rivers with periodic heavy debris loads or low angle river banks with large river width fluctuations leading to failure of other gauging techniques.

Observing the river reach with an optical device enables capturing qualitative and quantitative information beyond the water stage or flow velocities. For instance, ephemeral rivers and extended flooded areas become observable, and snow coverage and ice growth can be assessed. Furthermore, the vegetation evolution might be monitored, or the river cross-section changes determined to ensure a continuous and reliable water stage-discharge curve. Thus, camera observations can practically support hydrological monitoring.

Of course, quality constraints have to be considered when using camera-based approaches (e.g., Elias et al., 2020).

If the right setup is chosen, the images can be processed to extract the water stage information, ideally automatically. Several image-based approaches have already successfully been applied to camera gauges. However, most of them are designed for specific scenarios making it difficult to transfer them to other situations. The studies by Leduc et al. (2018) and Young et al. (2015) require vertical stage boards or rock sides for reliable measurements with errors of about 3 cm in the former study. Ran et al. (2016) apply image processing algorithms for automatic edge detection and image classification, and Stumpf et al. (2016) exploit the temporal texture to identify water regions. However, both studies are based on traditional image processing methods and do not perform an error assessment with independent reference measurements. Eltner et al. (2018) also consider the temporal texture of the changing water surface and estimate water stages with errors of 1.5 cm, but their approach solely focused on small regions of interest in the images to decrease the noise in their measurements. Pan et al. (2018) compared three different image-based algorithms and verified that the one based on convolutional neural network (CNN) revealed the best results with an average error of 9 mm when checked to reference water stage measurements. However, the authors also solely focused on a small region in the image, making it less suitable, for instance, for large flood scenarios. Although considering another environmental application, Kopp et al. (2019) illustrated the superiority of a CNN-based approach to another conventional image processing approach to measuring snow depth with errors below 5 cm and with high flexibility using a measuring rod.

All of these image-based approaches, except for the usage of CNN, inherit the challenge that they are sensitive to changing environmental conditions, for example, considering: calm and clear water during low flow versus turbulent and opaque water during flood events; or overexposed images with strong shadow and low contrast imagery during foggy conditions. Thus, a robust approach is needed to cope with different lighting conditions, such as strong shadows and changes in the surrounding, and that is transferable to different locations.

CNN is a class of deep learning, and in general, is applied to image analysis. CNNs were applied successfully in computer vision to identify objects with high reliability and robustness, and their popularity in the field of remote sensing is continuously growing due to the increased simplification of their application (Heipke & Rottensteiner, 2020). Previous works showed the potential of CNNs to classify water pixels in satellite imageries (Chen et al., 2018; Fang et al., 2019; Feng et al., 2018; Isikdogan et al., 2017; Jiang et al., 2018). However, in these studies, multispectral information was used, which significantly contributes to delineate water bodies because, in the infrared region, the water presents a strong absorption, contributing to its distinct differentiation regarding other targets. In this study, we adapt two state-of-the-art CNN structures to be applicable for robust water segmentation, that is, classification of each pixel, using RGB imagery. RGB sensors are cheaper when compared to multispectral sensors, making our work more reproducible.

The novel contribution is the proposal of an approach, which combines deep learning and photogrammetric methods for water stage estimation. CNNs are used to automatically and robustly segment water surfaces in images, which are used to generate waterlines via contour extraction of the segmented area. Afterward, photogrammetric approaches are applied to intersect the identified waterlines with the 3D terrain model to retrieve the water stage. Another novel contribution of our study is providing a labeled data set that allows for the assessment of ongoing approaches in this particular application of water stage measurement, which is so far not available. Thus, the data set is aimed to allow for the future development of a transferable CNN to be valid for water segmentation at different locations. The method proposed in this study is not to replace existing water stage measurement approaches but to complement them by densifying observations at lower costs.

2. Methods

In this study, two CNNs, fully convolutional network (FCN; Long et al., 2015) and SegNet (Badrinarayanan et al., 2017), are tested to evaluate their suitability to segment, and thus classify, water surfaces in imagery automatically. The boundary of the segmented water area (i.e., waterline) is intersected with a digital elevation model (DEM) of the observed river reach to acquire water stage measurements. The workflow

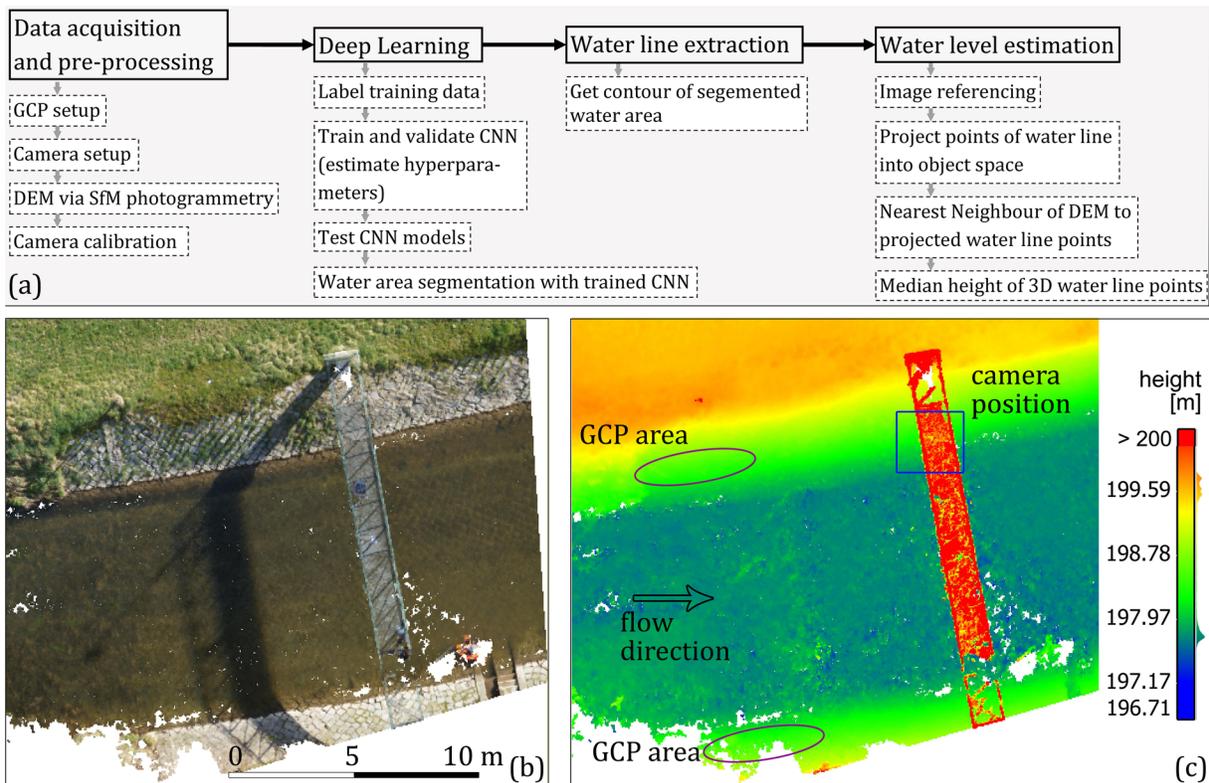


Figure 1. (a) Workflow to automatically extract the water stage from images. (b) Orthophoto and (c) DEM of the area of interest, including an indication of the camera position and ground control points (GCPs) locations.

for preparing the study area, camera setup, CNN approach, and water stage estimation is displayed in Figure 1a. A list of the terminology, originating from computer vision and photogrammetry, used in this manuscript to describe the methods is provided in supplement A to facilitate the understanding of the introduced approaches.

2.1. Study Area

The study area is located at the medium scale river Wesenitz, which flows through eastern Germany. The area of interest is situated at the standardized gauge station Elbersdorf to ensure reliable reference measurements to our image-based approach. The river width is about 10 m. Average water stages and discharge at the reference gauge are 45 cm and $2.14 \text{ m}^3\text{s}^{-1}$, respectively. At the gauge, water stages are measured automatically with a pressure gauge and averaged for 15 min.

Recent advancements in UAV (unmanned aerial vehicle) and 3D reconstruction from images enable easy, flexible, and affordable calculation of high-resolution topography data, which has led to a significant increase in their combined application in environmental sciences (e.g., Eltner et al., 2016). A DEM and orthophoto of the area of interest have been reconstructed (Figures 1b and 1c) using the computer vision techniques structure-from-motion (SfM; Ullman, 1979) and multiview stereo (MVS; Seitz et al., 2006), summarized here as SfM photogrammetry, from a field campaign in March 2017 (Eltner et al., 2018). Thereby, 20 overlapping images were captured with the UAV Asctec Falcon 8 equipped with a Sony Nex 5N (with a 6-mm fixed lens) at a flight altitude of about 25 m. Afterward, the DEM, comprised of a 3D point cloud, was generated with Agisoft PhotoScan software with a 3D error below 14 mm compared to independent terrestrial light detection and ranging (LiDAR) data. If no UAV is available, terrestrially captured images can also be used to calculate the DEM of the area of interest due to the platform independence of SfM photogrammetry. In total, 17 GCPs were surveyed with a total station and located around the area of interest with mm-accuracy to scale the image measurements. The resulting dense 3D point cloud was processed

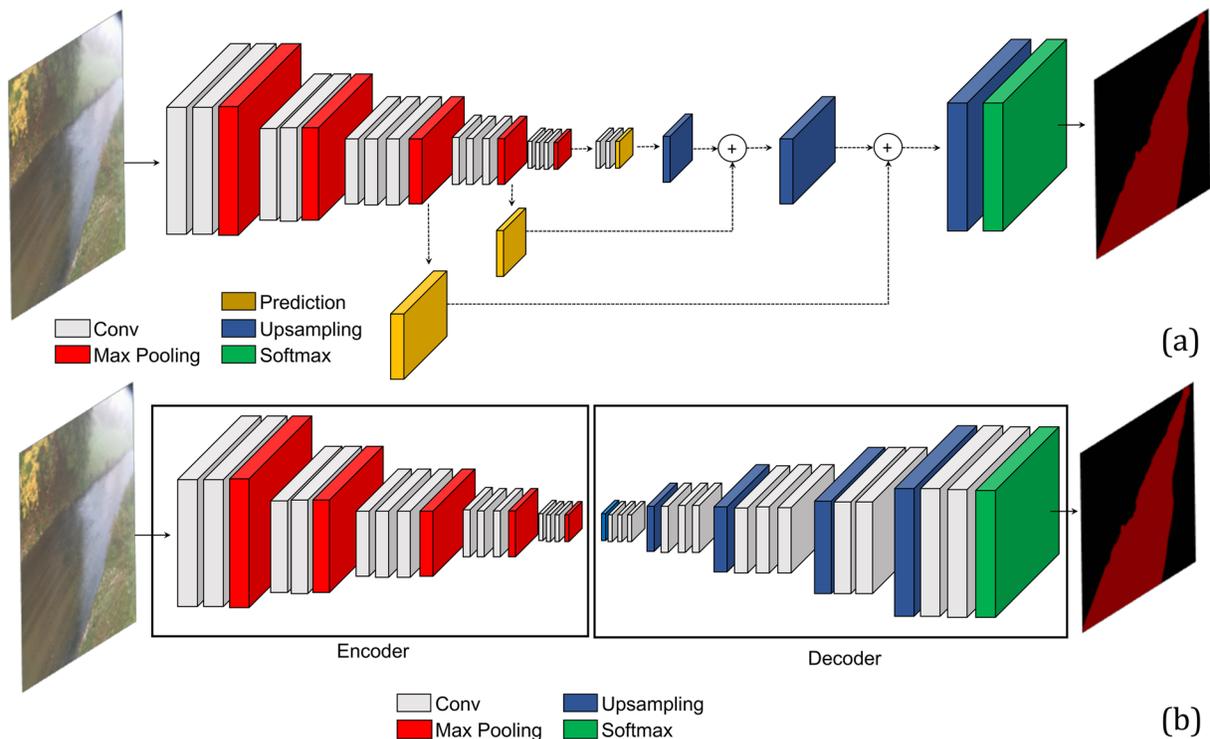


Figure 2. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation. (a) fully convolutional network (FCN) architecture; the figure is adapted from Long et al. (2015). (b) SegNet architecture is composed of two parts: encoder and decoder. The encoder extracts a low-resolution activation map while the decoder upsamples it to obtain a pixel-wise classification. The figure is adapted from Badrinarayanan et al. (2017).

to correct the underwater areas for the refraction effect, considering the multimedia photogrammetry tool developed by Dietrich (2017). However, if this approach is used to calculate the 3D model beneath the water surface, it has to be considered that the water needs to be calm and that the river bed has to be visible in the images. The resulting precision of the underwater area was 2.7 cm compared to total station point measurements (Eltner et al., 2018).

GCPs, represented by white circles on black background, were distributed at the left and right shore at the river region observed by the camera (Figure 2). They were measured with a total station with mm-accuracy in the same local coordinate system of the 3D model. Due to strong vegetation growth in the later observation period, additional natural GCPs were extracted at stone corners, and corresponding 3D coordinates were retrieved from the referenced UAV data (Eltner et al., 2018).

2.2. Image Acquisition

A Raspberry Pi RGB camera module was used to capture image sequences from March 30, 2017 to April 30, 2018. The corresponding single-board computer controls the camera. This low-cost setup is chosen to test the suitability of such tools as these cameras might be very suitable to densify hydrological observation networks due to their low resource consumption (Kröhnert & Eltner, 2018). The camera was mounted onto a lantern pole about 4 m above the ground. If daylight was sufficient, automatically measured with a light sensor controlled by the Raspberry Pi, 15 images were taken every 30 min. The camera module has a $2,592 \times 1,944$ pixel resolution and a fixed nominal focal length of 2.9 mm.

Due to various system failures at this observation spot, cameras had to be changed three times during the observation period. A new calibration of the interior geometry of the camera was necessary for each change to project the image measurements into object space correctly, allowing the metric water stage retrieval (Section 2.3.3). The camera calibration was performed with an in-house built calibration board. Images

were captured from different perspectives to estimate the interior orientation parameters (IOPs), minimizing correlations (Luhmann et al., 2014). In the end, the focal length, principal point, and radial distortion parameters were calculated within a bundle adjustment in the software Aicon 3D Studio. The radial distortion parameters are especially important for the potential erroneous projection of image measurements in the image corners.

2.3. Water Area Segmentation

The water area is segmented with two CNNs: FCN (Long et al., 2015) and SegNet (Badrinarayanan et al., 2017). In general, CNN is designed to learn the spatial composition of relevant characteristics from the raw input data. Learning is performed by multiple layers, the main one being the convolution layer that learns a set of filters automatically. In the first step, the acquired RGB data is labeled. Afterward, the network is trained with training and validation datasets. Finally, the trained CNN is applied to the test data to evaluate the performance of the estimated CNN, predicting the water area in so far unseen images.

2.3.1. Application of CNN to Water Segmentation

Deep learning allows automatic methods to learn patterns from raw data through multiple layers of processing (LeCun et al., 2015) in artificial neural networks (ANN). Each layer transforms the input representation into a higher-level representation. In this way, the deeper layers learn aspects of the raw data that are most important to the task while discarding irrelevant variations (higher-level representation).

Deep learning has offered great advances in many fields. CNN is a type of ANN, which is mainly used in image analysis. CNN is composed of convolution, pooling, and fully connected layers. A convolution layer receives an input volume (e.g., an image), which convolves with a set of learnable filters to produce an output volume, also called a feature map. Filters are trained to highlight relevant features on the activation map during learning. After the convolution layer, it is common to apply an activation function such as ReLU (Rectified linear unit) that utilizes a nonlinear function $f(x_0) = \max(x, 0)$. Pooling layers (e.g., max or average) are applied to reduce the computational cost by reducing the resolution while maintaining the relevant features. The max-pooling layer reduces the resolution maintaining only the maximum value in a given region (usually 2×2). Similarly, the average-pooling layer maintains only the average value for a region. Fully connected layers are applied after several convolution and pooling layers to classify the input volume in predefined categories. The last layer of CNN uses a softmax activation function to produce classification probabilities. The learning of layer filters is performed via stochastic gradient descent (SGD; LeCun et al., 2015). The main idea of SGD is to change the filters to minimize the loss function, which measures the discrepancy between the predicted category and the ground-truth (LeCun et al., 2015). The learning rate controls the magnitude of updating the filters.

Typical utilization of CNN is image segmentation, which refers to dividing pixels into regions with similar properties. In this work, the objective is to classify the image into binary classes: river and background. To automatically identify the river pixels in the images, we compared two state-of-the-art semantic segmentation methods: FCN and SegNet.

FCN converts CNNs used for classification tasks (composed of convolutional and fully connected layers) into fully convolutional networks that produce coarse activation maps. In this way, FCN can produce a class for each pixel instead of a class for the entire image. In the FCN method, the CNN structure of the model VGG16 (Simonyan & Zisserman, 2014) is remodeled by discarding the last layer (final classifier) and by converting all fully connected layers into convolutions. Thus, the original decision-making layer is replaced by (learnable) filters allowing for the input of different sized images. In FCN, convolutional layers with filters of size 1×1 are appended to the last layer to predict scores for all classes (yellow layers in Figure 2a), that is, in this study, water and nonwater, at each coarse output locations (scales). Finally, an upsampling layer is used to bilinearly upsample the coarse outputs. This layer increases the resolution by replicating the values of the neighbors. To refine the spatial precision, FCN fuses the prediction layer with shallower layers of the network by summing predictions and applying a softmax function at the end (as shown in Figure 2a) (Long et al., 2015; Torres et al., 2020).



Figure 3. Examples of original (first row) and labeled images (second row) using the annotation software LabelMe.

SegNet architecture consists of a sequence of layers (encoder) and a corresponding set of layers (decoder) followed by a pixel-wise classifier (Figure 2b). Given the input image, the encoder part provides a low-resolution activation map, which describes the most important features. Then, the decoder reconstructs the segmented image from the coarse activation map obtained from the encoder, as can be seen in Figure 2b. In SegNet, the encoder part is composed of the convolutional and max-pooling layers of the VGG16 (fully connected layers are not used). The decoder is composed of upsampling and convolutional layers that use the max-pooling indices from the encoder to upsample the low-resolution activation map (Badrinarayanan et al., 2017; Noh et al., 2015). Each decoder layer upsamples an input (doubles its resolution) by placing the input values in the locations indicated by the max-pooling indices and zero in the other positions. Since the upsampled maps are sparse (i.e., composed of a large number of zeros), convolution layers are applied to produce dense activation maps. After the convolution layer, the positions with zero will be filled with values learned by the filters. Using max-pooling indices provides important detail conservation and a significant reduction in the number of training parameters (Badrinarayanan et al., 2017). The detail preservation can be especially important to map the delineation between water and shore with good accuracy. Finally, the softmax activation function is applied to obtain a pixel-wise classification with probabilistic values.

FCN and SegNet are segmentation methods already used in several applications, e.g., for scene understanding (Badrinarayanan et al., 2017) and tree segmentation (Torres et al., 2020). In this work, both methods use the same initial set of convolutional layers to extract a lower activation map. This set of layers is known as the backbone. In FCN, the upsampling of the lower activation map to the original image size is performed during only three steps (8X, 16X, 32X). In contrast, SegNet upsamples using several blocks that use corresponding pooling indices (Figure 2b). The segmentation methods were coded using Keras-Tensorflow (Chollet, 2015) on the Ubuntu 18.04 operating system.

2.3.2. Image Data Set

We manually annotated 20,309 images from March 30, 2017 to April 30, 2018 using the software LabelMe (Wada, 2018; see examples in Figure 3). The processed images reflect different periods of the day in different seasons for about 1 year allowing the analysis of the river in different situations. The high number of images during different environmental scene representations is important to reduce the overfitting probability of

the model. This data set is referred to as a full data set (FD) in the experiments. To assess the best resolution for resizing the images, which is necessary when adopting CNN methods, we built a subset consisting of 3,407 images from March 30, 2017 to May 16, 2017 (first months).

However, not all images were used for later assessment of accurate water stage retrieval due to several camera failures; cameras were changed or repaired and afterward installed again, leading to changing interior and exterior camera geometries, respectively. Therefore, keeping all images within a single time series analysis complicates the performance assessment of water stage detection because of potential errors due to camera geometry and image segmentation intertwine. Furthermore, sometimes images were solely available for a few subsequent days, also complicating the time series analysis. Therefore, to assess how good water stage changes are captured, we focus on four intervals. The image sequences in these intervals were captured continuously for several weeks with the same setup to enable suitable statistical analysis avoiding the impact of changing camera geometry or single-day measurements with potential outliers. In the end, we evaluated the water stage estimation performance for the periods April 5–April 26, 2017 (spring), May 15–June 22, 2017 (early summer), June 23–July 7, 2017 (summer), and August 29–September 19, 2017 (autumn).

2.3.3. Experimental Setup to Train the CNNs

In this study, each image is downsampled to a fixed resolution due to memory consumption during the CNN training. We evaluated the best image resolution with a subset of the image data set composed of 3,407 images from March 30, 2017 to May 16, 2017. During the experiments, we evaluated resolutions of 256×256 and 512×512 pixels. The image data set was randomly divided into training (60%), validation (20%), and test sets (20%). The training set is used to train segmentation methods, while the validation set was used to tune hyperparameters (learning rate and the number of epochs). We refer to Goodfellow et al. (2016) for more information on CNN training. Finally, the test set is used to report the results of the proposed approach. After identifying the more suitable image resolution of 512×512 (Section 3.1), the FD was used for training following the same workflow as performed for the subset. For each day, the images were randomly divided into training (60%), validation (20%), and test sets (20%). As the test set has images of different dates and seasons, this set is suitable for evaluating the methods with respect to their accuracy and generability. Finally, the classified water pixels are converted into a single water line by extracting the boundary of the water segment based on a traditional image processing technique proposed by Suzuki and Abe (1985).

Before training, the weights of the encoder (i.e., VGG16 layers) of all methods were initialized with values pretrained at ImageNet, a procedure known as transfer learning. The stochastic gradient descent optimizer with a learning rate of 0.001, a momentum of 0.9, and a weight decay of 0.0005 were used for training all layers of both methods. Each method was trained through 30 epochs when the loss function stabilized in training and validation sets.

The performance of the methods in image space was measured by two metrics: pixel accuracy and intersection over union (IoU) (Long et al., 2015). Both compare manually measured ground truth (GT) data, which is the manually annotated (with LabelMe) water area, to the predicted data (Figure 4a). The pixel accuracy indicates the percentage of correctly classified pixels, contrasting true positives and true negatives (i.e., correctly segmented class belonging and nonclass belonging pixels) to all classified pixels (resulting in true positives and true negatives as well as false positives and false negatives). A value of 1 indicates that all pixels were classified correctly, and with increasing error, the value will decrease. The IoU metric calculates the ratio between the number of intersecting pixels of ground truth and predicted mask and the number of unified pixels of both masks. If both masks match exactly, the value will be 1, and if there are deviations, the value will decrease.

In addition to the annotated and automatic approach, single points were picked manually in the images at the right shore using the software ImageJ (Schneider et al., 2012). The image coordinates of the individual points indicate the border between water and shore. These single points are in contrast to the annotated water lines that stretch along the entire observed river reach. These observations were used to distinguish between the influence of the error of camera geometry estimation and errors of the automated measurements in the images on the accuracy of the water stage calculation. The well-controlled, single-point approach has

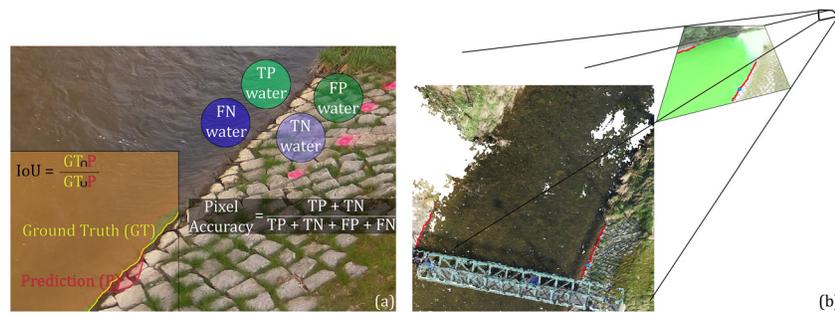


Figure 4. (a) Error metrics (IoU and pixel accuracy) and corresponding equations used to estimate the performance of image segmentation with CNNs. (b) Oriented camera in the same reference system (object space) as 3D point cloud displaying area of interest to transform image measurements (red water line) to water stage information. The water line, which can be considered as a collection of pixels, is projected through the camera projection center into object space and then intersected with the 3D model to eventually retrieve the water stage as median of all intersected points (i.e., pixels) of the water line. Blue dot indicates the manually measured single point that represents the border between water and shore.

the advantage that potential selection errors during the image labeling due to outliers along the shore-river border are minimized.

2.4. Image Measurements Referencing

To transform the image measurements in pixels into metric values of water height, the information about the interior and exterior camera geometry and the topography and bathymetry of the observed river reach is needed. The latter two are combined in a single dense point cloud with minimal point distances of 0.5 cm. The interior geometry was calibrated for each new camera setup with the in-house calibration field. The exterior geometry (camera pose) was estimated via spatial resection considering the GCP information and the calibrated IOPs. Thereby, the 2D image coordinates and the corresponding 3D coordinates of the GCPs are used in a Levenberg-Marquardt optimization, as more GCP information is given than needed to estimate the camera pose (e.g., Kraus, 2007; Luhmann et al., 2014). Camera pose was calculated for each captured image to account for camera movement due to wind or sun insolation.

After retrieving all the necessary parameters, the image points of the water line are projected into object space to intersect these points with the 3D point cloud describing the river reach (Figure 4b). In the 3D point cloud, the nearest neighbor point to the projected image water line point is chosen as a valid point. The Z-coordinate of these water line points corresponds to the water height. The water lines are solely intersected at the location of the paved river reach and thus mostly vegetation-free areas. Vegetated areas are not valid for the intersection approach because plant growth changes the 3D appearance of the river reach. However, for the measurements, a stable 3D object is assumed. If this is not the case, the 3D model of the area of interest would need to be updated for each image-based water stage retrieval. We intersect the image information, i.e. the water contour, with the left and the right river shore. The results are intersected water line points for the left and right shore, separately. The differentiation between both river sides is chosen to evaluate the potential influence of object to camera distance on the water stage calculation error. To estimate for each image a single water stage value from all intersected water line points, we calculate the median of all height values. Furthermore, a local robust weighted regression filter (Cleveland, 1979) is applied to each of the time series to smooth the temporal water stage change detection to mitigate the impact of strong outliers during the data comparison to the gauge measurement.

To assess the performance of water stage retrieval, we use the two error metrics accuracy, that is, average (mean) difference between the reference gauge and the image-based water stages, and precision, that is, the standard deviation of water stage differences between reference and camera gauge.

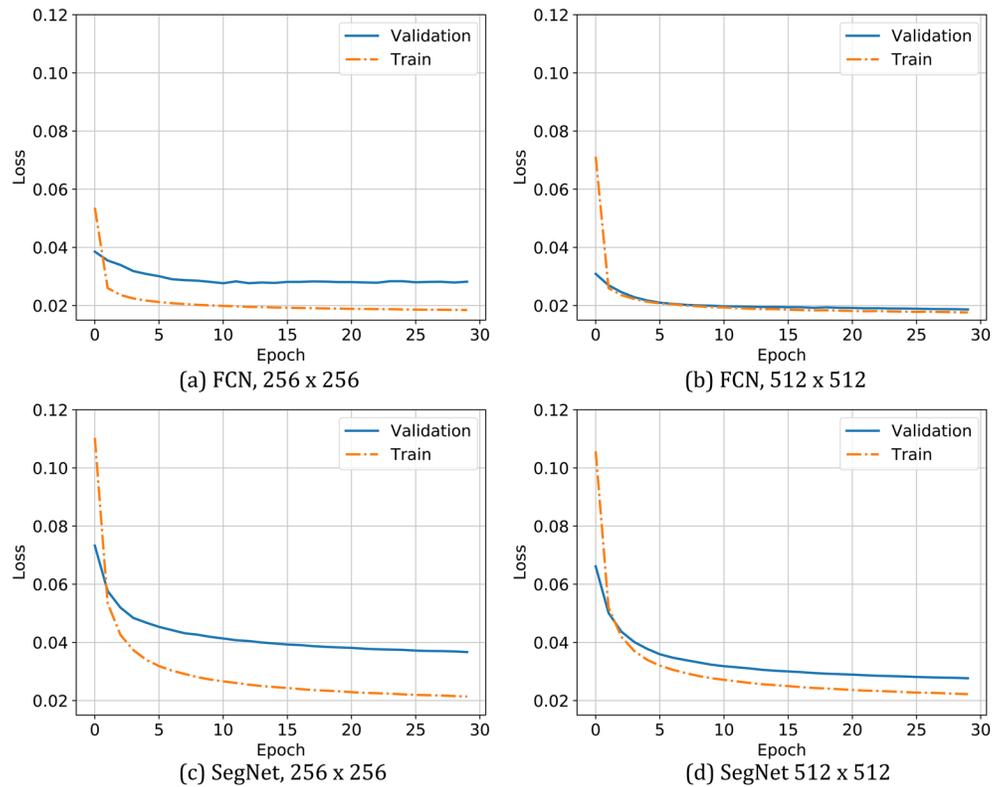


Figure 5. The loss function for two segmentation methods using resolutions of 256×256 and 512×512 pixels training with the subset image data set.

3. Results

First, we display the water segmentation results in the images considering both CNNs and different image resolutions. Afterward, the performance of water stage estimation was evaluated, comparing the image-based results to the reference gauge measurements.

3.1. Water Segmentation in the Images

To assess water segmentation, we varied the resolution of the input image from the subset of the image data set composed of 3,407 images. Figure 5 presents the loss function for both segmentation methods using resolutions of 256×256 and 512×512 pixels. As can be seen, the loss functions of SegNet and FCN using the resolution of 256×256 pixels showed indications of overfitting because the loss during validation with unseen images remains higher than during training. In contrast, the loss functions of both methods for resolution of 512×512 pixels indicate low overfitting (especially for FCN), as the loss values in training and validation were similar. In any case, the loss function for both methods stabilized with the chosen number of 30 epochs.

Table 1 presents the results using pixel accuracy and IoU for different resolutions. We observed that increasing the resolution from 256×256 to 512×512 improved the results of both semantic segmentation methods. The higher the resolution up to a certain limit (not investigated in the current work due to memory limitations), the more important details can be learned. Furthermore, a slightly higher performance of the SegNet approach can be observed compared to the accuracy achieved with FCN.

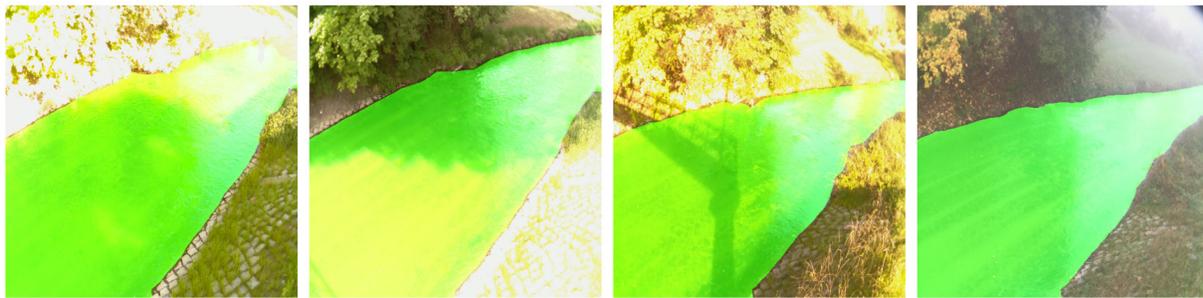
Given that 512×512 pixels achieved the best results for both methods, we trained them in the complete image data set (FD) with 20,309 images. This data set has several challenges, such as changing lighting and camera position. Although this data set has images with more variations, the results were similar to the subset (3,407 images), as shown in Table 1. Figure 6 shows the segmentation of test images for different

Table 1
Evaluation of the Image Resolution Using Pixel Accuracy and Intersection Over Union (IoU) in the Subset of Images and FD

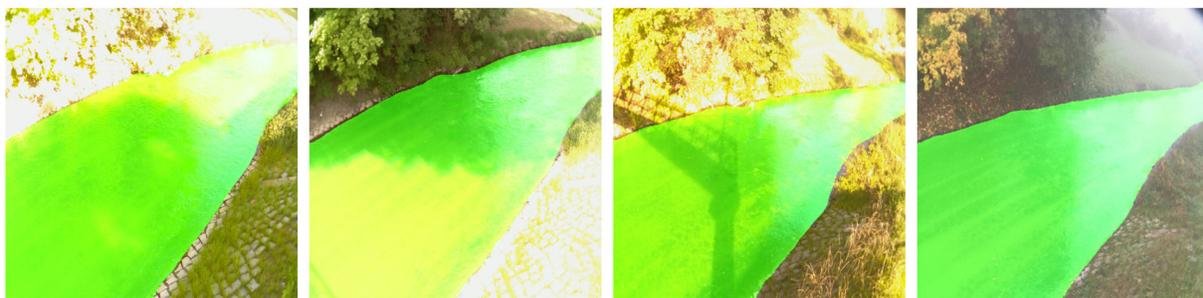
Method	Resolution	Pixel accuracy		IoU	
		Background	River	Background	River
SegNet	256 × 256	0.9880 (±0.006)	0.9890 (±0.004)	0.9750 (±0.005)	0.9795 (±0.005)
	512 × 512	0.9920 (±0.005)	0.9916 (±0.004)	0.9821 (±0.005)	0.9852 (±0.005)
	512 × 512 (FD)	0.9903 (±0.006)	0.9897 (±0.016)	0.9817 (±0.006)	0.9800 (±0.017)
FCN	256 × 256	0.9956 (±0.003)	0.9825 (±0.004)	0.9745 (±0.004)	0.9790 (±0.004)
	512 × 512	0.9952 (±0.004)	0.9900 (±0.003)	0.9830 (±0.004)	0.9861 (±0.003)
	512 × 512 (FD)	0.9820 (±0.006)	0.9804 (±0.018)	0.9819 (±0.006)	0.9802 (±0.018)



(a) RGB



(b) FCN



(c) SegNet

Figure 6. Examples of different illuminations and view in (a) test images segmented by (b) FCN and (c) SegNet.

lighting and camera position to illustrate the learning generalization. We verified that the images differ visually, although both segmentation methods separated the river accurately.

3.2. Water Stage Estimation

The application of two different CNNs resulted in comparable performances for water segmentation, and consequently for the water stage estimation, as can be verified in Figure 7. Both reveal capabilities to measure water stage robustly, indicated by small quartiles of differences between the reference gauge and the image-based water stage (Figure 7a). Considering all four intervals of measurement, the average deviation for FCN amount at the left and right shore were -1.1 ± 3.1 and -3.6 ± 2.0 cm, respectively. For SegNet, deviations were -3.1 ± 2.8 and -3.2 ± 2.3 cm. At the left shore, the accuracy was higher when using FCN. However, the precision was lower, indicated by a larger spread of the quartiles compared to the measurements at the right shore. Therefore, FCN revealed lower repeatability or robustness of the measurement at the left shore. The SegNet results revealed smaller quartile ranges of deviations and similar accuracies at both shore sides, which was in contrast to FCN. However, it can be noted that the differences in performance between SegNet and FCN are nevertheless small. The GT depicted a similar average performance as the predicted data at the left and right shores (-2.2 ± 2.8 and -3.7 ± 3.2 cm, respectively). The deviation of the manual point-based measurements at the left shore is also in the range of the automatic approaches (1.7 ± 2.3 cm; supplement B).

Besides the absolute comparison of water stages between the reference gauge and the image-based approaches (i.e., camera gauge), we also considered how well water stage changes were captured by the camera gauge (Figure 7b). Therefore, the spearman correlation coefficient was calculated, revealing a very good performance of FCN and SegNet regarding the measurement of water stage variations. The lowest correlation value was 0.87 (left shore, FCN), and the highest correlation coefficient amounts of 0.93 (right shore, FCN). Interestingly, the GT performs lower than both prediction methods. Assessing the absolute deviations and capturing of water stage change reveals that neither FCN nor SegNet can be considered as outperforming the other. The single point-based manual measurement of the water-shore-border depicts a correlation (0.89); in the range of CNN approaches. The similar performance of the point-based and CNN approaches indicates an overall strong impact of the accuracy of intersecting the image measurements with the scaled real-world point cloud at the error of water stage estimation regardless of the water delineation approach.

3.3. Seasonal Performance

In the next step, we took a closer look at different intervals (explained in Section 2.3.2) separately to check if different factors, such as lighting or vegetation growth, on the performance become obvious and to investigate if one CNN is preferable during different environmental conditions. During the first two seasons (spring and early summer), the water stage estimation showed, in most cases, lower accuracies if the right shore is used to intersect the image measurements (Table 2). In the third and fourth observation periods (summer and early autumn), measurements performed with the left shore reveal lower accuracies, probably due to strong vegetation growths between the cobblestone joints, which did not occur at the more strongly shaded right shore.

Considering the correlations between the CNN-based water stage estimation and the reference gauge for each period separately (Figure 7c) revealed that the best results are achieved at the left shore during early summer and at the right shore during autumn, which partly confirms the descriptive statistical findings (Table 2). The lowest correlation coefficients were calculated for the summer period at both shores. Also, the manual measurements using single points show the lowest accordance to the reference gauge in that period.

In general, considering the time series of each period, we observed an overestimation of the water stage at the right shore, especially during the summer, except for the last period in autumn, where the finding inverts (Figure 8). Nevertheless, the time series indicate that the temporal changes were sufficiently detected. Thus, the systematic offsets can be mitigated, considering a single independent reference measurement. Furthermore, short-term fluctuations were captured by the cameras, which were, however, averaged out by

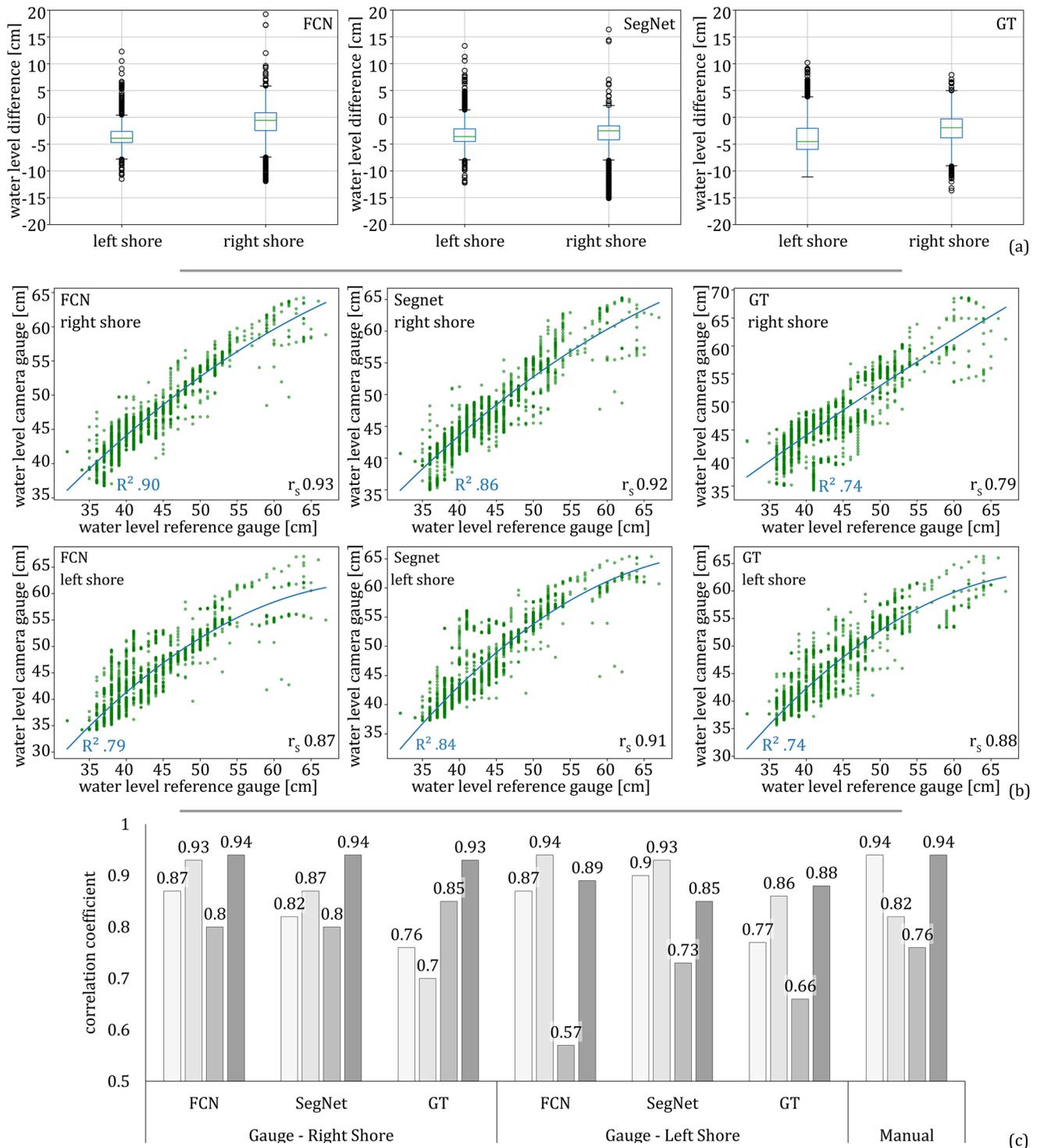


Figure 7. (a) The error of image-based water stage estimation using FCN, SegNet, and GT considering all four time series; the water stage differences between the reference gauge and the camera gauge are displayed. (b) Spearman correlation coefficient r_s and scatterplot for each image-based approach compared to gauge reference. Blue line corresponds to second degree polynomial fit to the scatter points. (c) Spearman correlation coefficients between water stage measured at the gauge and image-based approaches at both shore sides for each time series.

Table 2
Differences of Water Stage Estimation Comparing Image-Based Approaches to the Gauge Reference (Mean Difference, i.e., Accuracy, in [cm] and Standard Deviation (Std) of Difference, i.e., Precision, in [cm])

		FCN		SegNet		GT		Manual
		Right	Left	Right	Left	Right	Left	Left
Spring 2017	Mean	−3.01	−0.69	−3.74	−3.87	−3.73	−2.1	−1.88
	Std	1.8	2.56	2.71	1.78	3.94	3.02	1.58
Early summer 2017	Mean	−4.58	0.98	−3.88	−1.99	−5.35	−0.44	−0.30
	Std	0.87	0.85	1.12	0.88	1.72	1.27	1.21
Summer 2017	Mean	−4.56	−3.96	−3.88	−2.63	−3.27	−4.03	−4.30
	Std	1.92	5.10	1.64	2.97	2.02	2.55	1.61
Autumn 2017	Mean	−2.2	−4.4	−0.62	−6.84	−2.29	−4.37	−4.59
	Std	2.67	3.81	2.7	4.4	3.42	3.69	2.86

the reference gauge measurements. These fast but short term changes in the water stage were due to mill activities upstream, which regulate the flow frequently and strongly, leading to fast water stage changes within a few minutes that were not captured by the camera with 30 min resolution imagery. Therefore, correlations of 1 between the image-based and reference gauge can never be expected, even under perfect observation conditions.

Having a detailed look at the time series of all seasons (Figure 8) and the corresponding image classification results enables the identification of problematic cases (supplement C). During the spring period, a high number of outliers were given at the right shore considering water stages estimated with both CNN approaches. Two strong outliers occur in the period from June 24 to June 26 (summer) at the left shore for the FCN approach. In the middle of October, strong overestimations of water stages were observed at the left shore for both CNN approaches, also confirmed by the lowest accuracies during the last observation period (Table 2). An especially large overestimation of the water stage occurred on October 17 for the FCN approach. Higher errors were also observed during autumn for the manual measurement approach (Table 2, supplement D).

4. Discussion

The evaluation of FCN and SegNet revealed that, in general, both methods presented similar results in terms of the range of deviations and correlations to the reference measurements. However, FCN experienced stronger outliers (Figure 8). Smoother upsampling from SegNet could be responsible for smaller outliers, as shown for general images in Badrinarayanan et al. (2017). The water stage accuracies generated using the manually labeled water areas were worse than the predictions with the CNNs. A potential reason for this outcome is that, although the water area was not as perfectly manually annotated at the shore, many more water pixels were correctly assigned in the significantly larger water area, making the training solution with the CNNs robust.

The CNN-based image segmentation provided highly accurate results, minimizing the effect of the water line detection accuracy on the water stage estimation accuracy. Thus, deviations in water stages between reference and camera gauge were mainly due to errors of the camera pose and interior geometry estimation leading to noisy referencing of the image measurements. Errors of camera pose estimation were amongst others due to insufficient retrieval of the IOPs, which is indicated by changing pose accuracies with changing cameras (supplement E). But the standard deviation of the estimated camera pose can only be an indication of the accuracy due to correlations with the IOPs. Another potential error source impacting the pose estimation was the achievable accuracy of GCP measurements in the images, which varied during the seasons due to changing target appearances. However, quantifying the proportion of error of GCPs was not straightforward as it was not possible to quantify how accurately the GCPs were measured in the images

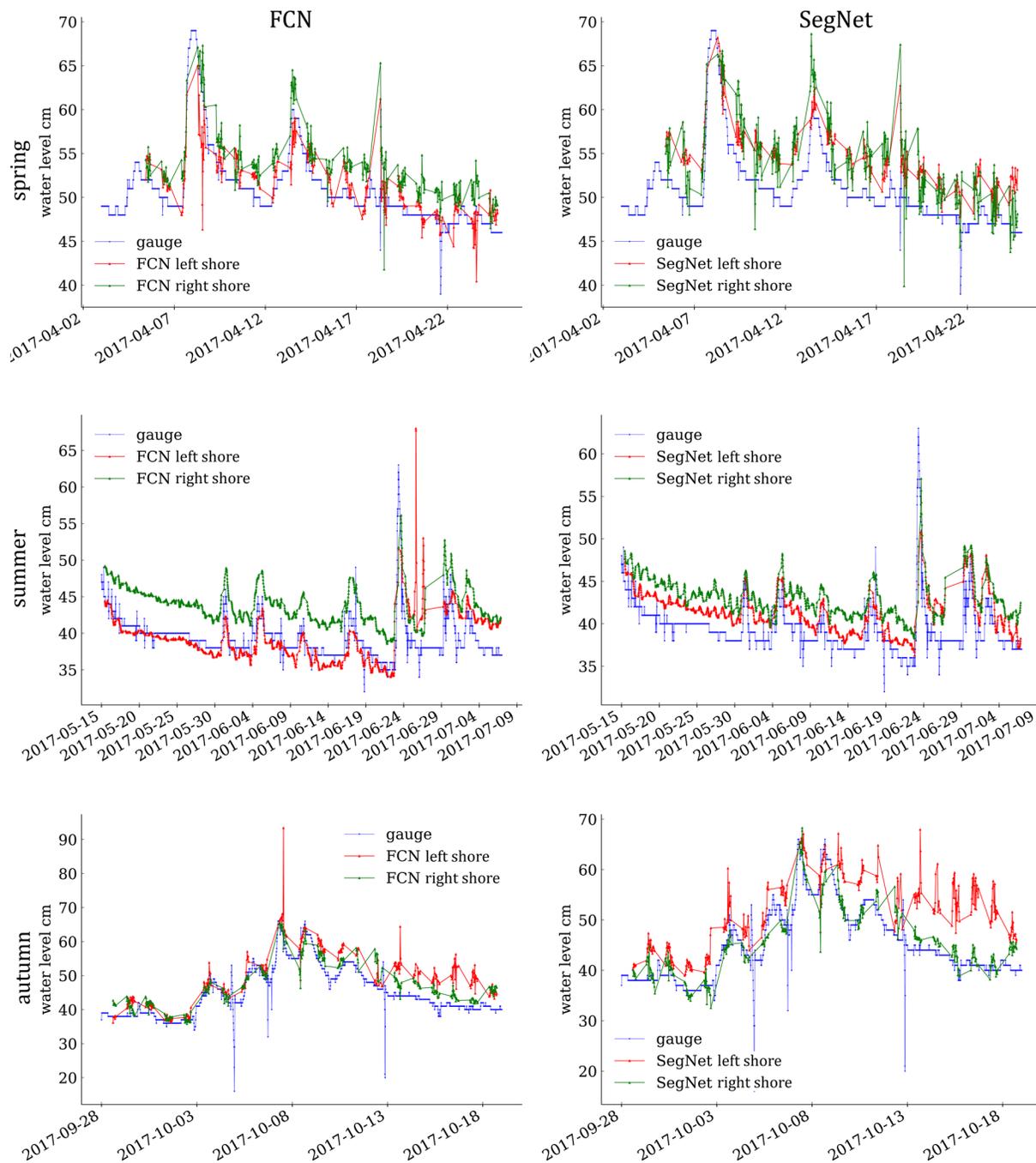


Figure 8. Time series of water stage detection with FCN and SegNet at left and right shore compared to gauge.

because they were used as observation during the adjustment. In future applications, independent check-points would be preferable for more reliable statements in this regard.

The distance between the camera and shores can influence the accuracy of the water stage measurement, as well. At the right shore (more distant from the camera), lower accuracies are obvious for both CNN approaches until early summer. Errors in the camera pose estimation can lead to bias during the intersection of the image observation with the 3D point cloud, and the higher distance from the camera increased those errors. Furthermore, the GSD (i.e., resolution) decreases with increasing distance, also leading to less distinguishable GCPs in the images. Future works can investigate the distance limit to achieve acceptable results.

Environmental conditions influenced the image content, thereby complicating the image segmentation. Very unfavorable lighting conditions, such as extreme shadows and overexposure, affect the image segmentation, and consequently, the water stage estimation. Vegetation growth and illumination difficulties during summer lead to the lowest correlation between reference and camera gauge in that season at both shores considering the automatic and manual water stage measurement approaches. During autumn, lower water stage accuracies and precisions were achieved at the left shore because most challenging light conditions on wet leaves and grass lead to ambiguities. Therefore, the faulty classification of the image content occurred, which was not the case at the right shore that was mostly shadowed and depicted little vegetation. The difficulties of seeing the water line also hindered the manual measurement leading to higher errors for that approach (Table 2, supplement D). Errors of image classification also occurred in summer and autumn, caused by sunny and foggy days, respectively. During the winter season, further challenges are expected. Days are shorter, and thus fewer measurements are available if no additional artificial light source is used. Furthermore, the river and shore might be covered by ice and snow. Temperature conditions have to be considered. For instance, temperature changes can influence the interior camera geometry (Elias et al., 2020).

Another influence on the accuracy of the water stage estimation is the quality of the 3D model. For instance, the high number of outliers of water stage values on the right shore during the spring season was potentially due to the circumstance that the camera was not sufficiently orientated during that period. Hence, only a small part of the paved river reach, to which the corresponding 3D model exists, was visible in the camera. Therefore, the averaging of the corresponding extracted water stages for each point along the water line in the image was more sensitive to outliers in the segmented water area because there were fewer points from which the water stage is estimated from. Furthermore, the resolution and precision of the 3D point cloud are important. The lower the point density, the lower will be the water stage precision. And the more erroneous points are present in the 3D model, the more outliers will be present in the estimated water stages.

The entire contour, that is, water line made of single points, was intersected with the shore, which is complex in this study considering the large cobblestones that lead to height variations in ranges of several centimeters, resulting in noisy water stage values from 3D point cloud intersection. Calculating the median eventually mitigated the impact of the rough terrain. However, using less complicated surfaces, for instance, concrete walls, would lead to higher accuracies because a plane could be fitted into the intersection area, also avoiding the influence of point densities of the 3D model.

Changes of the interest area, for instance, after a flood or due to vegetation growth, entail updating the 3D model for correct water stage measurements. In the future, a solution to this challenge can be the application of at least two cameras (with a known base and interior geometry) to utilize stereo-photogrammetry. Points of the segmented water contour would be matched between both images to let them intersect directly in 3D space without the need for a 3D model. However, as long as monoscopic (one camera) measurements are performed, locations should be preferred where there is no river reach change and minimal vegetation present. Also, approaches can be used that solely provide the river cross-section, for instance using ADCP, as long as the information is provided in the same coordinate system as the exterior camera geometry.

The accuracies achieved by the camera gauge were in the range of the demands of the German gauge manual (LAWA, 2018) in most scenarios, which states that the gauge has to measure the water stage with errors lower than 2.5 cm for 15 min intervals. However, a comparison to the full extent is not possible because the camera captures the water stage at a distinct point in time with a temporal resolution of 30 min. Therefore, the contrast to the reference gauge with 15-min averages of quasicontinuous measurements was limited. The observed bias in water stage measurements for the camera gauge, which is assumed to be due to errors of the camera pose, can be corrected by a single reference measurement if it remains a constant offset.

Previous camera gauge research (Eltner et al., 2018) at this location achieved a correlation of 0.87 for the summer period, revealing that our novel approach allows for similar or even better water stage accordance. Eltner et al. (2018) used an image-based classification attempt relying on the spatiotemporal water texture, and also required image sequences with further processing steps. The previous approach relied on carefully preselected areas for the water line intersection in object space and failed during very challenging environmental conditions. In contrast, the CNN-based methods do not require such preselection, making them more transferrable and more robust on different lighting conditions, facilitating their operationalization.

However, if the water is not visually distinguishable in the images, also the CNN methods will fail as they still rely on the image content. Furthermore, our approach has been developed for images captured during the day. To allow for a continuous river observation the image-based technique needs to be extended to night time photos.

Although the training of the initial CNN has high demands regarding knowledge and hardware, when the CNN has been trained, its application is simple, low-level, and can be used for inference in nonpowerful processors. The aim of this study was to demonstrate the potential of CNNs for robust water segmentation, and we anticipate that future application, and thus an extension of training data, will increase the transferability to many more river reaches to measure water stages. For future work, it is suggested to investigate unsupervised domain adaption and few-shot learning techniques to increase the generalization capacity of the models considering less or even no additional labeled data set from other rivers.

5. Conclusions

In this study, we proposed an image-based approach combining CNN and photogrammetric techniques for water stage retrieval. It was verified that it is possible to measure the water stage with high robustness, achieving error ranges from 1.1 to 3.6 cm. CNNs trained with a sufficiently large labeled data set can be used to segment water and nonwater classes in images. Transfer learning was applied by initializing the first layers of the methods (FCN and SegNet) through a pretrained network. The (with ImageNet) pretrained networks were used to have a good initialization of the weights and then all layers were trained again. This study indicated that regarding the CNN performance, the choice of the network structure (FCN or SegNet) is secondary, whereas the choice of the degree of downsampling of the captured images is important. Higher image resolution requires larger computing capacity but increases the accuracy of segmentation. Correlations between independent gauge measurements and the image-based methods were higher than 0.87 and reach up to 0.93. However, a perfect fit is not possible between reference and our introduced approach due to limits in camera geometry estimation, 3D model accuracies, reference gauge errors, and different temporal resolutions. In the future, the application could be extended to other river observations, including citizen science data, to increase the complexity of the training data set to eventually make the approach transferable to other rivers.

Data Availability Statement

The authors are grateful for data sources provided by the Saxon state company for the environment and agriculture. The processed data is provided by OPARA (<http://dx.doi.org/10.25532/OPARA-72>), and the raw and labeled images are provided by Harvard Dataverse (<https://doi.org/10.7910/DVN/ONOZRW>).

Acknowledgments

The authors thank the European Social Fund (ESF) for funding this project (grant number 100270097). These investigations are part of the research project “extreme events in small and medium catchments (EXTRUSO).” This research was partially funded by CNPq (grant numbers 433783/2018-4, 303559/2019-5) and CAPES Print (grant number 88881.311850/2018-01). The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and CAPES (Finance Code 001). Furthermore, the authors are grateful for the support by Diana Spieler, Felix Binder, Thomas Singer, Jens Grundmann, and Hans-Gerd Maas.

References

- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- Chen, Y., Fan, R., Yang, X., Wang, J., & Latif, A. (2018). Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water*, 10(5), 585. <https://doi.org/10.3390/w10050585>
- Chollet, F. (2015). *Keras*. GitHub Repository. Retrieved from <https://github.com/fchollet/keras>
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368), 829–836. <https://doi.org/10.1080/01621459.1979.10481038>
- Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, 10(4), 2233–2243. <https://doi.org/10.1109/TII.2014.2300753>
- Dietrich, J. T. (2017). Bathymetric structure-from-motion: Extracting shallow stream bathymetry from multi-view stereo photogrammetry. *Earth Surface Processes and Landforms*, 42(2), 355–364. <https://doi.org/10.1002/esp.4060>
- Elias, M., Eltner, A., Liebold, F., & Maas, H. G. (2020). Assessing the influence of temperature changes on the geometric stability of smartphone-and Raspberry Pi cameras. *Sensors*, 20(3), 643. <https://doi.org/10.3390/s20030643>
- 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. <https://doi.org/10.1029/2018WR023913>
- Eltner, A., Kaiser, A., Castillo, C., Rock, G., Neugirg, F., & Abellan, A. (2016). Image-based surface reconstruction in geomorphometry – merits, limits and developments. *Earth Surface Dynamics*, 4, 359–389. <https://doi.org/10.5194/esurf-4-359-2016>
- Eltner, A., Sardemann, H., & Grundmann, J. (2020). Technical Note: Flow velocity and discharge measurement in rivers using terrestrial and UAV imagery. *Hydrology and Earth System Sciences*, 24, 1429–1445. <https://doi.org/10.5194/hess-24-1429-2020>

- Fang, W., Wang, C., Chen, X., Wan, W., Li, H., Zhu, S., et al. (2019). Recognizing global reservoirs from Landsat 8 images: A deep learning approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9), 3168–3177. <https://doi.org/10.1109/JSTARS.2019.2929601>
- Feng, W., Sui, H., Huang, W., Xu, C., & An, K. (2018). Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model. *IEEE Geoscience and Remote Sensing Letters*, 16(4), 618–622. <https://doi.org/10.1109/LGRS.2018.2879492>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Heipke, C., & Rottensteiner, F. (2020). Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. *Geo-spatial Information Science*, 23(1), 10–19. <https://doi.org/10.1080/10095020.2020.1718003>
- Herschty, R. W. (2008). *Streamflow measurement* (3rd ed., p. 510). CRC Press.
- Isikdogan, F., Bovik, A. C., & Passalacqua, P. (2017). Surface water mapping by deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11), 4909–4918. <https://doi.org/10.1109/JSTARS.2017.2735443>
- Jiang, W., He, G., Long, T., Ni, Y., Liu, H., Peng, Y., et al. (2018). Multilayer perceptron neural network for surface water extraction in Landsat 8 OLI satellite images. *Remote Sensing*, 10(5), 755. <https://doi.org/10.3390/rs10050755>
- Kopp, M., Tuo, Y., & Disse, M. (2019). Fully automated snow depth measurements from time-lapse images applying a convolutional neural network. *Science of The Total Environment*, 697, 134213. <https://doi.org/10.1016/j.scitotenv.2019.134213>
- Kraus, K. (2007). *Photogrammetry: Geometry from images and laser scans* (2nd ed., p. 459). Berlin, Germany: Walter de Gruyter.
- Kröhnert, M., & Eltner, A. (2018). Versatile mobile and stationary low-cost approaches for hydrological measurements. *ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2, 543–550. <https://doi.org/10.5194/isprs-archives-XLII-2-543-2018>
- LAWA Bund/Länder-Arbeitsgemeinschaft Wasser (2018). *Leitfaden zur Hydrometrie des Bundes und der Länder – Pegelhandbuch*. Stuttgart, Germany: Kulturbuch-Verlag GmbH.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Leduc, P., Ashmore, P., & Sjogren, D. (2018). Stage and water width measurement of a mountain stream using a simple time-lapse camera. *Hydrology and Earth System Sciences*, 22(1), 1–11. <https://doi.org/10.5194/hess-22-1-2018>
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Paper presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, MA: CVPR.
- Luhmann, T., Robson, S., Kyle, S., & Boehm, J. (2014). *Close-range photogrammetry and 3D imaging* (2nd ed., p. 683). Berlin, Germany: Walter de Gruyter.
- Morgenschweis, G. (2010). *Hydrometrie* (p. 582). Berlin Heidelberg: Springer-Verlag.
- Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Paper presented at the 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: ICCV.
- Pan, J., Yin, Y., Xiong, J., Luo, W., Gui, G., & Sari, H. (2018). Deep learning-based unmanned surveillance systems for observing water levels. *IEEE Access*, 6, 73561–73571. <https://doi.org/10.1109/ACCESS.2018.2883702>
- Ran, Q. H., Li, W., Liao, Q., Tang, H. L., & Wang, M. Y. (2016). Application of an automated LSPIV system in a mountainous stream for continuous flood flow measurements. *Hydrological Processes*, 30(17), 3014–3029. <https://doi.org/10.1002/hyp.10836>
- Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671–675. <https://doi.org/10.1038/nmeth.2089>
- Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., & Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Paper presented at the 2006 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY.
- Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556.
- Stumpf, A., Augereau, E., Delacourt, C., & Bonnier, J. (2016). Photogrammetric discharge monitoring of small tropical mountain rivers: A case study at Rivière des Pluies, Réunion Island. *Water Resources Research*, 52(6), 4550–4570. <https://doi.org/10.1002/2015WR018292>
- Suzuki, S., & Abe, K. (1985). Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1), 32–46.
- Torres, D. L., Feitosa, R. Q., Happ, P. N., La Rosa, L. E. C., Marcato Junior, J., Martins, J., et al. (2020). Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. *Sensors*, 20(2), 563. <https://doi.org/10.3390/s20020563>
- Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153), 405–426. <https://doi.org/10.1098/rspb.1979.0006>
- Wada, K. (2018). *Labelme: Image Polygonal Annotation with Python*. GitHub Repository. Retrieved from <https://github.com/wkentaro/labelme>
- Young, D. S., Hart, J. K., & Martinez, K. (2015). Image analysis techniques to estimate river discharge using time-lapse cameras in remote locations. *Computers & Geosciences*, 76, 1–10. <https://doi.org/10.1016/j.cageo.2014.11.008>