Federal Public Service
Ministry of Education
Federal University of Mato Grosso do Sul
Postgraduate Program in Environmental Technologies

Pedro Alberto Pereira Zamboni

ENHANCING HYDROLOGICAL MONITORING THROUGH DEEP LEARNING AND PHOTOGRAMMETRY

Campo Grande, MS
November, 2023

Federal University of Mato Grosso do Sul
Faculty of Engineering, Architecture and Urbanism, and Geography
Postgraduate Program in Environmental Technologies

Pedro Alberto Pereira Zamboni

ENHANCING HYDROLOGICAL MONITORING THROUGH DEEP LEARNING AND PHOTOGRAMMETRY

Thesis submitted to the Faculty of Engineering, Architecture and Urbanism, and Geography, Federal University of Mato Grosso do Sul, in partial fulfillment of the requirements for the Degree of Doctor in Science: Water Resources and Environmental Sanitation.

Advisor: Prof. Dr. José Marcato Junior
Co-Advisor: Prof. Dr. Wesley Nunes Gonçalves
Co-Advisor: Profa. Dra. Anette Eltner

Approved in: 22/11/2023

Examination board

Dr. José Marcato Junior
President

Prof. Dr. Hemerson Pistori
UCDB

Profa. Dra. Ana Paula Marques Ramos
Unesp

Prof. Dr. Paulo Tarso Sanches de Oliveira
UFMS

Dr. André Almagro
UFMS

Campo Grande, MS
November, 2023

DEDICATION

To everyone who came before me. To everyone who believed in me.

ACKNOWLEDGMENTS

First and foremost I am extremely grateful to my advisors, Prof. Dr. José Marcato Junior and Prof. Dr. Wesley Nunes Gonçalves for their support, advice, and patience during my PhD study. Their knowledge and passion were a great source of inspiration. I will always be grateful for the patience and the availability to share their knowledge without vanity and selfishness. The meetings to show me help me to resolve some of my problems with coding to all the conversations about life helped me immensely to stay strong during the last almost four years. I finish my PhD studies being a better researcher and a better person.

I am also deeply grateful to Prof. Dr. Anette Eltner and Geosensor Systems research group for receiving, for the partnership, unconditional support, and shared experiences with me during my exchange in Germany. Without the help of Anette and her group finishing my PhD will not be possible. My time there not only changed my personal but also my professional life. I will always remember my time in Germany and the friends that I carried with me.

To my colleagues in Geomatics and Computer Vision Research Group, with a special thanks to my friends Maxillian, João and Vitor, thank you or all for all the friendship and help. For my friends from Hydrology, Erosion and Sediment Laboratory (HEroS) and Hydrology, Water Security research group, and EESC/USP: Prof. Dr. Paulo Tarso, Dr. André Almagro, Dr. Jamil Anache, Dr. Gabriela Gesualdo, Dr. Jullian Sone, and MS.c. Rodrigo Bahia, my special thanks for all these years of friendship, for all the time, field work, and coffees that we had together. You guys have a special place in my life.

I would like to thanks the professors, colleagues, technical staff, and secretaries from the FAENG and the Federal University of Mato Grosso do Sul, for the support during the doctoral course, to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the financial support, and the Dresden University of Technology for receiving me for my exchange.

Last but not least, I would like to thank my father, my mother and siblings. I have an amazing family, unique in many ways, and their support has been unconditional all these years. Their support and love has enabled me to achieve everything that I have. It was a long road until we got here.

"I think it's much more interesting to live not knowing than to have answers which might be wrong. I have approximate answers and possible beliefs and different degrees of uncertainty about different things, but I am not absolutely sure of anything and there are many things I don't know anything about ..."

(Richard Feynman)

RESUMO

Zamboni, P. (2021). **Melhoria do monitoramento hidrológico através da aprendizagem profunda e da fotogrametria**. Tese de Doutorado, Faculdade de Engenharias, Arquitetura e Urbanismo, e Geografia, Universidade Federal de Mato Grosso do Sul, Campo Grande, MS. Brasil.

A observação dos componentes do ciclo hidrológico podem ser desafiadoras devido à escala em que ocorrem e ao custo dos sensores. Medir a formação de escoamento superficial e vazão são chave para a compreensão da dinâmica da água, uma vez que influencia também as atividades humanas, a fim de manter ecossistemas naturais equilibrados. O principal objetivo desta tese de doutorado é propor abordagens de aprendizagem profunda combinadas com fotogrametria para medir automaticamente a formação de escoamento superficial e vazão. Nossos resultados sugerem que considerar o desequilíbrio de classe e a incerteza do rótulo durante o treinamento de aprendizagem profunda para segmentar áreas de poças de água é mais importante do que a própria rede, bem como ensembles. Área, número e conectividade das poças de água e a sua foram comparados com medida da vazão, onde foram encontrados diferentes comportamentos em relação à geração de escoamento superficial. Em relação à vazão, nossos resultados mostraram que tanto STCN quanto SAM utilizando pontos fixos e SAM combinado com Dino alcançaram resultados satisfatórios para segmentação de água, mesmo com conjunto de dados de rótulo mínimo ou não anotado. As medidas dos níveis de água utilizando estas máscaras resultam num bom ajuste com os dados de referência, sendo capazes de capturar alterações no fluxo de água, especialmente para níveis de água mais elevados. Nas imagens dinâmicas, STCN e SAM Dino obtiveram resultados semelhantes, entretanto a escolha do primeiro frame influencia os resultados da STCN. Os resultados encontrados nesta tese de doutorado abrem uma nova fronteira para hidrólogos e tranquilizadores da ciência do solo com a possibilidade de medir diretamente a formação de escoamento superficial e uma solução mais barata e escalável para medição de vazão.

Palavras-chave: hidrologia, escoamento superficial, vazão, aprendizagem profunda, fotogrametria.

ABSTRACT

Zamboni, P. (2021). **Enhancing hydrological monitoring through deep learning and photogrammetry** Doctoral Thesis, Faculty of Engineering, Architecture and Urbanism, and Geography, Federal University of Mato Grosso do Sul, Campo Grande, MS. Brazil.

Observing components of the hydrological cycle can be challenging due to the escalation that occurs and the cost of sensors. Measuring the formation of surface runoff and flow is fundamental to understanding water dynamics, as it also influences human activities in order to keep natural ecosystems balanced. The main objective of this doctoral thesis is to propose deep learning approaches combined with photogrammetry to automatically measure surface teaching formation and flow. Our results suggest that considering class imbalance and label uncertainty when training deep learning to segment water pocket areas is more important than the network itself as well as ensembles. Area, number and connectivity of water pools and their comparison with the flow measurement, where different behaviors were found in relation to the generation of surface runoff. Regarding flow rate, our results demonstrated that both STCN and SAM using fixed points and SAM combined with Dino achieved overwhelming results for water segmentation, even with minimal or unannotated label dataset. Measurements of water levels using these masks resulted in a good fit with reference data, being able to capture changes in water flow, especially at higher water levels. In dynamic images, STCN and SAM Dino obtained similar results, however the choice of the first frame influenced the STCN results. The results found in this doctoral thesis open a new frontier for hydrologists and soil science practitioners with the possibility of directly measuring surface runoff formation and a cheaper and more scalable solution for flow measurement.

Keywords: hydrology, runoff, streamflow, deep learning, photogrammetry.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

LIST OF TABLES

CHAPTER 1

CHAPTER 2

GENERAL INTRODUCTION

1. Background and problem statement

Through the water cycle, rainfall is partitioned into surface runoff, subsurface runoff, and underground flow (Le Mesnil et al., 2021). When rainfall intensity overcomes infiltration capacity, water starts to accumulate into soil depression forming ponding. Subsequently, runoff is generated once rainfall surpasses the depression storage and infiltration capacity (Yang et al., 2015). Surface runoff influences water resources, ecosystem and human well-being and safety (Gulahmadov et al., 2021), and it is responsible for triggering processes such as soil erosion. For instance, soil erosion causes loss of soil nutrients, reducing field productivity. Moreover, it can lead to an increase of greenhouse gas emissions (Lugato et al., 2018).

Runoff generation processes can be observed in different scales, from micro-scale up to catchment scale, being the rainfall simulation on small-scale plots one of the most common methods (Falcão et al., 2020; Abudi et al., 2012). Nevertheless, these runoff generation vary at a very fine scale. It is well known that micro-topography is one of the most significant features controlling the generation of surface runoff (Dunne et al., 1991; Govers et al., 2000; Kværner and Kløve, 2008; Turunen et al., 2020). Nevertheless, limitations in measurement methods limit the understanding of this process, especially at a micro-scale. Traditional approaches are unable to observe and measure direct effects of micro-topography during rainfall partitioning, ponding formation, its connection and runoff generation. Eventually surface runoff is concentrated into the natural drainage (river network) and becomes streamflow. Understanding runoff generation is key to better comprehend streamflow patterns since different mechanisms lead to different floods (Stein et al., 2020).

Streamflow results from precipitation with a time delay, yet this connection is influenced by factors like basin size, topography, soils, vegetation and spatial heterogeneity (Bales et al., 2018). Floods triggered by rainfall stand out as one of the most prevalent forms of natural disasters, possessing the highest potential for damage compared to all other natural disasters globally (WMO, 2013). River floods not only cause significant immediate damages and loss of life but also entail broader and more prolonged adverse economic repercussions

(Koks and Thissem, 2016). To model and plan water resources effectively, it is essential to establish hydrological observation networks that are dense in both space and time, ensuring a comprehensive database (Eltner et al., 2021). However, water stage monitoring networks usually are composed of a small number of sensors. In many instances, measurements in hydrology primarily focus on large-scale catchments, or gauge positions are selected with a focus on water management concerns rather than addressing hydrological research questions (Kirchner, 2006). In Brazil, urban water stage monitoring networks are almost non-existent due to difficulties in deploying such networks in urban areas. Traditional river monitoring sensors (pressure gauges) require maintenance and are prone to oxidation, may be lost during a flood event (Eltner et al., 2021) and prone to vandalism (Vitry et al., 2019). Moreover, this kind of sensor can be expensive, mainly in developing countries.

In this context, images emerge as an remote sensing alternative with a high potential to improve our understanding of such phenomena, allowing for a direct measurement. For instance, images that capture soil surface during rainfall events hold significant potential for enhancing the comprehension of runoff generation and infiltration processes. This monitoring allows for direct observation of the formation of standing and flowing water. In order to build operation tools capable of processing this data, automatic methods need to be explored. In recent years, mainly Convolutional Neural Networks (CNNs) and Transformer based models have greatly impacted environmental sciences as deep learning techniques (Persello et al., 2022, Heipke & Rottensteiner, 2019). Works have explored deep learning approaches in several hydrological applications such as rainfall measurement (Yin et al., 2023), water stage retrieving (Vandaele et al., 2021; Virty et al., 2019) and surface water velocity measurement (Ansari et al., 2023). Even though it is a powerful methodology, deep learning and image processing only get us so far with measurements in terms of pixels.

Photogrammetry has fastly grown as an essential tool in geosciences (Eltner et al., 2016), playing an important role in monitoring natural hazards (Blanch et al., 2023). Photogrammetry methodologies allow us to transform measurement in a pixel resolution originated from images into metric values. One of the main ideas relies on the projection on points from an 2D space (e.g., water contour automatic segmented using deep learning and images) into a 3D point cloud, assessing the n number of nearest points from the 3D point cloud to the project 2D point. Combining deep learning and photogrammetry for automatic image processing opens a new world of possibilities in terms of hydrological monitoring as demonstrated by Eltner et al. (2021). Notwithstanding, further development needs to be done to extend the impact of this combined methodology. Regarding deep learning, training strategies need to be studied and further developed to deal with challenge scenarios, i.e. segmenting water ponding, and strategies to reduce the need of large annotated dataset.

To the best of our knowledge, no methodology allows for a direct observation and measurement of the process involved in runoff generation. Additionally, there is a large gap in observed hydrological data, especially water stage. Given the importance of water resources for the general well-being of natural ecosystems to support human activities, it is necessary to develop new and cheaper ways of hydrological monitoring. This doctoral thesis aims to advance methods to measure and understand components of the hydrological cycle related to runoff generation and streamflow by combining deep learning methods and photogrammetry. By exploring different approaches combining state-of-the-art deep learning methods and photogrammetry, this research seeks to address the challenge posed in hydrological monitoring on different scales. The adoption of these methods can increase the understanding of how rainfall is transformed into surface runoff and later into streamflow, further, helping to adopt measures to reduce impact caused by both.

2. Objectives

2.1. General objective

The main objective of this study is to propose deep learning approaches combined with photogrammetry to automatically measure runoff formation and streamflow.

2.2. Specific objectives

i. To evaluate the performance CNNs to water segmentation in time-lapse rainfall simulation plots, evaluating label imbalance, patch spatial correlation during training, and ensemble models;

ii. To evaluated runoff generation pattern in different plots under simulated rainfall;

iii. To evaluate potential of pre-trained large models in water segmentation of river images;

iv. To assess the potential of using water masks segmented by video object segmentation and pre-trained large models in automatic water stage retrieve.

3. Organization of thesis

The Thesis is organized into two chapters. In **Chapter 1**, we investigate the potential of deep learning in the segmentation of water areas in time-lapse images captured during rainfall simulations. First, we evaluated three different CNNs, trained considering label unbalanced as well spatial correlation between samples. Ensemble models were also evaluated. We evaluated the transferability of all trained models in two different plots. Using the best model, water pixel areas were identified. Finally, we compared the water area with measured runoff, and computed the ponding time and the number of connected components. In **Chapter 2** we evaluated CNNs methods that require minimum to non annotated dataset to retrieve water masks from camera gauge stations and Unmanned Aerial Vehicles datasets. First, a video object segmentation network used to segment time-lapse images was evaluated. This method only used one annotated image (the initial frame) to produce masks for the entire image sequence. Further, we evaluated state-of-the-art image segmentation deep learning approaches that work with user input (i.e. points, bounding box, or text). Segmented water masks were evaluated both qualitatively and quantitative. Finally, we used segmented water masks for river delineation and water stage retrive.

4. References

Le Mesnil, M., Moussa, R., Charlier, J.-B., & Caballero, Y. (2021). Impact of karst areas on runoff generation, lateral flow and interbasin groundwater flow at the storm-event timescale. Hydrology and Earth System Sciences, **25** (3), 1259–1282. https://doi.org/10.5194/hess-25-1259-2021

Persello, C., Wegner, J. D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., & Camps-Valls, G. (2022). Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. IEEE Geoscience and Remote Sensing Magazine, **10** (2), 172-200. https://doi.org/10.1109/MGRS.2021.3136100

Vandaele, R., Dance, S. L., and Ojha, V. (2021). Deep learning for automated river-level monitoring through river-camera images: an approach based on water segmentation and transfer learning, Hydrol. Earth Syst. Sci., **25**, 4435–4453. https://doi.org/10.5194/hess-25-4435-2021.

Yang, W.-Y., Li, D., Sun, T., & Ni, G.-H. (2015). Saturation-excess and infiltration-excess runoff on green roofs. Ecological Engineering, **74**, 327–336. https://doi.org/10.1016/j.ecoleng.2014.10.023

Moy de Vitry, M., Kramer, S., Wegner, J. D., and Leitão, J. P. (2019). Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network, Hydrol. Earth Syst. Sci. **23**, 4621–4634. https://doi.org/10.5194/hess-23-4621-2019

Gulahmadov, N., Cheng, Y., Gulakhmadov, A., Rakhimova, M., Gulakhmadov, M (2021). Quantifying the relative contribution of climate change and anthropogenic activities on runoff variations in the central part of Tajikistan in Central Asia, Land. **10** (5), 525. https://doi.org/10.3390/land10050525

Stein, L., Pianosi, F., Woods, R. (2020). Hydrol. Process. 34, 1514–1529

Lugato, E., Smith, P., Borrelli, P., Panagos, P., Ballabio, C., Orgiazzi, A., Fernandez-Ugalde, O., Montanarella., L., Jones, A. (2018). Soil erosion is unlikely to drive a future carbon sink in Europe, Sci. Adv. **4**. https://doi.org/10.1126/sciadv.aau3523

Dunne, T., Zhang, W., Aubry, B. F. (1991). Effects of rainfall, vegetation, and microtopography on infiltration and runoff, Water Resour. Res., **27** (9), pp. 2271-2285

Govers, G., Takken, I., Helming, K. (2000). Soil roughness and overland flow, Agronomie, **20** (2), pp. 131-146

Kværner, J., Kløve. B. (2008). Generation and regulation of summer runoff in a boreal flat fen, J. Hydrol., **360** pp. 15-30, 10.1016/j.jhydrol.2008.07.009

Turunen, M., Turtola, E., Vaaja, M.T., Hyväluoma, J., Koivusalo. H. (2020). Terrestrial laser scanning data combined with 3D hydrological modeling decipher the role of tillage in field water balance and runoff generation, Catena, **187**, 104363. https://doi.org/10.1016/j.catena.2019.104363

Bales, R. C., Goulden, M. L., Hunsaker, C. T., Conklin, M. H., Hartsough, P. C., OGeen, A. T., Hopmans, J. W., & Safeeq, M. (2018). Mechanisms controlling the impact of multi-year drought on mountain hydrology. Scientific Reports, **8**, 690. https://doi.org/10.1038/s41598-017-19007-0

Flood Forecasting and Early Warning. World Meteorological Organization, 2013.

Koks, E. E. & Thissen, M. A. (2016). Multiregional impact assessment model for disaster analysis. Econ. Syst. Res. **28**, 429–449. https://doi.org/10.1080/09535314.2016.1232701

Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research, **42**. https://doi.org/10.1029/2005WR004362

Eltner, A., Kaiser, A., Castillo, C., Rock, G., Neugirg, F., Abellán, A. (2016). Image-based surface reconstruction in geomorphometry – merits, limits and developments, Earth Surf. Dynam., **4**, 359–389, https://doi.org/10.5194/esurf-4-359-2016,

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

Abudi, I., Carmi, G., Berliner, P. (2012). Rainfall simulator for field runoff studies, Journal of Hydrology, **454-455**, 76-81. https://doi.org/10.1016/j.jhydrol.2012.05.056

Falcão, K., Panachuki, E., Monteiro, F., Menezes, R., Rodrigues, D., Sone, J., Oliveira, P. (2020). Surface runoff and soil erosion in a natural regeneration area of the Brazilian Cerrado, International Soil and Water Conservation Research, **8** (2), 124-130. https://doi.org/10.1016/j.iswcr.2020.04.004

Yin, H., Zheng, F., Duan, H., Savic, D., Kapelan, Z. (2023). Estimating Rainfall Intensity Using an Image-Based Deep Learning Model, Engineering, **21**, 162-174. https://doi.org/10.1016/j.eng.2021.11.021

Ansari, S., Rennie, C. D., Jamieson, E. C., Seidou, O., & Clark, S. P. (2023). RivQNet: Deep learning based river discharge estimation using close-range water surface imagery. Water Resources Research, 59, e2021WR031841. https://doi.org/10.1029/2021WR031841

CHAPTER 1

Measuring water ponding time, location and connectivity on soil surfaces using time-lapse images and deep learning

Zamboni, P., Blümlein, M., Lenz, J., Gonçalves, W. N., Junior, J. M., Wöhling, T., Eltner, A. Measuring water ponding time, location and connectivity on soil surfaces using time-lapse images and deep learning. Under revision in International Soil and Water Conservation Research (Impact factor: 6.4)

Abstract

Rainfall simulations are an established method to gain knowledge on small-scale hydrological processes like infiltration, ponding and the formation of surface runoff. Due to limitations in measuring methods, these processes must usually be understood to happen homogeneously within the bounded plot area while it is well known that they actually vary on a subplot scale. Within this study we took high resolution time-lapse images of several plots to observe and quantify the subplot processes of ponding and the formation of connectivity and surface runoff. We investigated the potential of deep learning in the segmentation of water ponding areas in time-lapse images during rainfall simulations and to estimate the ponding time. We trained three different Convolutional Neural Networks (CNNs), considering classification uncertainty and imbalance of the ground-truth data (water pixels) as well as ensemble modeling and spatial correlation between samples. Our findings suggest that addressing ground-truth annotation uncertainty and imbalance was more important in our study than the choice of the CNNs itself, and ensemble models increase the model performance leading to more robust predictions. Overall, our results suggest that CNNs have a great potential to segment ponding areas, and thus it is possible to observe their spatio-temporal evolution. When comparing the evolution of water ponding areas to runoff, different behaviors across the plots were observable, which could be related to differences in initial soil moisture and infiltration behaviors. Further, our image-based deep learning approach allows for direct measurement of the ponding time, and can be considered a first step to spatially and temporally resolved mapping of infiltration rates.

Keywords: Climate change, general circulation model, rainfall, regional climate model.

1. **Introduction**

Enhancing our understanding of runoff processes is a central objective within hydrological and geomorphological sciences (Beven, 2021; Sone et al., 2020; Vlček et al., 2022; Zhao et al., 2018). One goal is to comprehend the mechanisms involved in the partitioning of precipitation into the surface, subsurface, or underground flow (Le Mesnil et al., 2021). However, infiltration and runoff processes on soils are complex and controlled by various factors, e.g., soil type, slope, microtopography, land use and cover, and rainfall intensity. Traditional approaches to measure runoff on sloping soils mostly provide local values, and observing the spatiotemporal dynamics of water on the soil surface (e.g., water pond formation) continues to pose a challenge. For instance, time to ponding is usually a subjective field measurement due to unclear definitions, resulting often in a rough estimation by differencing infiltration and runoff rate (Fiener et al., 2011). In this sense, images capturing the soil surface during rainfall events have a high potential to improve the understanding of runoff generation and also infiltration processes as the formation of standing and flowing water can be observed directly. The analysis of such images remains a tedious work, if done manually - which this study overcomes by providing automatic workflows for image analysis and therefore derived quantifications for relevant processes.

Rainfall water surplus accumulates in soil depressions leading to ponding formation when the rainfall intensity and water accumulation exceeds the infiltration capacity. The ponding time of soil represents the time period between the initiation of rainfall and the occurrence of surface ponding. It indicates when the infiltration capacity of the soil is exceeded, which is a valuable parameter for runoff and infiltration modeling (Assouline et al., 2007). Once the rainfall intensity exceeds the depression storage capacity (DSC) and infiltration capacity, the water ponds start to connect, and runoff is generated (Yang et al., 2015). DSC can be used to support water balance measurements and corresponding models that estimate the time delay of the onset of overland flow (e.g., de Roo et al., 1996).

Information about the temporary water storage and infiltration in surface depressions also helps to track changes in the hydrological connectivity (Darboux and Huang, 2001; Peñuela et al. 2016; Wang et al., 2018). Hydrologic connectivity describes the connection of individual isolated water patches on hillslopes, which is a necessary condition for the formation of runoff (Bracken & Croke, 2007). In addition, it is also an important aspect to consider in soil erosion studies (Baartman et al., 2013). Its development depends on the spatial distribution of water on the soil surface (McNamara et al., 2005; Wilson et al., 2016), micro-topography (Ali and Roy, 2009), and its changes in four dimensions; longitudinal, transversal, vertical, and temporal (Wu et al., 2021).

In addition to measuring water ponds and their interconnections during rainfall events, simply identifying the area covered by water can be an important information for soil erosion studies. Many raster based soil erosion models internally estimate the areas covered by water due to its relevance for soil particle detachment and transport (e.g. RillGrow - Favis-Mortlock et al., 1998, LISEM - Jetten & De Roo, 2001, EROSION-3D - von Werner & Schmidt, 1996). Thus, the image-based water coverage observation can be an important parameter to validate and calibrate such models.

Recent developments in computer vision and deep learning, especially Convolutional Neural Networks (CNNs), have greatly impacted environmental sciences (Persello et al., 2022, Heipke & Rottensteiner, 2019). Researchers have explored the potential of CNNs for scene classification (Carvalho et al., 2022; Su et al., 2021), image segmentation (Brandt et al., 2020; Nguyen et al., 2022), and object detection (Higa et al., 2022; Jing et al., 2021; Zamboni et al., 2021), in images captured from satellites, unmanned aerial vehicles (UAVs) and with stationary cameras. Regarding water segmentation, CNNs have been used to automatically identify water in images (Wagner et al., 2023) to perform water level measurements with camera gauges (Eltner et al., 2021; Vandaele et al., 2021), and flood mapping with aerial images, either using UAV (Gebrehiwot et al., 2019; Ichim & Popescu, 2020) or satellite (Konapala et al., 2021; Muñoz et al., 2021) platforms. Nevertheless, the application of these methods to observe ponding and runoff formation at soil surfaces still needs to be seen, especially considering the challenges of distinguishing wet from water-covered soil.

Compared to other image water segmentation cases (e.g., Eltner et al., 2021), the water area on the soil surface during a simulated rainfall event, as conducted in this study, is considerably small. Furthermore, identifying and labeling water ponding areas presents a significant challenge due to their small scale, complex shapes, similar color gradients at the class boundaries, and class imbalance. In images from rainfall simulations, water represents a small fraction of the pixels for the majority of the rainfall event, leading to an imbalance when compared to the rest of the image pixels. The performance of CNNs can decrease by imbalances in the training sample distribution across classes (López et al., 2013). Therefore, it is necessary to investigate methodologies that consider class imbalance and labeling uncertainty while learning the CNN models.

Moreover, when using deep learning in geospatial applications, the spatial correlation of sampled objects has to be considered. Usually, samples, e.g., images, for training and testing are drawn randomly across a region of interest, assuming independence. However, employing images not used during the model training does not ensure that this sample is independent. Dependence among samples can emerge due to spatial proximity because close objects are more correlated than distant ones (Tobler, 1970). Generally, using non-independent data with a different distribution (e.g., nearby samples in a spatial cluster) during testing can lead to biased results and overly optimistic model results (Kattenborn et al., 2022; Meyer & Pebesma, 2022), with differences in performance up to almost 50% (Schratz et al., 2019).

The contribution of this study is the demonstration of the potential of computer vision to estimate the water ponding time and assess runoff formation behaviors. CNNs are adapted to segment water ponds on the soil surface in time-lapse images to differentiate between water retention and surface runoff measured at the plot outlet. CNNs were trained to automatically segment the water areas on the soil, acknowledging class imbalance and labeling uncertainty. We assess the impact of spatial correlation between training and testing samples and the transferability of the models to new and unseen plots. The best models were used to segment water in orthorectified time-lapse images allowing for the scaled estimation of water coverage and assessing changes in the connectivity of individual water ponds. Finally, we estimated the ponding time and compared it with ponding time derived using rainfall simulation. The proposed method can help to better understand how runoff is generated by understanding how ponding and connected ponds occur at different temporal and spatial scales. This study is a proof of concept of how CNN-driven, spatiotemporal separation of ponded plot areas can help a better understanding of runoff-generating processes.

2. Material and methods

Our work was carried out in four steps (Figure 1). First, we conducted rainfall simulations, collecting soil surface images, measuring the runoff, and collecting ground control points (GCPs). Second, we manually annotated 83 images from three plots (plots 1 to 3) with binary masks (background and water areas), created two datasets, and performed model training and testing. The first dataset was created by randomly separating sample image patches into training, validation, and test sets. The second dataset was split considering spatial correlation of samples. An additional dataset of eight labeled images was annotated from two further unseen plots (from here on referred to as plots 4 and 5), which was not used during training to assess model transferability. We used three well-known deep learning models for water segmentation, i.e., VGG-16, U-Net EfficientNetB0, and U-Net ResNet101. Furthermore, two weighted equations were used for class imbalance and label uncertainty. Finally, the different models were combined into ensembles. Overall, six models (three considering weighted equations and three without this consideration) and four ensembles (considering weighted equations) were generated.

In the third step, the best model was used during inference to segment the water ponding area (producing binary masks) for each plot for all images. Furthermore, we investigated the models' robustness and transferability using unseen plots 4 and 5. Water pixels were further processed to identify connected components, i.e. individual water ponds. Afterwards, the image measurements were scaled via orthorectification. Finally, we compared the predicted water area, pond number and the measured runoff at the plot outlet. We also assessed the ponding time using the discharge measured during rainfall simulation and compared them to the ponding time estimated using the CNN-based water segmentation.

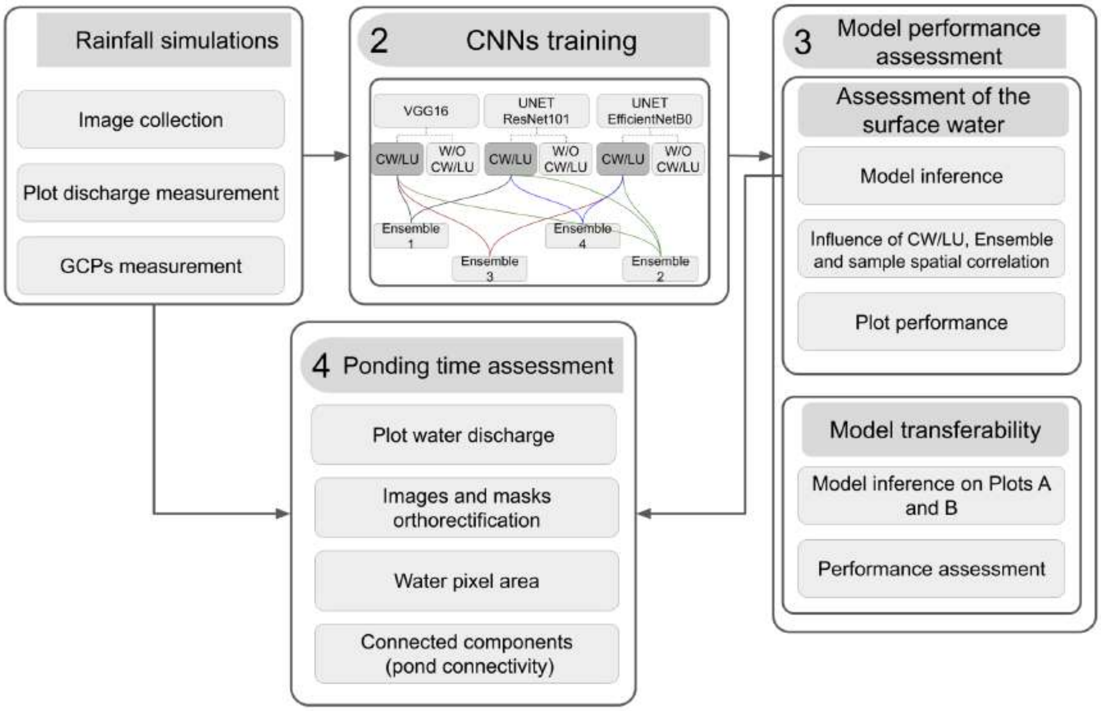


Figure 1: Study workflow. GCPs are ground control points, CW denotes class weight, and LU means label uncertainty.

2.1. Study area and rainfall simulations

Three different soil erosion plots were considered to develop the workflow. The plots are situated in landscapes in Germany, which are prone to soil erosion; i.e., in the quaternary loess belt in Saxony and in geology from the Keuper period in Thuringen. The landscapes are hilly and exhibit grain size distributions dominated by silt or clay in the former and latter regions, respectively. The plots are situated on farmed land managed by different soil tillage practices (tillage or grubber).

Experimental plots with a size of about 3 m² were used. The plots were exposed to artificial rainfall (Figure 2a, e.g., Hänsel et al., 2016) with different intensities ranging from 0.6 to 0.9 mm/min (Table 1). In the first plot, the rainfall experiment lasted about 140 minutes; with a break in between after about one hour, lasting for one hour, and some additional rainfall afterward for another 20 minutes. In the second plot, the rainfall experiment lasted for about 55 minutes. In the third plot, data was captured for the rainfall experiment that ran for 85 minutes. Discharge was measured at the outlet (at the bottom of the plot) every minute during all experiments.

Table 1: Parameters of the single soil erosion plots used during rainfall simulations.

Plot	Capture date	Tillage practice	Cover	Rainfall intensity [mm/min]	Grain size	Initial soil moisture [Vol-%]	Soil bulk density [g/cm3]
1	06.10.2020	Strip till	Bare (10% mulch)	0,6	Tu2	40,5	1,45
2	20.07.2021	Grubber	Bare	0,6	Us	24,9	1,25
3	13.05.2020	Conserving	Field bean (10% mulch)	0,8	Ut3	25,24	1,21
4	06.05.2020	Grubber	Bare (5% mulch)	0,8	Ut3	26,63	1,35
5	22.05.2020	Grubber	Field bean (10% mulch)	0,9	Ut3	25,99	1,29

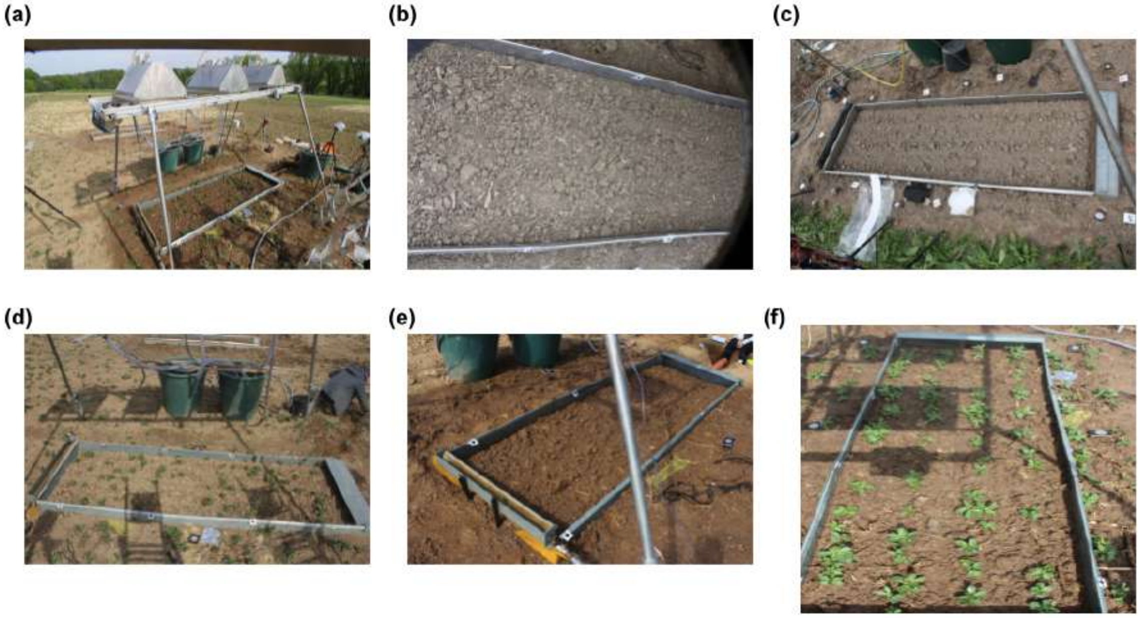


Figure 2: (a) Rainfall simulator system. (b) Plot 1, c) Plot 2, (d) Plot 3, (e) Plot 4, and (f) Plot 5.

2.2. Image acquisition and dataset generation

A SLR camera Canon EOS 450D was used to capture images in plots 1 and 2. The camera has a resolution of 4274x2848 pixels and was equipped with a fixed focal length of 24 mm at plot 1 and 18 mm at plot 2. For the third plot, a Canon 1200D, with a resolution of 5184x3456 pixels and a focal distance of 18 mm, was used. The cameras were mounted at the side of the plots on tripods at heights of about 2 to 3 meters above the ground. At plot 1 images were recorded every 20 seconds during the rainfall and every 2 minutes during the rainfall break. At plot 2 and 3 images were captured every 20 and 10 seconds, respectively. Moreover, a Canon EOS 600D was used to acquire images from plots 4 and5 with a 5184x346 pixels resolution and focal distances of 23 and 9.1 mm.

We used a total of 1388 images (338, 429, and 613 at plots 1, 2, and 3, respectively), from which the water area was annotated manually in 83 images. Additionally, three primary images from plot 4 and five from plot 5 were labeled. To produce a more generalized and robust model, the images include different stages during the rainfall simulation, different perspectives at the area of interest, varying illumination conditions, various soil colors, and different coverages of vegetation.

Two datasets were created for model training using annotated images from plots 1 to 3. The first dataset was created without considering the spatial correlation among samples (Figure 3a). For the second dataset, patches were selected considering the spatial correlation among samples (Figure 3b). We cropped the labeled images in patches with a size of 1024x1024 pixels due to the high resolution of the original images and to guarantee that the water class occupies a feasible amount of pixels, discarding all the patches with no water pixels. An additional dataset, composed of the eighth annotated images from plots 4 and 5, was used to assess model generalization and transferability.

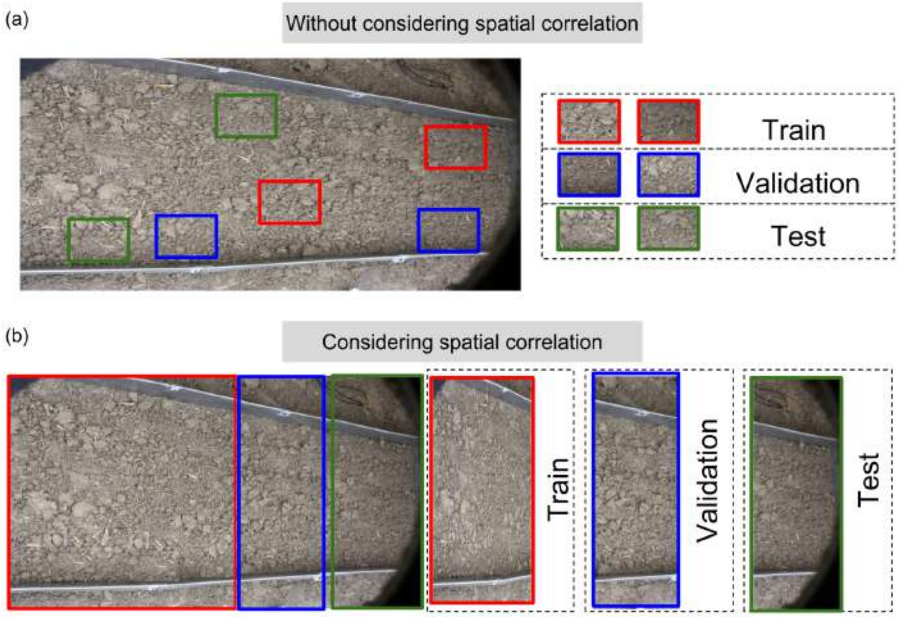

(a)
Without considering spatial correlation
Train
Validation
Test
(b)
Considering spatial correlation
Train
Validation
Test

Figure 3: Example of the splitting process in order to create a dataset without (a) and with (b) consideration of spatial correlation between samples. (a) patches were randomly selected from images to be used for training, validation, and testing of the models. (b) patches were not chosen randomly across the entire plot, but from pre-set regions; the first portion of the images was used for training, the middle part for validation, and the last portion for testing the models.

For the dataset not considering the spatial correlation between samples, a total of 304 patches were generated, where 110, 69, and 125 patches were from plots 1, 2, and 3, respectively. The patches were randomly split into training (60%), validation (20%), and test sets (20%).

Commonly, the performance of deep learning models is assessed using independent test samples. However, this is special for geospatial tasks due to correlations between close pixels and patches or even overlapping patches in a time-series of images. We ensure spatial independence during the random splitting of the patches by keeping training, validation, and test regions apart. We used the first portion of each image for training, the middle part for validation, and the last portion for testing. Thereby, patches used for testing are most distant from training patches. For this dataset, 410 patches (50% for training, 25% for validation, and 25% for testing) were generated.

2.3. Deep learning approach

Models

The water area was segmented in the images using three different CNNs. The first one was based on VGG-16 (Simonyan & Zisserman, 2015), and the second and third ones were U-Net (Ronneberger et al., 2015) networks, using ResNet101 (He et al., 2015) and EfficientNetB0 (Tan & Le, 2019) as backbones. In CNN architectures, backbones are networks used to extract relevant features to encode the input into a feature representation.

CNNs for semantic segmentation are commonly composed of convolutions, batch normalization, activations, pooling, upsampling, and fully convolutional layers. Convolutional layers convert an input volume (i.e., image) into an output volume or feature map by convolving the input volume with a set of learnable filters, where the filters are trained to extract useful information. Usually, convolution layers are followed by batch normalization and an activation function, i.e., Rectified linear unit or ReLU (a nonlinear function defined by f(x)=max(x,0)). Pooling layers reduce the dimensions of the feature map, typically by maintaining the maximum value of the region (max pooling) or the average value of the region (average pooling). Upsampling layers are applied to increase the feature map dimension since pooling layers reduce the dimensions of the feature map. A convolution layer with a kernel of 1 by 1 is used to map the feature map to the desired number of classes.

VGG-16 is a well-known semantic segmentation network, easy to implement, and provides competitive results. For the VGG-16-based architecture, an encoder-decoder network was built. The encoder, responsible for extracting relevant features that characterize the image content, has five convolution blocks with convolutional layers, batch normalization, activation (Relu activation function), and maximum pooling layers. The blocks are composed of different numbers and orders of layers. The decoder, used for upsampling the feature map results from the encoder, is composed of five blocks made of upsampling, convolution, batch normalization, and activation layers, again with varying numbers and order. The final layer of the network is an activation layer using the softmax activation function.

U-Net, along with U-Net variants, is one of the most often-used networks for semantic segmentation (Hu et al., 2021; Shamsolmoali et al., 2020; Ye et al., 2022). U-Net presents a U shape network with a contracting path, i.e., the encoder, and a symmetrical expanding path, i.e., the decoder. We adopted two CNN-based architectures as backbones; ResNet101 (He et al., 2015) and EfficientNetB0 (Tan & Le, 2019). Thus, we build the two U-Net-like networks implementing ResNet101 and EfficientNetB0 as encoders and designing symmetrical versions of them as decoders.

Regarding the U-Net backbone, residual networks, or ResNets, have been explored in image segmentation, revealing good performance (Huang et al., 2020). Using connections across feature maps, ResNets allow the gradient to flow through the skip connection, thus solving the vanishing gradient problem. The vanishing gradient problem is an issue in deep networks, where the gradient, computed from the loss function and used in backpropagation, tends to zero after multiple applications of the chain rule and thus hindering the optimization of parameter weights. In ResNets, skip connections allow for a better gradient flow from the initial filters. ResNets consist of five convolution blocks, followed by average pooling, a fully connected layer, and a final softmax activation function.

EfficientNets is a family of models that introduced a compound scaling method that uniformly scales the network dimensions, i.e., width, depth, and resolution, with a fixed set of coefficients, presenting state-of-the-art results in image segmentation (Atila et al., 2021; Baheti et al., 2020). Thus, new network structures are learned, which depending on the training task, can be potentially more efficient as the networks can turn out smaller and faster. EfficientNet architecture uses mobile inverted bottleneck convolutions (MBConv). MBConv applies an inverted strategy, applying an initial 1x1 convolution, followed by a deep-wise convolution with a kernel size of 3x3 or 5x5, and another 1x1 convolution. Here, we adopted the EfficienNetB0 due to memory limitations; EfficienNetB0 is the lightest model of the EfficientNet family.

Dataset imbalance and label uncertainty

In deep learning methods, the class imbalance can lead to a biased segmentation towards the more dominant class during the inference procedure (López et al., 2013). To reduce the imbalance problem, different strategies have been proposed, e.g., uniformly sampling the dataset (Deng et al., 2009), and rebalancing the dataset by oversampling the minority class or undersampling the majority class (as used by Ravuri et al., 2021). However, these strategies change the data distribution and can disturb training and inference procedures (Dal Pozzolo et al., 2015). Annotating accurate labels for the water area can be challenging for the human eye due to complex boundary gradients between water and soil, as well as the resolution of the images. Figure 4 illustrates the challenge of the labeling task due to the complex shapes of the water ponds and similar color schemes and textures, as well the class imbalance.

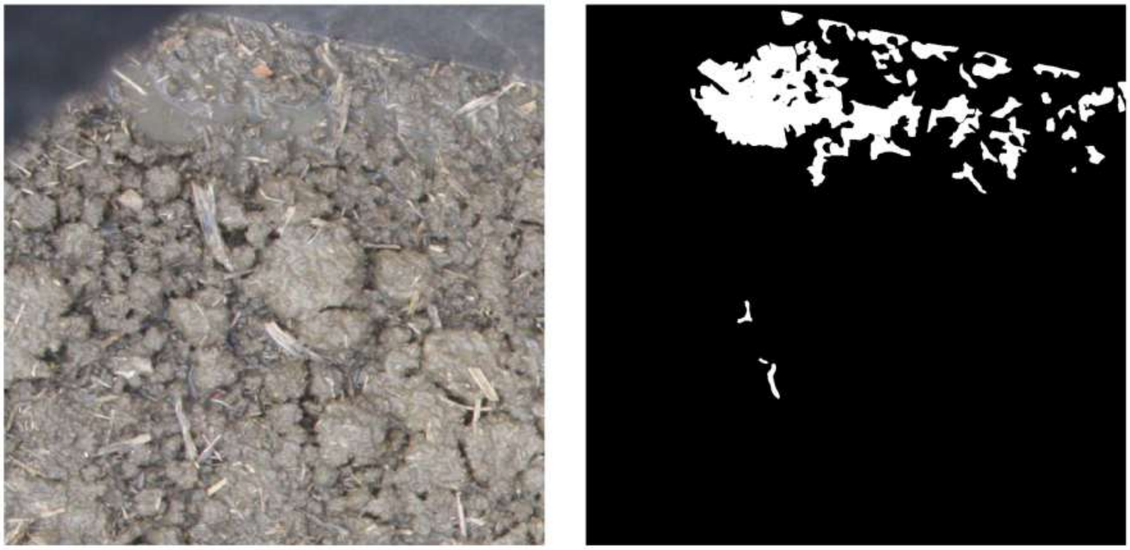


Figure 4: Example of a water mask: on the left, the RGB patch, and, on the right, the corresponding annotated mask is displayed. Due to the complex appearance of water on the soil, creating labels that distinguish between water and background is challenging. The example further shows the class imbalance, where it can be seen that more pixels were assigned as background (shown in black) instead of water (white pixels).

To overcome the issues of class imbalance and label uncertainty, Bressan et al. (2022) proposed a new approach introducing a strategy that considers weights for each image pixel during the loss calculation (Figure 5, Equation 4). Loss or cost functions measure the difference, or error, between the prediction of a neural network and the ground-truth data, where the goal is to minimize the error. In this strategy, pixels that belong to the minority class (water) receive a higher weight to increase their importance. Moreover, pixels close to the object border have higher uncertainty; thus, they have less importance, and their weight is decreased. The equation is expressed as follows:

$$L(\widehat{M},M) \;=\; \frac{1}{n}\sum_{x=1}^{n} \omega(x) \,\cdot\, L(\widehat{M}(x), M(x))\#4$$

Where M is the ground-truth mask, M is the predicted mask, L is the loss function, n is the number of pixels and ω(x) is the pixel weight.

The pixel weight is calculated according to Equation 5.

$$\omega(x) \;=\; \varphi(c(x)) \,\cdot\, \delta(x)\#5$$

where φ(c(x)) refers to class imbalance with c(x) being the class labeled for a given pixel x, and δ(x) is related to the label uncertainty.

The first part, $\varphi(x)$, is calculated using Equation 6.

$$\varphi(x) \;=\; \frac{m}{C * n^{c}}\#6$$

where m is the total number of pixels in all training labels, C is the number of classes, and n^{c} is the number of pixels that belong to class C.

Note that classes with fewer pixels have higher importance. Furthermore, for the pixel in the same class c, the weight φ(x) is the same. The second term, δ(x), is modeled according to Equation 7, where d(x) refers to the distance from a given pixel x to the closest border pixel and is the standard deviation describing the uncertainty buffer of the object border. We used sigma equal to 2, based on the findings of Bressan et al. (2022).

$$\delta(x) \;=\; 1 \;-\; e^{-\frac{d(x)^{2}}{2\sigma^{2}}}\#7$$

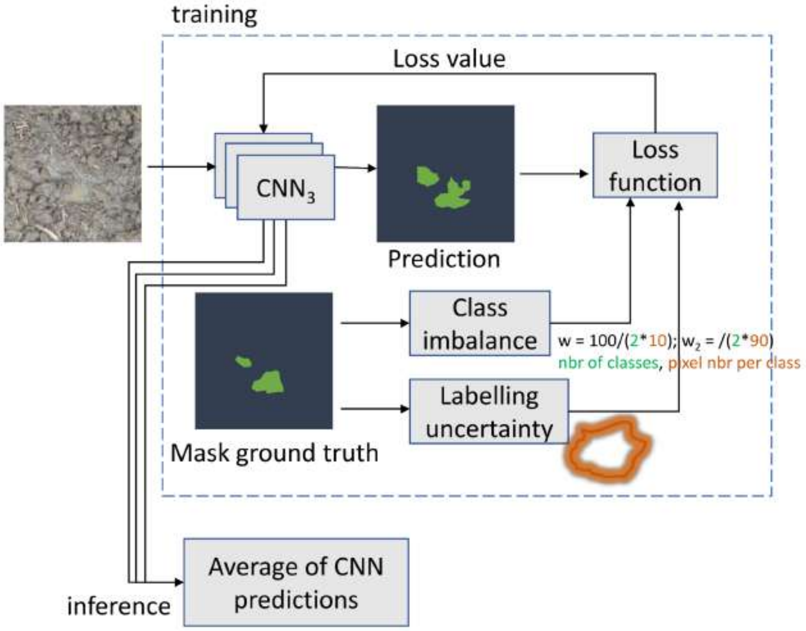


Figure 5: The experimental scheme used in the training of CNNs. The ground-truth mask is used to determine the pixel weights for the CNNs training, considering the class imbalance and the label uncertainty. During inference, we averaged the output of the trained models using the weighted loss function to produce ensemble predictions.

Experimental setup

Training and test procedures were conducted in Google Colab Pro using Keras-Tensorflow. We used categorical cross-entropy as a loss function and Stochastic Gradient Descent as an optimizer. The learning rate was set to 0.001, considering a momentum of 0.9 and decay of 0.005. The batch size was 2. All the models were trained over 100 epochs and we started the training with pre-trained weights from ImageNet.

To evaluate the models, we applied pixel accuracy (ACC) and intersection over union (IoU); two standard metrics for semantic segmentation. The ACC (Equation 8) compares the true positives and true negatives (in other words, the correct and wrongly segmented pixels assigned to a class) and all classified pixels (true positives, true negatives, false positives, and false negatives). ACC is the percentage of pixels correctly assigned to each class. An ACC equal to 1 indicates that all pixels of a class were correctly classified, and a value of 0 indicates that all pixels were wrongly classified. IoU (Equation 9) is the ratio between the intersection and the union of the predicted and the ground-truth area. In other words, if both masks predict and ground-truth match perfectly, the IoU will be equal 1.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \#8$$

$$IoU = \frac{GT \cap P}{GT \cup P} \#9$$

2.4. Water area and its connectivity

The predicted masks, i.e., the number of pixels classified as water, are thus far solely provided in the image space. To quantify the area of water in the plot during the rainfall simulation, the image measurements need to be scaled. Therefore, the predicted masks were orthorectified. GCPs (i.e., marked targets) surrounding the plot were measured with a total station or using a measuring tape. The GCPs were also measured in the images to derive the parameters of a homography to perform a projective transformation eventually. Thus, the image measurements were projected into a plane defined by the GCPs resulting in pixels with known ground sampling distance. Assuming the surface is a plane is a strong simplification as we do not consider the microtopography. However, to generally illustrate the potential of our method it is sufficient. Eventually, each mask is used to compute the total water area in square meters.

The orthorectified masks were further processed to identify connected water pixels. The aim is the measurement of individual water ponds and to assess their changing number throughout the rainfall experiment. The ponds were extracted using the connected component algorithm that enables the identification of the connectivity of pixels and pixel clusters with a structuring element (a kernel). Due to some artifacts in the water masks, i.e., false one-pixel wide border lines that resulted from the need to clip the large input images into smaller patches, it was necessary to close these lines. Therefore a simple morphological operator, i.e., dilatation was applied. Thereby, again a structuring element is used, but this time to probe and expand the water ponds and hence closing such one-pixel wide gaps. Eventually, the identified water ponds were counted in each image. To verify as a pond, at least 50 and 100 connected pixels had to be present at plot 2 and 3 and at plot 1, respectively. These thresholds were set to account for outlier pixels identified as very small water patches. For the time-series assessment the water pixel count and water pond count were smoothed due to some outliers using a median kernel with the size of five.

3. Results and discussion

First, we present the water segmentation results for all test datasets applied to the three networks analyzing the influence class weights and label uncertainty (CW/LU) for the models and ensembles and without considering the spatial correlation. Then, we analyze the performance of the models of each individual plot. Afterwards, we assess the impact of spatial correlations of patches on the model performance and we evaluate the performance of each model, when applied to new and unseen images to assess model transferability. At the end, we compare the water area growth at the different plots, we assess the ponding time estimated by our image approach, and we evaluate the interplay between water pond connectivity, water area growth and discharge. We refer to networks trained using CW and LU as VGG-16 CW/LU, U-Net EfficientNetB0 CW/LU, and U-Net ResNet101 CW/ LU.

3.1. Pond segmentation without considering spatial correlation of samples

For the dataset without considering the spatial correlation among samples, we evaluated the performance of different network configurations and loss functions for pond segmentation. Using CW and LU during training consistently improved performance, indicated by an increase in ACC and IoU (Table 2, Figure 6). Figure 7 shows examples of segmented images from each network and the ensemble models.

Table 2: Impact of CW and LU and ensemble models on pixel accuracy (ACC), Intersection over Union (IoU) for the dataset without considering the spatial correlation of samples.

ID	Model	CW/ LU	Background		Water	
			ACC	IoU	ACC	IoU
Model 1	VGG-16	True	0.959	0.947	0.795	**0.466**
Model 2	VGG-16	False	0.989	0.963	0.535	0.453
Model 3	U-Net EfficientNetB0	True	0.916	0.916	**0.872**	0.376
Model 4	U-Net EfficientNetB	False	0.988	0.960	0.502	0.418
Model 5	U-Net ResNet 101	True	0.954	0.939	0.736	0.412
Model 6	U-Net ResNet 101	False	0.971	0.943	0.499	0.332
Ensemble 1	Average model 1 and 5	True	0.959	0.947	0.778	0.458
Ensemble 2	Average models 1, 3 and 5	True	0.964	0.954	0.822	**0.509**
Ensemble 3	Average models 1 and 3	True	0.950	0.432	**0.859**	0.465
Ensemble 4	Average models 3 and 5	True	0.961	0.949	0.797	0.477

When considering single models, U-Net EfficientNetB0 with CW and LU achieved the highest ACC value for segmented water (0.872). However, this network also had the lowest background segmentation ACC, indicating that it incorrectly assigns a greater number of background pixels as water leading to an overestimation of the water area. Because ACC only considers the ratio of correctly assigned pixels to the total number of pixels, these incorrectly background pixels do not affect water ACC. VGG-16 with CW and LU revealed the highest water segmentation IoU (0.466).

Generally, we observed a decrease in image segmentation performance when rainfall was captured. The rain seemed to act as noise in these images, making it more challenging for the neural networks to correctly segment images in such cases, especially considering that images with rain were not implemented during the training phase of the models. Thus, in future training also, images with captured rainfall should be included.

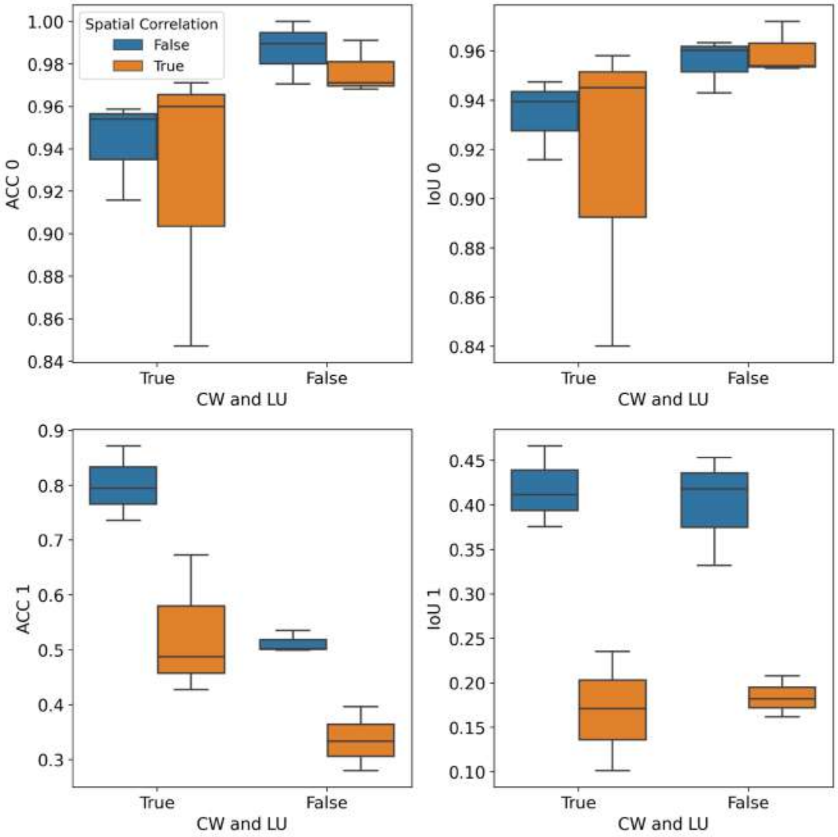


Figure 6: Segmentation performance of all single models for the water area (ACC1, IoU1) and background (ACC0, IoU0) considering CW and LU and sample spatial correlation (True orange, False blue). Note, the y-axis range is different for the different plots to improve visibility of differences.

Influence of CW and LU

When considering CW and LU, the models showed a higher increase of the ACC for the water class (median ACC of 0.794 ± 0.05 and IoU of 0.411 ± 0.04; Figure 6) and a lower segmentation performance for the background. If CW and LU are used, the importance of the dominant class, background, is lowered in the loss equation, shifting it to the less dominant class (i.e., water). Thus, the weights of the network were adjusted, focusing on the ACC of the water class. As our goal was to correctly segment water, the decrease in the performance of background segmentation was neglected. It is evident that models trained with CW produced a better correlation between the predicted number of water pixels and the ground-truth data for the test dataset (Figure 7). When comparing the same model with and without CW and LU, the use of these two parameters further increased the correlation between the predicted number of water pixels and the ground-truth data.

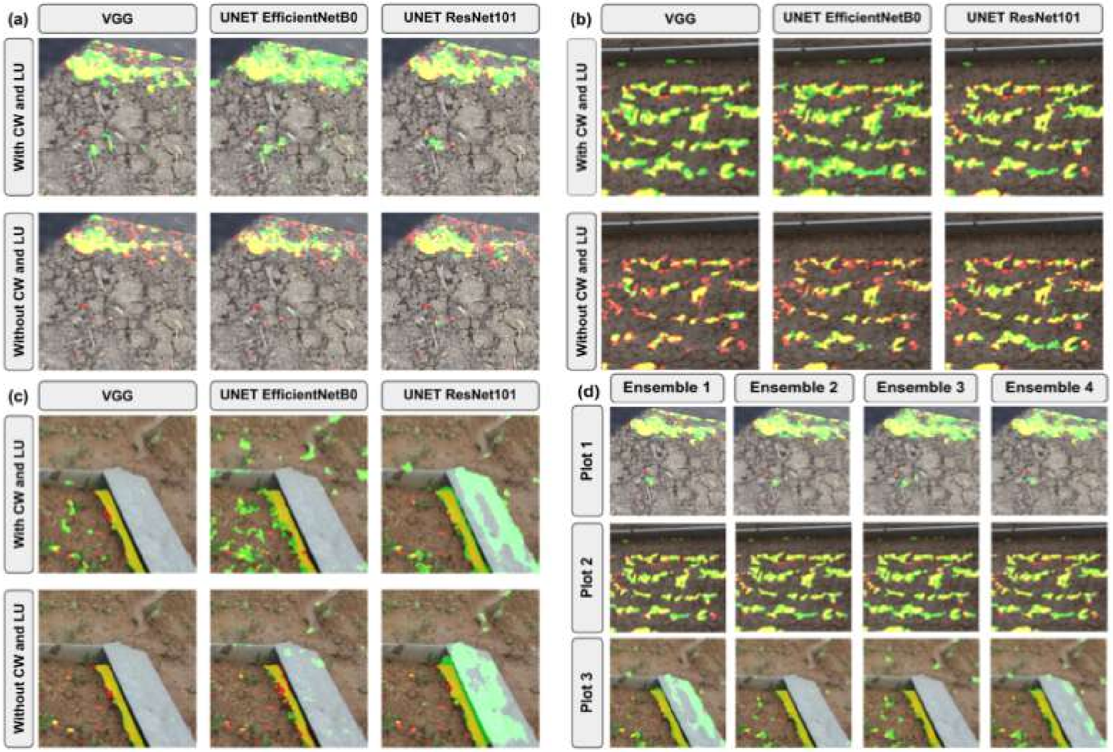


Figure 7: Examples of segmentation results for plot 1 (a), plot 2 (b), and plot 3 (c) with and without CW and LU and considering ensembles of the three models with CW/LU (d) for the dataset without considering spatial correlation between patches. Ground-truth data is shown in red, while green represents pixels wrongly classified as water and yellow represents pixels correctly classified as water.

The enhanced performance of CW/LU models underscores the importance of addressing class imbalance and labeling uncertainty during training. Using Equation 5 we can assess the weight of each class. Our training set had a weight of 0.528 for the background and 9.184 for the water class, while the validation and test sets had weights of 0.527 and 0.529 for the background and 9.670 and 9.037 for the water class, respectively, indicating a significant imbalance between the two classes. Bressan et al. (2022) also found that considering pixel weights improved classification performance for imbalanced datasets in vegetation mapping tasks using different CNN models, demonstrating the effectiveness of this approach across a range of network architectures and segmentation tasks. The increased agreement between the models that consider the pixel weights indicates that, for the task of soil surface water segmentation, it is more important to consider the class imbalance and the uncertainty of the annotations than the choice of the network architecture.

Considering ensembles

Ensembles have been used in different computer vision tasks (Qummar et al., 2019; Thambawita et al., 2021), presenting better performance than individual models. In our ensembles, we averaged the output of the models, which were trained considering CW and LU. We created four different ensemble models, each compromised a different combination of the single models. Generally, the ensemble models revealed the highest values of ACC and IoU (Table 2). Ensemble 2, considering all architectures, achieved the highest water class IoU (0.509), and ensemble 3, considering VGG-16 CW/LU and U-Net EfficientNetB0 CW/LU, was the second best model in terms of water class ACC (0.859). For ensemble 2, water IoU showed an increase in performance from 4.27% up to 13.35% and for ensemble 3 the difference in performance considering water ACC ranged from -1.26% to 12.34% when compared to the models trained with CW/LU. Ensembles presented a higher median value for water class ACC and IoU (0.809 ± 0.03 and 0.471 ± 0.02, respectively) when compared to single models (0.794 ± 0.05 and 0.411 ± 0.04, respectively). Furthermore, ensembles presented a lower standard deviation for both classes and both metrics, highlighting an increase in the robustness of the predictions (Figure 8).

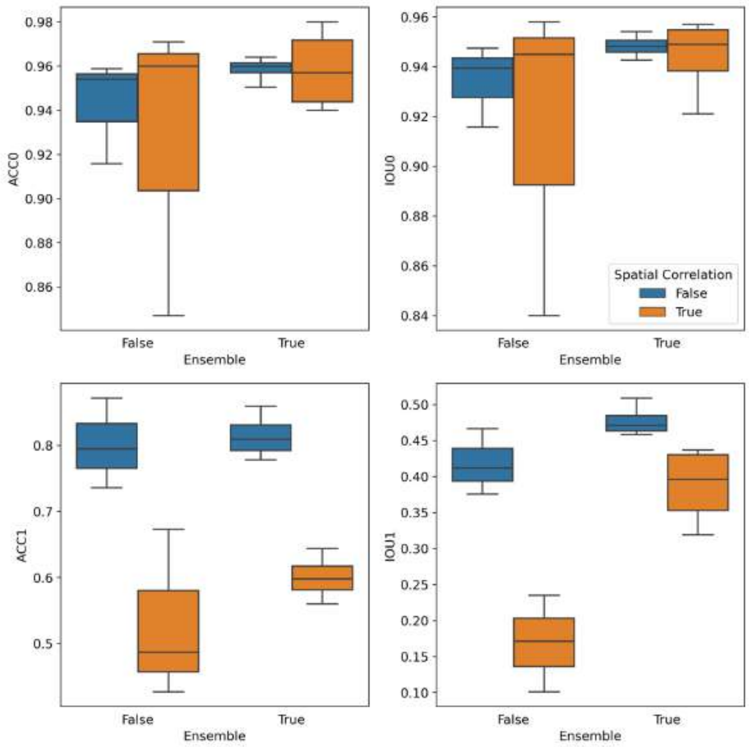


Figure 8:Segmentation performance water area (ACC1, IoU1) and background (ACC0, IoU0) considering ensemble models or not (columns True and False, respectively) and spatial correlation of samples (True orange and False blue).

Looking at the individual plots

At the individual plot scale the improved performance, when considering CW and LU, becomes obvious again (Figure A1). Water class ACC increased from 0.575 ± 0.027 to 0.821 ± 0.032, 0.452 ± 0.030 to 0.755 ± 0.036, and 0.504 ± 0.027 to 0.798 ± 0.068 for Plot 1, 2 and 3, respectively. Thus, average performance increased, although the standard deviation increased slightly. Generally, there was no obvious difference in performance in all individual models.

Figure 9 presents the results of the ensemble models for each plot. Again, ensembles are showing superior average performance and a decrease in standard deviation. Furthermore, we observed a better performance of plot 1 compared to the other two plots, even though in plot 1 we did not use more images (110 for plot 1 versus 69 and 125 images for plots 2 and 3 respectively) or a greater number of water pixels (3.581.510 for plot 1 versus 4.272.067 and 9.381.495 pixels for plot 2 and 3 respectively) during the training. The difference in performance might be caused by the camera position. As can be seen in Figure 2, for plot 1, the camera was positioned such that it had nearly a nadir view of the region of interest. In contrast, in plots 2 and 3 images were captured with slanted views, leading to more concealed regions behind soil aggregates.

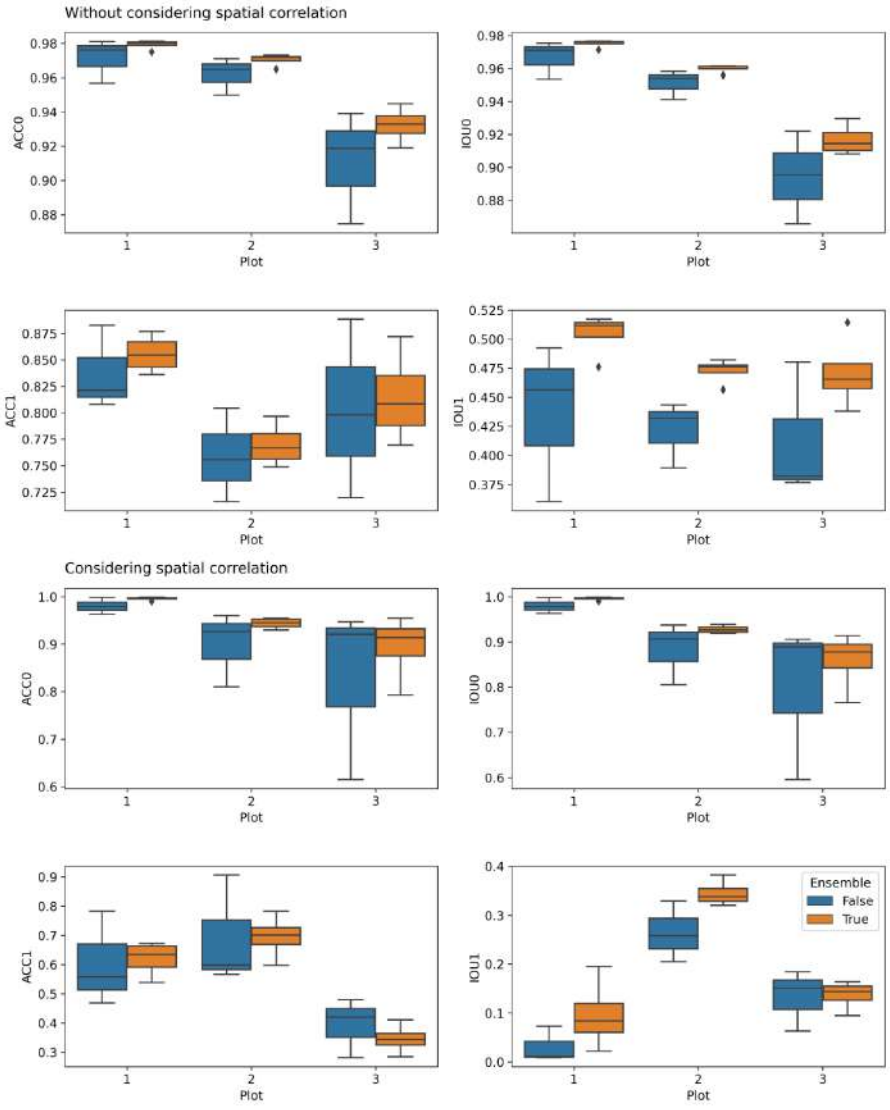


Figure 9: Segmentation performance of all study plots for the water and background accuracy (ACC0 and ACC1, respectively) and background and water IoU (IoU0 and IoU1, respectively) considering ensemble models and the spatial correlation between samples.

3.2. Pond segmentation considering spatial correlation of samples

Table 3 presents the performance of models and ensemble trained considering the spatial correlation between samples, with and without the use CW and LU, and the difference in performance compared to their contra parts. In general, similar to models trained not considering the spatial correlation, by using CW and LU, models trained considering the spatial correlation achieved higher values of water ACC and IoU. Consistent gains in performance were observed for water ACC by applying the pixel weights for single models with increases in performance of 0.208, 0.277, and 0.094 for, respectively, VGG-16, U-Net EfficientNetB0, and U-Net ResNet101. For water IoU, the use of pixel weight did not produce a significant impact on performance, and for U-Net EfficientNetB0 a slight decrease in performance. U-Net EfficienetNet reached the best value of water ACC, although the lowest water IoU, and VGG-16 the best water IoU. Using CW and LU also increases the performance for the background class for both metrics. By using ensembles, we achieved better performance than single models. Except for U-Net EfficientNetB0, ensembles achieved higher water ACC, with emphasis on ensembles 2 and 3. Likewise, ensembles reached the highest values of water IoU, especially Ensembles 1 and 2.

Table 3: Performance and comparison of models trained considering the spatial correlation among samples. In bold, the difference between the performance of models trained considering not the spatial correlation.

Model	CW/LU	Background		Water	
		ACC	IoU	ACC	IoU
VGG-16	True	0.971 **-0.012**	0.958 **+0.011**	0.487 **-0.308**	0.235 **-0.231**
VGG-16	False	0.991 **-0.002**	0.972 **+0.009**	0.279 **-0.256**	0.208 **-0.245**
U-Net EfficientNetB0	True	0.847 **-0.069**	0.840 **-0.076**	0.673 **-0.198**	0.101 **-0.274**
U-Net EfficientNetB	False	0.968 **-0.032**	0.953 **-0.007**	0.396 **-0.106**	0.182 **-0.237**
U-Net ResNet 101	True	0.960 **-0.006**	0.945 **+0.006**	0.427 **-0.309**	0.171 **-0.241**
U-Net ResNet 101	False	0.971 **0.000**	0.954 **+0.011**	0.333 **-0.166**	0.162 **-0.170**
Ensemble 1	True	0.980 **+0.021**	0.957 **+0.011**	0.588 **-0.191**	0.437 **-0.021**
Ensemble 2	True	0.975 **-0.019**	0.954 **0.000**	0.608 **-0.214**	0.428 **-0.081**
Ensemble 3	True	0.940 **-0.010**	0.921 **-0.022**	0.644 **-0.215**	0.319 **-0.146**
Ensemble 4	True	0.969 +0.008	0.944 **-0.005**	0.560 **-0.237**	0.364 **-0.112**

Figures 10 and A2 show the performance of models trained considering and not considering the spatial correlation among samples. Models trained considering spatial correlation achieved a lower performance compared to the contra parts in all cases. We observed an average difference in the performance of 22% and almost 18% for water ACC and water IoU, respectively, when comparing the two different training strategies. Regarding the background class, models achieved similar performance with margin differences (positives and negatives) between training strategies. Our data shows a lower difference in performance for models trained without CW and LU, with an average difference in the performance of almost 18% and 22% for water ACC and IoU, respectively, compared to 27% and 25% for models trained with CW and LU. The gap in performance, considering models trained with

CW and LU, for water ACC ranged from 20% for U-Net EfficientNetB0 to almost 30% for VGG-16. On the other hand, for water IoU, the decrease in performance ranged from 23% for VGG-16 to 27% for U-Net EfficientNetB0. For ensembles, the average difference in performance was 21% for water ACC and 9% for water IoU. Regarding water ACC, the difference in performance was between 19% for Ensemble 1 and 24% for Ensemble 4. For water IoU, the gap ranged from 2% for Ensemble 1 to 15% for Ensemble 3.

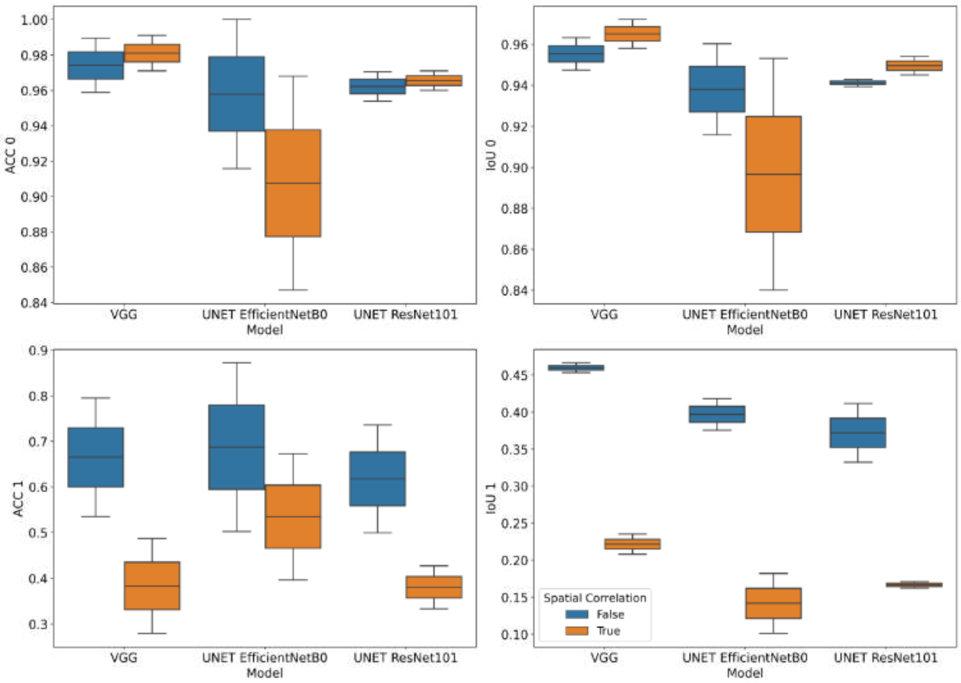


Figure 10: Segmentation performance of single models trained with and without CW and LU considering and not the spatial correlation among samples.

Our findings suggest over-optimistic models when training procedures are conducted without considering the spatial correlation, confirming previous studies (Kattenborn et al., 2022; Meyer & Pebesma, 2022; Schratz et al., 2019). Even though models trained not considering spatial correlation were "over-fitted" for the plots used to train, this does not make the results invalid. Predictions produced by CNNs trained and tested with spatially correlated data are not invalid and the predictions can be unbiased for balanced input data (Kattenborn et al., 2022).

3.3. Model performance on unseen plots

Figures A3 present the performance of Ensemble 2 for plots 4 and 5. It is worth emphasizing that plots 4 and 5 were not used during training and testing procedures, being used only to assess model robustness and transferability to unseen areas. Our results do not show a specific pattern regarding the best strategy to be used to train the models, being models trained not considering the spatial correlation more transferable to plot 4, and models considering the spatial correlation more transferable to plot 5. This fact could be related to dataset size and difference and image perspective.

For plot 4, models trained using non spatial correlated patches and considering CW and LU showed an overall best performance for water accuracy. In terms of water IoU, models trained with spatially correlated patches and that consider CW and LU have a slightly better performance. U-Net EfficientNetB0 and Ensemble 3, trained considering CW and LU, reached the highest performance, with similar values for models trained considering and not the spatial correlation between samples (being the models trained not considering the spatial correlation 5% and almost 2% better, respectively). Nevertheless, for these two cases, models trained considering the spatial correlation reached better performance in terms of water IoU, with an increase of 9% and 13%, respectively. Ensemble 3 and VGG-16 with CW and LU, both models considering the spatial correlation, reached the highest values of water IoU (35% and 32%, respectively).

For plot 5, models trained considering spatial correlation among samples showed a better performance in terms of water ACC and water IoU for most cases. U-Net EfficientNetB0 and Ensemble 3 trained using CW and LU present the highest values of water ACC. Although, for plot 5, by considering the spatial correlation, we increase the performance of U-Net EfficientNetB0 and Ensemble 3 by almost 4%. In terms of water IoU, Ensemble 1 and VGG-16 without CW and LU, both considering the spatial correlation, reached the best values (17% and 16%, respectively).

3.4. Interaction between discharge, water pixel area and connectivity

This section demonstrates how the segmented and orthorectified water area from ensemble 2 can be used to improve the hydrological interpretation of experimental results. We compare the water area identified by the CNN to the number of connected water segments and the measured runoff for each plot (Figures 12, 13 and 14). It is important to highlight that we present a rough estimate of the water coverage area. To provide a more accurate estimation, it is necessary to consider microtopography. Furthermore, the pixel measurements were scaled, projecting the images into a plane defined by the GCPs, which were not exactly at ground level, leading to some overestimation of the pixel size.

CNN-based image segmentation also enables the observation of the spatiotemporal evolution of ponding areas. Thus, it can be used to track and understand the onset of hydrologic connectivity, such as the interconnection of isolated ponding areas which is illustrated in Figure 11. The observation of the connectivity among the ponding areas can help to estimate when surface runoff occurs and where it is likely to form. It should be noted that the method is perfectly scalable and therefore of potential use in a variety of hydrologic applications.

Segmenting small flow channels that emerge once the depressions overflow is challenging due to their limited water volume. However, the spatial distribution of the depression can indicate where the flow channels are likely to occur (between the ponding areas in a downhill direction).

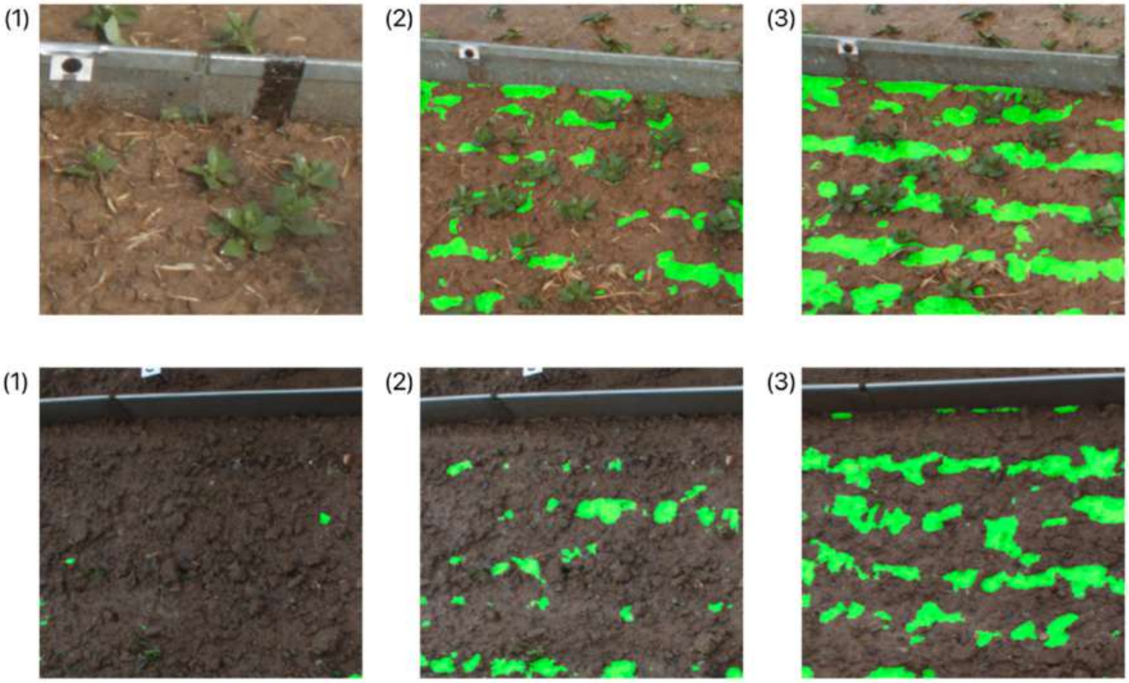


Figure 11: Examples that highlight the growth of water ponds, measured by CNN based segmentations (green patches) at the soil surface during different points in time at plot 2 (lower row) and plot 3 (upper row).

Figure 12 shows the segmented water area and the measured discharge for plot 2. We highlighted different stages of water ponding and runoff formation. In the initial stage of the rainfall simulation, the infiltration capacity is higher than the rainfall intensity, and all water infiltrates.

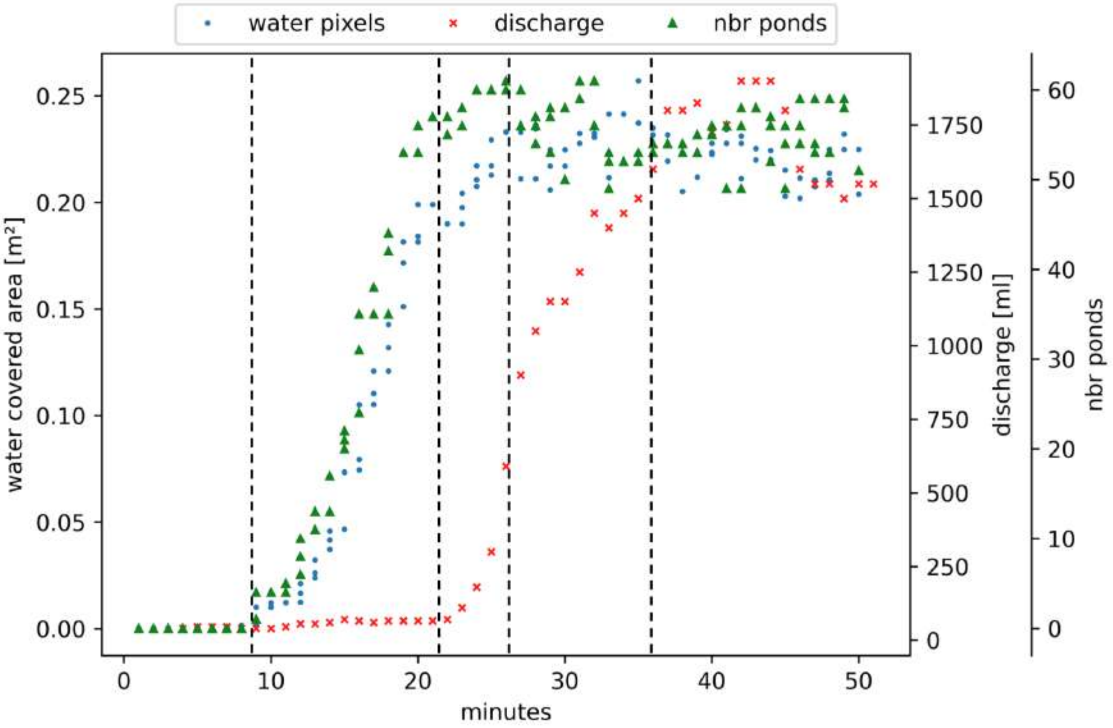


Figure 12: Water coverage area and measured discharge for plot 2.

In the second stage i.e., after about 10 minutes, the rainfall intensity exceeds the infiltration capacity at certain areas of the soil and the excess water starts to accumulate in surface depressions leading to an increase in the water coverage area. However, no discharge is measured yet, and the time gap between water ponding and runoff can be observed. The ongoing rainfall causes more soil areas to reach their maximum infiltration capacity, leading to the formation of new ponding areas as well as the spatial expansion of already existing ones, which in turn leads to an increase in the water coverage area. Once a depression reaches its storage capacity, water spills over and flows downhill, generating flow channels, and thus initiating the first transport of soil particles (Yang & Chu, 2013). The water is then captured and collected in the next downstream depression that has not yet reached its storage capacity. This process leads to the interconnection of different ponding areas, which is highlighted by a decrease of the growth of the number of water ponds. During this phase a strong settling of the freshly tilled soil is visible in the orthophoto time series (see supplementary material), which causes changes in microtopography and the formation of water flow paths on the soil surface.

In stage three i.e. after about 25 minutes, an increase in discharge occurs. This is caused by ponding areas, which have developed a connection with the outlet of the plot. The hydrological connectivity increased and therefore also the drainage of the water to the outlet. The water area growth decelerates and no new water ponds develop. A further extension of water covered area seems hindered by the comparable rough soil surface with larger aggregates remaining from soil tillage.

In stage four the water coverage area has reached a steady state. This happens when the storage capacity of each depression is exceeded, and maximum connectivity among the ponding areas is reached. Thus, the number of water ponds starts to slowly decrease. At this stage, the maximum spatial extent of water is reached and the whole plot contributes to the discharge. Furthermore, after about 35 to 40 minutes also the water area decreases, which coincides with another increase in discharge, potentially due to the emptying of water ponds. In the final (5th) stage, the discharge reaches a steady state.

The assessment of water ponding demonstrates that the microrelief delays the onset of runoff as excess water is first captured in depressions rather than immediately running off the soil surface. Water that is temporarily stored in depressions has more time to infiltrate, resulting in an increased cumulative infiltration. The discharge starts shortly before the water coverage area reaches a steady state suggesting that the CNN-based image segmentation can allow a rough estimation for the runoff onset.

At plot 3 we can also observe a similar behavior of water area and number of ponds growth as well as of discharge (Figure 13). However, the timing and intensities are different. Again, a time delay between the start of water ponding and the first discharge can be seen. The water coverage quickly increases after about 15 minutes. Already after about 20 minutes the growth of the number of water ponds starts to stall, while the water area continues to grow strongly, so individual ponds are extending their area rather than new ponds are being formed. The water area reaches a constant value around the time when discharge accelerates and the number of water ponds starts to decrease due to a higher connectivity between ponds, i.e., after about 50 minutes.

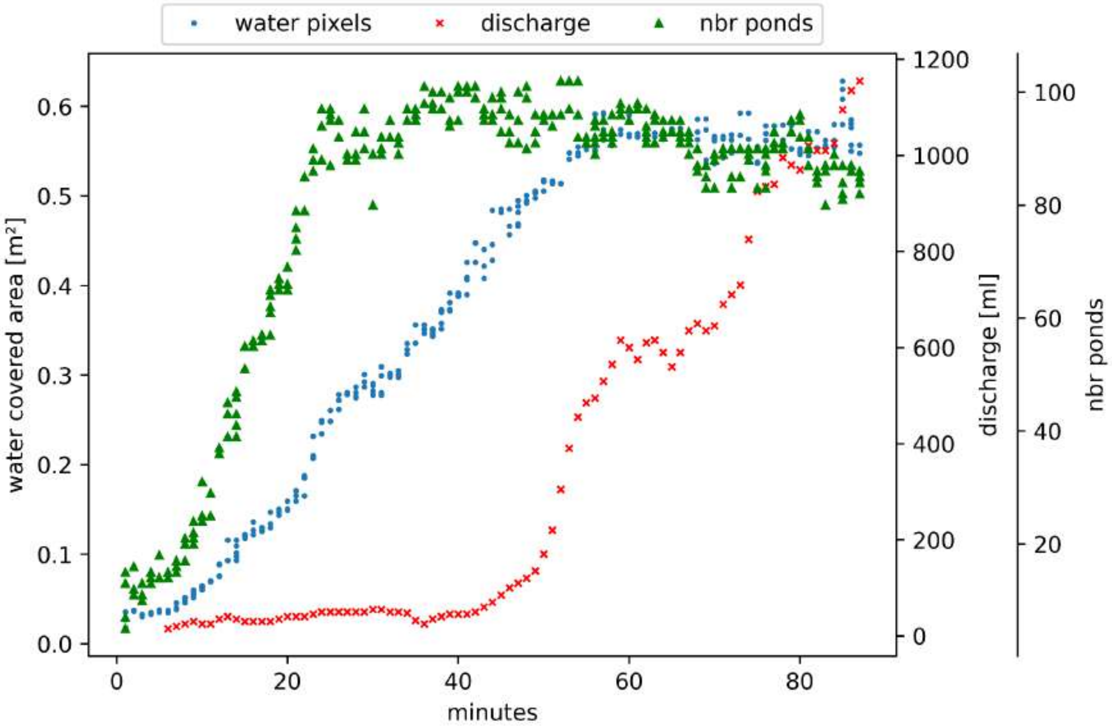


Figure 13: In blue, the water coverage area at plot 3 is predicted using ensemble two and time-lapse images, and in red the measured discharge at the outlet of the rainfall simulations. Green markers highlight the number of water ponds.

The water ponds grow a lot slower at plot 3 when compared to plot 2 (0.1 to 0.15 m² in about 20 minutes versus 0.1 to about 0.15 m² in about 10 minutes), although the rainfall intensity was higher in plot 3. The initial soil moisture was similar in both plots. The different speed of water growth indicates a different change in infiltration rates, i.e. a faster decrease in plot 2. However, a total higher water coverage is observed at plot 3 (~0.2 m² versus ~0.7 m²). In contrast to plot 2, with freshly tilled soil, the growing bean plants on plot 3 indicate that a larger time span has passed since the last soil tillage. Consequently, the surface shows less large aggregates, as these are already destructed by previous rainfall impact and the available water storage in depressions is lower than on plot 2. This suggests also a higher initial connectivity provided by the micro-relief of the plot. Due to the smoother surface, the initiating ponds can spread wider, so the number of ponds does not increase while the water covered area still grows (minutes 20 to 45).

At plot 1, different behavior in regard to water ponding was observed (Figure 14). Very early on (after about 10 minutes), a strong increase in discharge was measured, while water ponding did not accelerate until about 20 minutes after rainfall started. In general, water on the surface did not grow as strongly when compared to the other two plots, i.e. water covered less than 0.05 m² after 25 minutes, whereas after the same duration at plots 2 and 3 water coverage reached at least 0.15 m². This might indicate a larger infiltration rate at the beginning of plot 1 compared to plots 2 and 3. Although similar rainfall intensities were given in plots 1 and 2, the different temporal behavior of water ponding became obvious and might be attributed to heterogeneities, e.g. in soil moisture content, across the plot area at the beginning of the rainfall. Within the orthophoto time series (see supplementary material) it can be seen that a single pond is forming near the plot outlet, which causes the increase in runoff. The rest of the plot shows a later - and nearly - homogeneous formation of ponds. Their number increases with the increase of water area and, in contrast to plot 2 and 3, no stagnation or even decrease during the first phase of rainfall becomes obvious.

After the hour break and second start of rainfall, the number of water ponds increases strongly over a very short interval, as does the water area. However, afterwards the number of water ponds very strongly decreases again, occurring around the same time, i.e. after about 125 minutes, when discharge increases strongly. This corresponds to the processes observed in the other two plots, i.e. some decrease of water ponds number after connectivity increased. The delayed onset of the discharge indicates that the heterogeneities seen in the first rainfall phase have dissolved. The decreasing number of water ponds during onset of discharge shows again how hydrological connectivity emerges between ponds, as in plots 2 and 3.

Roughly between the 60th and 120th minutes, during the rainfall simulation break, an increase in the water pixel area and the number of ponds can be observed. However, no actual changes happened at the soil surface. The error in the measurement is related to changes in the lighting conditions leading to a misclassification of water pixels. During this period the water pixel area slightly increased, whereas the number of ponds increased strongly, which indicates a large number of small ponds that can be considered as outliers.

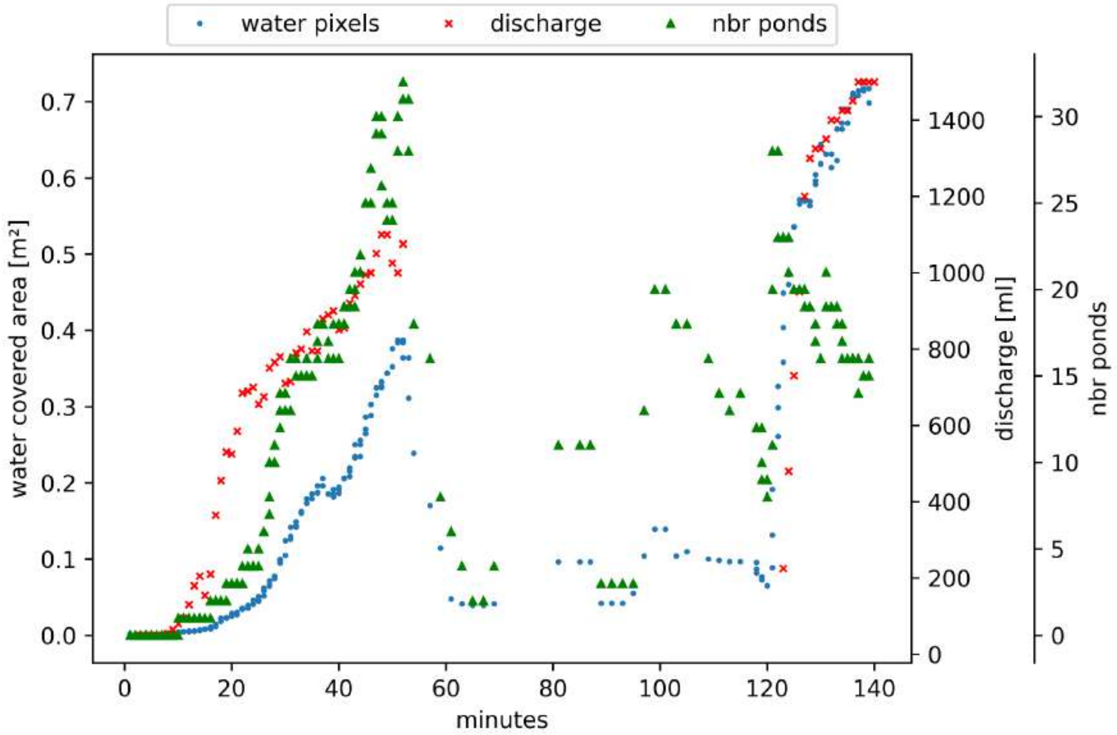


Figure 14: In blue, the water coverage area at plot 1 is predicted using ensemble two and time-lapse images, and in red the measured discharge at the outlet of the rainfall simulations. Green markers highlight the number of water ponds.

Our work shows the potential of using CNNs to map soil surface water coverage to provide a direct and visual quantitative assessment of ponding. Water retention on the soil surface, estimated with high temporal and spatial resolution from images, can be used to approximate soil infiltration and the hydrological connectivity for a given runoff event, representing a step forward in understanding runoff formation at the plot scale. Our measurements highlight the importance of the micro-topography of the soil surface and the connectivity among the ponds, as it controls the generation of the runoff and its temporal variation. Limitations to our methods remain in the detection of small connecting channels, which might be covered by large aggregates (e.g. in plot 2) or by grown plants (e.g. plot 3) from the view of the camera. On the other hand, our proposed method can easily be upscaled. It is not limited to the plot scale and is potentially applicable at higher spatial resolutions such as the field scale.

In the future, we intend to increase our dataset with images captured from different perspectives from the plot, i.e. a multi-camera setup, to achieve a more robust water segmentation and connectivity assessment, considering redundant information. Furthermore, further deep learning models should be tested because there might be even better models available in the model zoo for the challenging task of segmenting shallow water on soil surfaces. Another future task should be the combination of the predicted water mask with a model of the surface microtopography to allow for actual water volume estimations and potentially measure spatially variable infiltration rates if pond catchment areas are taken into account.

4. Conclusions

We propose an image-based approach that employs CNNs and photogrammetric techniques to segment water in rainfall simulation plots to estimate the area of water coverage and to analyze ponding. Images from five different erosion plots were captured during rainfall simulations, along with the discharge and GCPs for image rectification. Our findings suggest that accounting for class imbalance and label uncertainty during network training leads to significantly improved performance. Our results indicate that, for the task of soil surface water segmentation, considering the pixel weights is more important than the model architecture to reach satisfactory performance. Furthermore, ensemble models lead to better results compared to single models. Our results further suggest that models trained considering the spatial correlation among samples can be slightly more transferable to unseen sites. However, the application of these models revealed the lowest performance when compared to the inference performed to the images of the plots for which the models were trained and tested. Thus, more training data might be needed.

The direct comparison of the measured discharge and the development of the ponding areas, i.e., measured by the number of water area pixels and the number of ponds, revealed the importance of ponding time and connectivity assessment to better understand the runoff formation. We could observe different behavior regarding the timing and intensity of ponding and discharge at all three plots. For instance, at one plot the interplay between water ponds connectivity, water coverage and discharge highlighted that due to an increase of hydrological connectivity, water coverage stagnated and discharge increased.

Visually observing and automatic quantification of water pond formation and development is a new frontier and a step forward in understanding runoff generation, providing a new and detailed data source. To the best of our knowledge, our approach is the first to allow the direct quantification of the spatial-temporal development of the water ponding.

5. Acknowledgments

We would like to thank Lea Epple, Anne Bienert, and Oliver Grothum for their support in the fieldwork. We are thankful for the soil data analysis by the TU Bergakademie Freiberg. The project was founded by the DFG (405774238), CNPq (p: 405997/2021-3, 308481/2022-4 and 305296/2022-1), CAPES Print (p: 88881.311850/2018-01) and Fundect (p: 71/009.436/2022).

6. Appendix

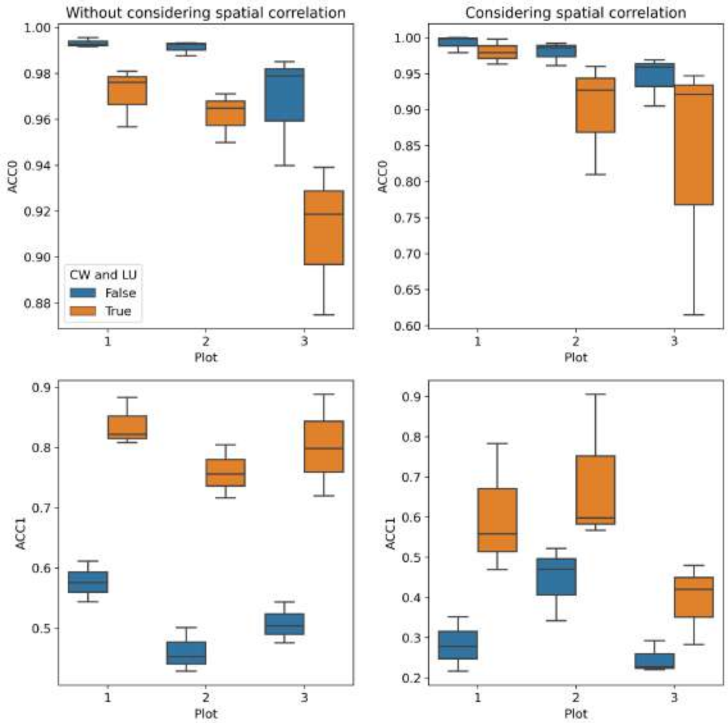


Figure A1: Model performance, background and water accuracy (ACC0 and ACC1, respectively) and IoU (IoU0 and IoU1, respectively) for individual plots considerindering and not patch spatial correlation.

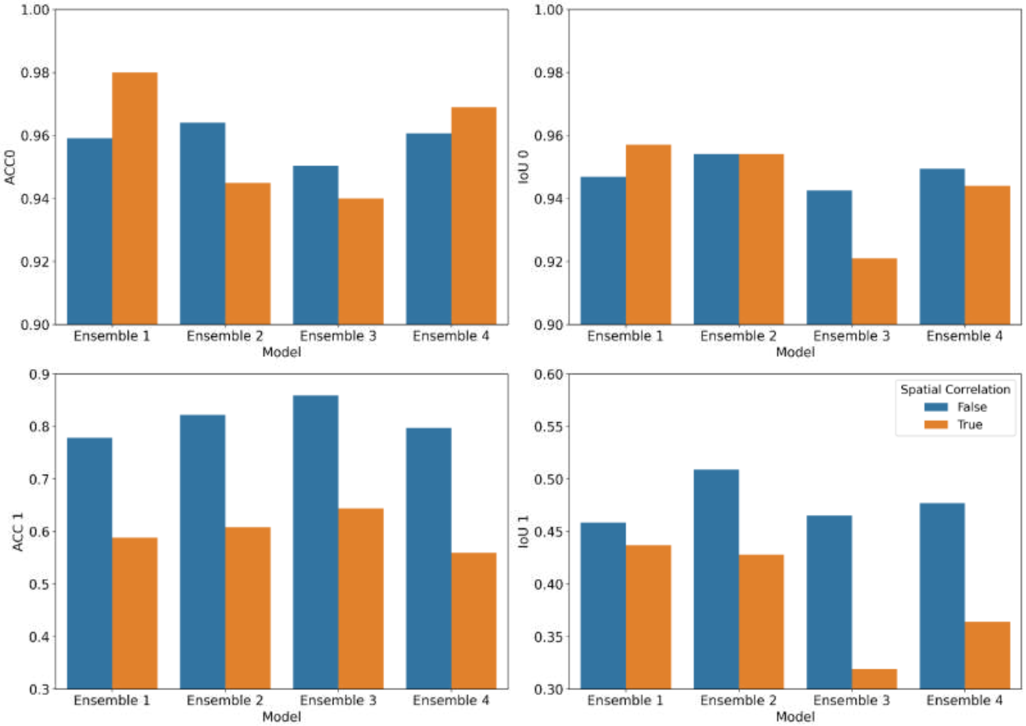


Figure A2: Segmentation performance of ensemble models trained considering and not the sample spatial correlation.

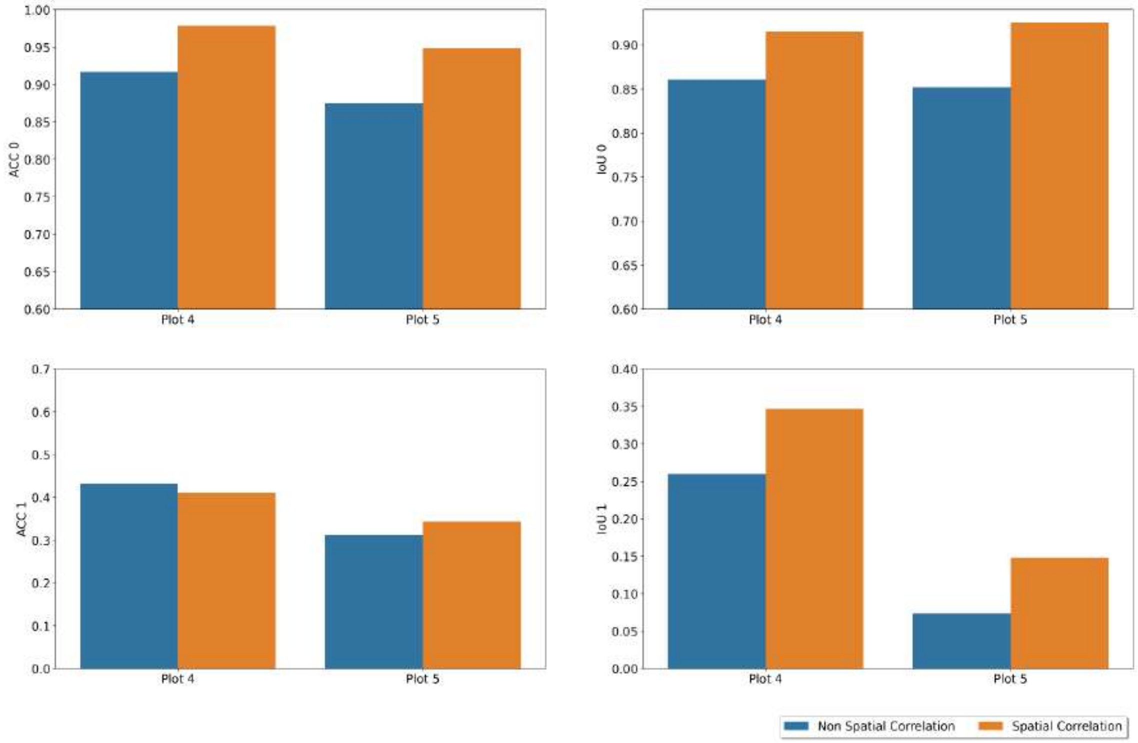


Figure A3: Model performance considering models trained with CW/LU and with and without considering spatial correlation for Ensemble 2 for plots 4 and 5.

7. References

Ahuja, L. R. (1983). Modeling Infiltration into Crusted Soils by the Green—Ampt Approach. Soil Science Society of America Journal, 47(3), 412–418. https://doi.org/10.2136/sssaj1983.03615995004700030004x

Ali, G.A., Roy, A.G. (2009). Revisiting hydrologic sampling strategies for an accurate assessment of hydrologic connectivity in humid temperate systems. Geography Compass, 3 (1), 350-374, 10.1111/j.1749-8198.2008.00180.x

Anache, J. A. A., Wendland, E. C., Oliveira, P. T. S., Flanagan, D. C., & Nearing, M. A. (2017). Runoff and soil erosion plot-scale studies under natural rainfall: A meta-analysis of the Brazilian experience. CATENA, 152, 29–39. https://doi.org/10.1016/j.catena.2017.01.003

Assouline, S. (2013). Infiltration into soils: Conceptual approaches and solutions. Water Resources Research, 49, 1755–1772. https://doi.org/10.1002/wrcr.20155

Assouline, S., Selker, J. S., & Parlange, J.-Y. (2007). A simple accurate method to predict time of ponding under variable intensity rainfall. Water Resources Research, 43(3).

https://doi.org/10.1029/2006WR005138

Anache, J. A. A., Wendland, E., Rosalem, L. M. P., Youlton, C., & Oliveira, P. T. S. (2019). Hydrological trade-offs due to different land covers and land uses in the Brazilian Cerrado. Hydrol. Earth Syst. Sci, 23, 1263–1279. https://doi.org/10.5194/hess-23-1263-2019

B. Baheti, S. Innani, S. Gajre and S. Talbar, "Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 1473-1481, doi: 10.1109/CVPRW50498.2020.00187

Baartman, J.E.M., Masselink, R., Keesstra, S.D. and Temme, A.J.A.M., Linking landscape morphological complexity and sediment connectivity. Earth Surf. Process. Landforms, 38, 1457-1471, 2013. https://doi.org/10.1002/esp.3434

Beven, K. (2021). The era of infiltration. Hydrology and Earth System Sciences,25, 851–866. https://doi.org/10.5194/hess-25-851-2021, 2021.

Bracken, L., & Croke, J. (2007). The Concept of Hydrological Connectivity and Its Contribution to Understanding Runoff-Dominated Geomorphic Systems. Hydrological Processes, 21. https://doi.org/10.1002/hyp.6313

Brandt, M., Tucker, C. J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., Chave, J., Rasmussen, L. V., Hiernaux, P., Diouf, A. A., Kergoat, L., Mertz, O., Igel, C., Gieseke, F., Schöning, J., Li, S., Melocik, K., Meyer, J., Sinno, S., … Fensholt, R. (2020). An unexpectedly large count of trees in the West African Sahara and Sahel. Nature, 587(7832), Article 7832. https://doi.org/10.1038/s41586-020-2824-5

Bressan, P. O., Junior, J. M., Correa Martins, J. A., de Melo, M. J., Gonçalves, D. N., Freitas, D. M., Marques Ramos, A. P., Garcia Furuya, M. T., Osco, L. P., de Andrade Silva, J., Luo, Z., Garcia, R. C., Ma, L., Li, J., & Gonçalves, W. N. (2022). Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping. International Journal of Applied Earth Observation and Geoinformation, 108, 102690. https://doi.org/10.1016/j.jag.2022.102690

Campbell, G. S. (1985). Soil Physics with BASIC: Transport Models for Soil-Plant Systems. Elsevier.

Carvalho, M. de A., Marcato, J., Martins, J. A. C., Zamboni, P., Costa, C. S., Siqueira, H. L., Araújo, M. S., Gonçalves, D. N., Furuya, D. E. G., Osco, L. P., Ramos, A. P. M., Li, J., de Castro, A. A., & Gonçalves, W. N. (2022). A deep learning-based mobile application for tree species mapping in RGB images. International Journal of Applied Earth Observation and Geoinformation, 114, 103045. https://doi.org/10.1016/j.jag.2022.103045

Casenave, A., & Valentin, C. (1992). A runoff capability classification system based on surface features criteria in semi-arid areas of West Africa. Journal of Hydrology, 130(1), 231–249. https://doi.org/10.1016/0022-1694(92)90112-9

Dal Pozzolo, A., Caelen, O., & Bontempi, G. (2015). When is Undersampling Effective in Unbalanced Classification Tasks? In A. Appice, P. P. Rodrigues, V. Santos Costa, C. Soares, J. Gama, & A. Jorge (Hrsg.), Machine Learning and Knowledge Discovery in Databases (S. 200–215). Springer International Publishing. https://doi.org/10.1007/978-3-319-23528-8_13

Darboux F, Huang C. (2001). Evolution of soil surface roughness and flowpath connectivity in overland flow experiments. Catena 46: 125– 139.

De Roo, A.P.J, Wesselink, C.G., Ritsema, C. (1996). LISEM: A single-event physically based hydrological and soil erosion model for drainage basins: Theory, input and output. Hydrological Processes 10: 1107–1117.

Dunkerley, D. (2012). Effects of rainfall intensity fluctuations on infiltration and runoff: Rainfall simulation on dryland soils, Fowlers Gap, Australia. Hydrological Processes, 26(15), 2211–2224. https://doi.org/10.1002/hyp.8317

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

Favis-Mortlock, D. (1998). A self-organizing dynamic systems approach to the simulation of rill initiation and development on hillslopes, Computers & Geosciences, 24(4), 353-372. https://doi.org/10.1016/S0098-3004(97)00116-7

Fiener, P., Seibert, S. P., & Auerswald, K. (2011). A compilation and meta-analysis of rainfall simulation data on arable soils. Journal of Hydrology, 409(1), 395–406. https://doi.org/10.1016/j.jhydrol.2011.08.034

Ganot, Y., Holtzman, R., Weisbrod, N., Nitzan, I., Katz, Y., & Kurtzman, D. (2017). Monitoring and modeling infiltration–recharge dynamics of managed aquifer recharge with desalinated seawater. Hydrology and Earth System Sciences, 21(9), 4479–4493.

https://doi.org/10.5194/hess-21-4479-2017

Gebrehiwot, A., Hashemi-Beni, L., Thompson, G., Kordjamshidi, P., & Langan, T. E. (2019). Deep Convolutional Neural Network for Flood Extent Mapping Using Unmanned Aerial Vehicles Data. Sensors, 19(7), Article 7. https://doi.org/10.3390/s19071486

Green, W. H., & Ampt, G. A. (1911). Studies on Soil Phyics. The Journal of Agricultural Science, 4(1), 1–24.

Hänsel, P., Schindewolf, M., Eltner, A., Kaiser, A., Schmidt, J. (2016), Feasibility of High-Resolution Soil Erosion Measurements by Means of Rainfall Simulations and SfM Photogrammetry. Hydrology, 3 (4), 38. https://doi.org/10.3390/hydrology3040038

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition (arXiv:1512.03385). arXiv. https://doi.org/10.48550/arXiv.1512.03385

Heipke C., Rottensteiner F. (2020): Deep learning for geometric and semantic tasks in photogrammetry and remote sensing, Geo-spatial Information Science, 23:1, https://doi.org/10-19. DOI: 10.1080/10095020.2020.1718003

Helalia, A. M., & Letey, J. (1988). Cationic Polymer Effects on Infiltration Rates with a Rainfall Simulator. Soil Science Society of America Journal, 52(1), 247–250. https://doi.org/10.2136/sssaj1988.03615995005200010043x

Higa, L., Marcato Junior, J., Rodrigues, T., Zamboni, P., Silva, R., Almeida, L., Liesenberg, V., Roque, F., Libonati, R., Gonçalves, W. N., & Silva, J. (2022). Active Fire Mapping on Brazilian Pantanal Based on Deep Learning and CBERS 04A Imagery. Remote Sensing, 14, 688. https://doi.org/10.3390/rs14030688

Holden, J., & Burt, T. P. (2002). Infiltration, runoff and sediment production in blanket peat catchments: Implications of field rainfall simulation experiments. Hydrological Processes, 16(13), 2537–2557. https://doi.org/10.1002/hyp.1014

Horton, R. E. (1939). Analysis of runoff-plat experiments with varying infiltration-capacity. Eos, Transactions American Geophysical Union, 20(4), 693–711. https://doi.org/10.1029/TR020i004p00693

Horton, R. E. (1941). An Approach Toward a Physical Interpretation of Infiltration-Capacity. Soil Science Society of America Journal, 5(C), 399–417. https://doi.org/10.2136/

Huang, H., Lin, F., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y. W., Wu, J. (2020). U-Net 3+: A Full-Scale Connected U-Net for Medical Image Segmentation. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1055-1059.https://doi.org/10.1109/ICASSP40776.202

Hu, Kai, Dongsheng Zhang, and Min Xia. 2021. "CDUNet: Cloud Detection UNet for Remote Sensing Imagery" Remote Sensing 13, no. 22: 4533. https://doi.org/10.3390/rs13224533

Ichim, L., & Popescu, D. (2020). Segmentation of Vegetation and Flood from Aerial Images Based on Decision Fusion of Neural Networks. Remote Sensing, 12, 2490. https://doi.org/10.3390/rs12152490

Jetten, V.G., de Roo, A.P.J. (2001). Spatial Analysis of Erosion Conservation Measures with LISEM . In: Harmon, R.S., Doe, W.W. (eds) Landscape Erosion and Evolution Modeling. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-0575-4_14

Jing, M., Cheng, L., Ji, C., Mao, J., Li, N., Duan, Z., Li, Z., & Li, M. (2021). Detecting unknown dams from high-resolution remote sensing images: A deep learning and spatial analysis approach. International Journal of Applied Earth Observation and Geoinformation, 104, 102576. https://doi.org/10.1016/j.jag.2021.102576

Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M. D., & Dormann, C. F. (2022). Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. ISPRS Open Journal of Photogrammetry and Remote Sensing, 5, 100018. https://doi.org/10.1016/j.ophoto.2022.100018

Konapala, G., Kumar, S. V., Ahmad, Exploring, S. K. (2021). Sentinel-1 and Sentinel-2 diversity for flood inundation mapping using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 180, 163-173. https://doi.org/10.1016/j.isprsjprs.2021.08.016

Kostiakov, A. N. (1932). On the dynamics of the coefficient of water-percolation in soils and on the necessity of studying it from a dynamic point of view for purposes of amelioration. Trans. 6th Cong. International. Soil Science, Russian Part A, 17–21.

Le Mesnil, M., Moussa, R., Charlier, J.-B., & Caballero, Y. (2021). Impact of karst areas on runoff generation, lateral flow and interbasin groundwater flow at the storm-event timescale. Hydrology and Earth System Sciences, 25(3), 1259–1282. https://doi.org/10.5194/hess-25-1259-2021

Lewis, M. R. (1937). The rate of infiltration of water in irrigation practice, Trans. Am. Geophys. Union, 18, 361–368.

López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences, 250, 113–141. https://doi.org/10.1016/j.ins.2013.07.007

McNamara, J., Chandler, D., Seyfried, M., & Achet, S. (2005). Soil Moisture States, Lateral Flow, and Streamflow Generation in a Semi-Arid, Snowmelt-Driven Catchment. Hydrological Processes, 19, 4023–4038. https://doi.org/10.1002/hyp.5869

Meyer, H., & Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. Nature Communications, 13(1), Article 1. https://doi.org/10.1038/s41467-022-29838-9

Muñoz, D. F., Muñoz, P., Moftakhari, H., & Moradkhani, H. (2021). From local to regional compound flood mapping with deep learning and data fusion techniques. Science of The Total Environment, 782, 146927. https://doi.org/10.1016/j.scitotenv.2021.146927

Nguyen, T.-A., Kellenberger, B., & Tuia, D. (2022). Mapping forest in the Swiss Alps treeline ecotone with explainable deep learning. Remote Sensing of Environment, 281, 113217. https://doi.org/10.1016/j.rse.2022.113217

Peñuela A, Darboux F, Javaux M, Bielders CL. (2016). Evolution of overland flow connectivity in bare agricultural plots. Earth Surface Dynamics Discussions 41(11): 1595–1613.

Persello, C., Wegner, J. D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., & Camps-Valls, G. (2022). Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. IEEE Geoscience and Remote Sensing Magazine, 10(2), 172-200. https://doi.org/10.1109/MGRS.2021.3136100

Qummar, S., Khan, F., Shah, S., Khan, A., Band, S., Rehman, Z., Khan, I., & Jadoon, W. (2019). A Deep Learning Ensemble Approach for Diabetic Retinopathy Detection. IEEE Access, 7, 1–1. https://doi.org/10.1109/ACCESS.2019.2947484

Ravuri, S., Lenc, K., Willson, M. et al. (2021) Skilful precipitation nowcasting using deep generative models of radar. Nature, 597, 672–677. https://doi.org/10.1038/s41586-021-03854-z

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation (arXiv:1505.04597). arXiv. https://doi.org/10.48550/arXiv.1505.04597

Schindewolf, M., & Schmidt, J. (2012). Parameterization of the EROSION 2D/3D soil erosion model using a small-scale rainfall simulator and upstream runoff simulation. CATENA, 91, 47–55. https://doi.org/10.1016/j.catena.2011.01.007

Schindewolf, M., & Schmidt, W. (2009). Prüfung und Validierung des neu entwickelten Oberflächenabflussmoduls des Modells EROSION 3D im Zusammenhang mit Maßnahmen des vorsorgenden Hochwasserschutzes auf landwirtschaftlich genutzten Flächen. Schriftenreihe des Landesamtes für Umwelt, Landwirtschaft und Geologie ; Heft 15/2009. http://nbn-resolving.de/urn:nbn:de:bsz:14-ds-1244617738468-42743

Schmidt, J. (1996). Entwicklung und Anwendung eines physikalisch begründeten Simulationsmodells für die Erosion geneigter landwirtschaftlicher Nutzflächen. In Herausgeberexemplar (FU Berlin). Selbstverl. des Inst. für Geograph. Wiss., Berlin. https://doi.org/10.23689/fidgeo-3199

Schmidt, J., Werner, M. v, & Michael, A. (1999). Application of the EROSION 3D model to the CATSOP watershed, The Netherlands. CATENA, 37(3), 449–456. https://doi.org/10.1016/S0341-8162(99)00032-6

Schratz, P., Muenchow, J., Iturritxa,E., Richter, K., Brenning, A. (2019). Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecological Modelling, 406, 109-120. https://doi.org/10.1016/j.ecolmodel.2019.06.002.

SCS. (1954). Hydrology guide for use in watershed planning, Soil Conservation Service, USDA, Washington, D.C.

Shamsolmoali, P., Zareapoor, M., Wang, R., Zhou, H., & Yang, J. (2019). A novel deep structure U-Net for sea-land segmentation in remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(9), 3219-3232

Sharma, M. L., Gander, G. A., & Hunt, C. G. (1980). Spatial variability of infiltration in a watershed. Journal of Hydrology, 45(1), 101–122. https://doi.org/10.1016/0022-1694(80)90008-6

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Sone, J. S., Oliveira, P. T. S., Euclides, V. P. B., Montagner, D. B., de Araujo, A. R., Zamboni, P. A. P., Vieira, N. O. M., Carvalho, G. A., & Sobrinho, T. A. (2020). Effects of Nitrogen fertilization and stocking rates on soil erosion and water infiltration in a Brazilian Cerrado farm. Agriculture, Ecosystems & Environment, 304, 107159. https://doi.org/10.1016/j.agee.2020.107159

Su, Y., Zhong, Y., Zhu, Q., & Zhao, J. (2021). Urban scene understanding based on semantic and socioeconomic features: From high-resolution remote sensing imagery to multi-source geographic datasets. ISPRS Journal of Photogrammetry and Remote Sensing, 179, 50–65. https://doi.org/10.1016/j.isprsjprs.2021.07.003

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

Thambawita, V., Hicks, S., Halvorsen, P., & Riegler, M. (2021). DivergentNets: Medical Image Segmentation by Network Ensemble. arXiv preprint arXiv:2107.00283.

Thompson, S. E., Katul, G. G., & Porporato, A. (2010). Role of microtopography in rainfall-runoff partitioning: An analysis using idealized geometry. Water Resources Research, 46(7). https://doi.org/10.1029/2009WR008835

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography, 46(sup1), 234–240. https://doi.org/10.2307/143141

Touma, J., & Albergel, J. (1992). Determining soil hydrologic properties from rain simulator or double ring infiltrometer experiments: A comparison. Journal of Hydrology, 135(1), 73–86. https://doi.org/10.1016/0022-1694(92)90081-6

Ümit Atila, Murat Uçar, Kemal Akyol, Emine Uçar, Plant leaf disease classification using EfficientNet deep learning model, Ecological Informatics, Volume 61, 2021, 101182, ISSN 1574-9541, https://doi.org/10.1016/j.ecoinf.2020.101182

Vandaele, R., Dance, S. L., & Ojha, V. (2021). Deep learning for automated river-level monitoring through river-camera images: An approach based on water segmentation and transfer learning. Hydrology and Earth System Sciences, 25(8), 4435–4453. https://doi.org/10.5194/hess-25-4435-2021

van Genuchten, M. Th. (1980). A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Science Society of America Journal, 44(5), 892–898.

https://doi.org/10.2136/sssaj1980.03615995004400050002x

Vlček, L., Šípek, V., Zelíková, N., Čáp, P., Kincl, D., & Vopravil, J. (2022). Water retention and infiltration affected by conventional and conservational tillage on a maize plot; rainfall simulator and infiltrometer comparison study. Agricultural Water Management, 271, 107800. https://doi.org/10.1016/j.agwat.2022.107800

von Werner, M. & Schmidt, J. (1996). Band III: EROSION-3D Modellgrundlagen - Bedienungsanleitung, in: EROSION 2d/3d Ein Computermodell Zur Simulation Der Bodenerosion Durch Wasser. Sächsische Landesanstalt für Landwirtschaft, Sächsisches Landesamt für Umwelt und Geologie, Dresden, Freiberg.

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/10.1016/j.jag.2023.103305)

Wang S, Strauss P, Yao A, Wang X, Yuan Y. (2018). Assessing hydrological connectivity development by using a photogrammetric technique with relative surface connection function (RSCf) in a plot-scale experiment. Journal of Soil and Water Conservation 73(5): 518– 532.

Wilson, T., Baker, B., Meyers, T., Kochendorfer, J., Hall, M., Bell, J., Diamond, H., & Palecki, M. (2016). Site-Specific Soil Properties of the US Climate Reference Network Soil Moisture. Vadose Zone Journal, 15. https://doi.org/10.2136/vzj2016.05.0047

Wu, Y., Zhang, Y., Dai, L., Xie, L., Zhao, S., Liu, Y., & Zhang, Z. (2021). Hydrological connectivity improves soil nutrients and root architecture at the soil profile scale in a wetland ecosystem. Science of The Total Environment, 762, 143162. https://doi.org/10.1016/j.scitotenv.2020.143162

Yang, W.-Y., Li, D., Sun, T., & Ni, G.-H. (2015). Saturation-excess and infiltration-excess runoff on green roofs. Ecological Engineering, 74, 327–336. https://doi.org/10.1016/j.ecoleng.2014.10.023

Ye, W., Lao, J., Liu, Y., Chang, C.C., Zhang, Z., Li, H., Zhou, H. (2022). Pine pest detection using remote sensing satellite images combined with a multi-scale attention-UNet model. Ecological Informatics, 72. https://doi.org/10.1016/j.ecoinf.2022.101906

Zamboni, P. A. P., Vieira, N. O. M., Carvalho, G. A., & Sobrinho, T. A. (2020). Effects of Nitrogen fertilization and stocking rates on soil erosion and water infiltration in a Brazilian Cerrado farm. Agriculture, Ecosystems & Environment, 304, 107159. https://doi.org/10.1016/j.agee.2020.107159

Zhao, L., Hou, R., Wu, F., & Keesstra, S. (2018). Effect of soil surface roughness on infiltration water, ponding and runoff on tilled soils under rainfall simulation experiments. Soil and Tillage Research, 179, 47–53. https://doi.org/10.1016/j.still.2018.01.009

CHAPTER 2

Few to Non Label Water Segmentation: Benchmark and for Water Stage Measurement

Zamboni, P., Junior J. M., Gonçalves, W. N., Blanch, XEltner, A. Few to Non Label Water Segmentation: Benchmark and for Water Stage Measurement. To be submitted in Hydrology and Earth System Sciences.

Abstract

Visual based solutions are a non-concat and cheaper solution for water mapping and water stage estimation. Therefore, image based solutions are an alternative in order to increase the hydrological monitoring network. Many of this solution utilizes deep learning for automatic segmentation of the water area, thus, relies on large annotated datasets. Here, we propose a deep learning approach for water segmentation and water level estimation that relies on minimum to non annotated dataset. To minimize the labeling effort, we considered a video object segmentation network (STCN), SAM, which is an open-set promptable image segmentation , and a combination between SAM and Grounding DINO (a promptable open-set detector). Three camera gauges time series images were used along with dynamic river images captured using unmanned aerial vehicles (UAV). Segmented water masks were evaluated both qualitatively and quantitatively, moreover, camera based water stages were compared to reference values. For the UAV dataset, our results suggest STCN and SAM achieved similar results. For the STCN, changes and the initialization of the model can significantly impact results. Our findings show similar results among the tested models for camera gauges cameras. Furthermore, models were capable of producing realistic water stages, capturing variations in flow, especially for high flows, with STCN achieving the best results, being a suitable option for sub-hour monitoring. Our results showed that the tested approaches can be viable tools for an ad hoc water stage measurement, proof being capable of producing good results in different stations, under different climate and illuminations conditions with minimum annotation.

Keywords: Water stage, water segmentation, deep learning, photogrammetry.

1. **Introduction**

Water segmentation is a key task in multiple water resources applications such as river and flood mapping. Traditional computer vision and image analysis algorithms have been explored to automatically extract water areas from images. Researchers explored spatial-temporal texture based algorithms to detect river water lines in images from stationary cameras (Eltner et al., 2018, Stumpf et al., 2016). Other utilized grayscale intensity, textural features, or motion segmentation or combination of these methods (Penã-Haro et al., 2021), and background subtraction, and morphological and color analysis for flash flood detection in surveillance cameras (Filonenko et al., 2015) for river water segmentation. However, traditional image processing approaches to extract water are prone to errors due to changing environmental conditions (Eltner et al., 2021), and more robust methods are required to deal with these limitations. In this context, deep learning techniques, notably Convolutional Neural Networks (CNNs), have garnered significant interest within the environmental science community due to their efficacy and resilience as tools for image processing.

Several studies have applied deep learning in the form of CNNs to derive water areas from remote sensing images (Ling et al., 2019; Mullen et al., 2023). In the case of camera gauges at rivers, deep learning gained attention in recent years because of the ability of adapting to challenging scenarios, such as different flow conditions, illuminations, sediment concentration, or vegetation growth. For instance, Vitry et al. (2019) utilized 1214 images from Chaudhary (2018) and 300 images from the Cityscapes (Cordts et al., 2016) dataset for water segmentation. In this study seven frames from each camera gauge station were used and data augmentation, to artificially increase the training dataset, and fine-tuning, to adapt already trained models to the new domain, were applied. Vandaele et al. (2021a), utilized two water dataset, i.e., LAGO (Lopez-Fuentes at al., 2017) with a total of 300 images and WATERDB (Vandaele et al. (2021b) composed of 12,684 images. The authors combined the datasets with two different networks, Resnet50-Upernet and DeepLab V3, along with transfer learning to extract water areas in images. In another camera gauge setup, Muhadi et al. (2021) used 710 images as training data for the DeeplabV3+ and SegNet models. The idea behind the transfer learning relies on using features learned from one problem and using it on a similar and new problem, helping the initialization of this new model. Model fine-tuning consists of re-training the model, or parts of the model, using new and small datasets with a very low learning rate. Wagner et al. (2023) studied the potential of offline and online augmentation using 32 different CNNs for water segmentation to improve their generalization with a dataset composed of 1,128 labeled images. Data augmentation is a commonly used technique during the training of deep learning models where the dataset is artificially increased using a series of transformations to avoid overfitting and to increase model robustness and generalization. Eltner et al. (2021), trained two CNNs, SegNet and FCN, using 20,309 annotated images captured by a Raspberry Pi camera.

One obstacle to the adoption of deep learning for image segmentation is the fact that large labeled datasets are required to properly train supervised models (Sung et al., 2018; Feyjie et al., 2021). Collecting and labeling data for semantic segmentation or video segmentation can be notably expensive (Liu et al., 2020). Main datasets, which are used as benchmarks for image segmentation, such as the Pascal VOC (Everingham et al., 2010), Cityscapes (Cordts et al., 2016), and ADE20K (Zhou et al., 2016, are composed of thousands of hand annotated images. In environmental applications, collecting and labeling data is even more difficult due to challenging conditions to acquire the data and due to the often fuzzy and complex nature of the objects of interest. Various environmental obstacles such as rain, fog, and unfavorable lighting have to be considered. Labeled datasets require samples for all different conditions to achieve a robustly trained model. Furthermore, models trained with a large dataset for a specific river or region can become “too good” for that specific dataset, being not that easily transferable to different rivers/regions due to overfitting. Thus, there is a need to study approaches that can be used to reduce the effort expended to create large training datasets in environmental applications. Reducing labeling effort in water segmentation for camera gauges and flood monitoring is a fundamental step towards a more dense network of monitoring networks, enabling better water resources management and planning, especially in face of the climate changes.

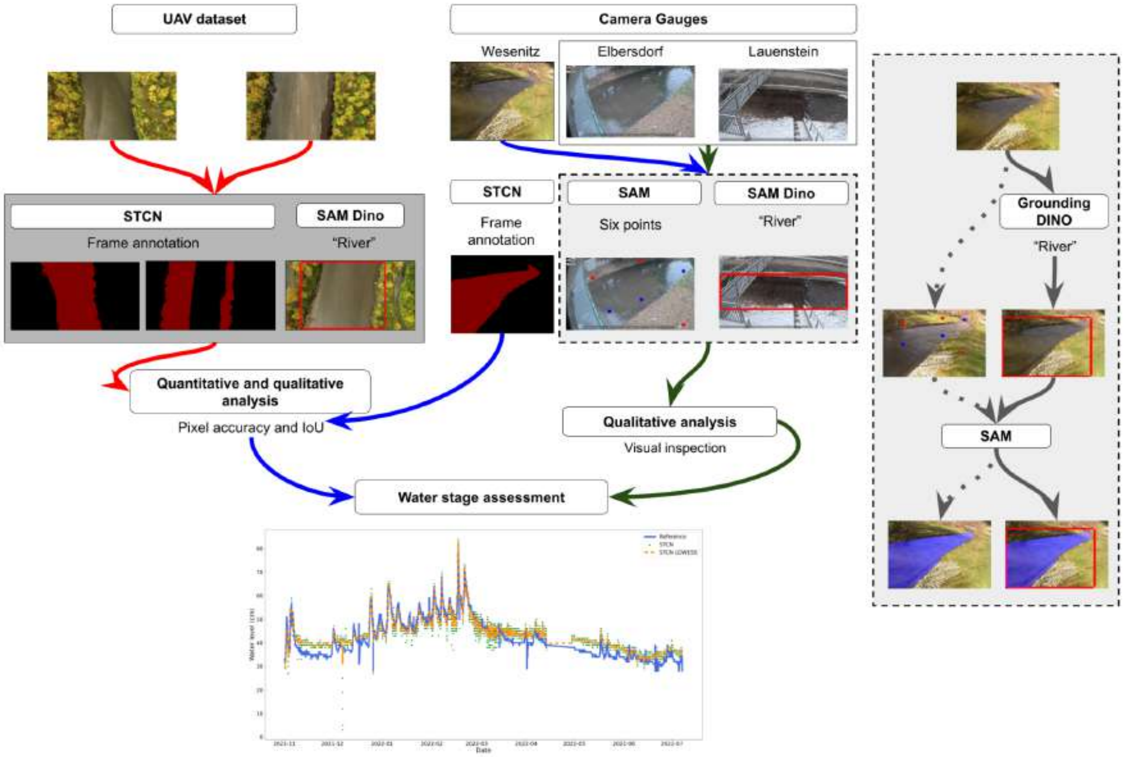
Recent developments in deep learning for image and video segmentation can be useful to tackle such issues. In video object segmentation (VOS), the goal is to produce segmentation for class-agnostic objects in a video (Ge at al., 2021). Several models have already been proposed, with emphasis on Space-Time Memory networks, or STM (Oh et al., 2019). STMs sequentially analyze video frames, starting from the second frame and using the annotation for the first frame. During frame processing, previously frames object masks are treated as memory frames. Space-Time Correspondence Networks, STCN, is a video segmentation network, proposed by Cheng et al., (2021), that works similar to STM, achieving high performance in terms of segmentation and speed. Main difference between STCN and STM relies on the memory bank and affinity calculation. In the STM, for each object in the video a specific memory bank is built and the affinity calculated, while in STCN a single affinity matrix is built.

Also in the context of minimizing labeling efforts, great advances were achieved in computer vision with the introduction of pre-trained open-set models with a high degree of generalization, such as Grounding Dino, proposed by Liu et al. (2023) for object detection and Segment Anything (SAM) proposed by Kirillov et al. (2023) for image segmentation. Grounding Dino combines language and visual models to achieve an open set detector that works with huma inputs (such as object names). In other words, Grounding Dino locates objects in images given a text prompt. Meanwhile, SAM is a foundation model for image segmentation, capable of producing segmentation masks with prompt input (points, text, or bounding box) by a user. Foundation models are neural networks trained with a large amount of raw data and can be adapted to several tasks. SAM has gained attention from researchers from different areas, e.g. medical images (Mazurowski et al., 2023; Chauveau and Merville, 2023), and crater mapping (Giannakis et al., 2024).

To the best of our knowledge, the application of neither VOS nor Groduning Dino and SAM were explored for water segmentation with the special focus on camera gauges and flood monitoring. The novel contribution of this study is to evaluate approaches for water segmentation in the context of water stage estimation, with minimum to non annotated dataset. Additionally, we evaluated these approaches in dynamic scenarios with river unmanned aerial vehicles (UAVs) dataset. We evaluated the performance of the STCN network and the combination of Grounding Dino and SAM, from hereon referred to as SAM Dino, to automatically extract the water mask from image sequences. Three different datasets of image time series, acquired by static camera gauges, were used to demonstrate the generalization power of such networks. In the same sense, a UAV dataset was used to show that the methods can handle image sequences captured by a moving platform and from a very different perspective. Eventually, we used the water masks from the camera gauges to derive the water height to assess the performance of the AI approach to measure water stages. Our work does not aim to replace existing water segmentation methods or camera gauge setups but to support them for water level measurement or in flood monitoring.

2. Material and methods

Our proposed approach can be seen on Figure 1. Using images for the camera gauges stations, we produce masks using STCNm SAM using points along the river and outside the river as input, and SAM Dino. We assess model performance both qualitatively and quantitatively. Afterwards, masks produced by all the models were used to estimate the water stage for each station and values were then compared to reference values for the nearest water stage sensor. For the UAV dataset, we evaluated STCN and SAM Dino, quantitatively and qualitatively. Once this dataset is dynamic, fixing points alongside the river was not possible, therefore, we were not able to use SAM with fixed points. We used the best results for the UAV dataset and combined river mask segmentation from the original images with a 3D point cloud of the river to segment the river shores and estimate water stage.



,

Figure 1: Study workflow. Masks for the UAV dataset were produced using STCN and SAM Dino and after that we assess the performance of both models in terms of pixel accuracy and IoU. For the camera gauges datasets, we produce water masks using STCN, SAM with six points and SAM Dino. For the Wesenitz dataset, we assess model performance in terms of pixel accuracy and IoU, and for the Elberdosf and Lauenstein datasets we conducted a qualitative analysis. For all three camera gauges datasets we assess the water stage and compare it with reference data.

2.1. Images datasets

Four datasets were explored in this work. The first one, proposed by Eltner et al. (2021), is composed of 20,309 images collected using a Raspberry Pi camera V2 with a resolution of 2,592 x 1,944 pixels. The camera was mounted at the Wesenitz river in eastern Germany. The images were collected from March 30th, 2017 until April 30th, 2018. Thereby, data was acquired in 15 frames every 30 min. During the observation period the camera position had been changed three times due to system failures. Due to the long observation period it was possible to capture images during different environmental conditions, which allows us to assess the models performance in different scenarios. Reference water stages were measured by a nearby pressure gauge. These measurements were averaged for 15 minute intervals. From here, we will refer to this dataset as the Wesenitz dataset.

The Elbersdorf dataset is part of the KIWA (Artificial Intelligence for Flood Warning) project (Blanch et al., 2022). The dataset contains 14.281 images of the Wesenitz river (Germany) acquired using an AXIS Q1645-LE surveillance camera with a 1980x1280 pixel resolution at a 4.3 mm focal length. The camera is located at the Elbersdorf gauging station of the Saxony HWIMS network, thereby enabling comparison with official measurements from the discharge station. The dataset consists of images captured every 15 minutes during daylight hours, spanning from August 23th, 2021 to July 07th, 2022. In addition, the Lauenstein dataset contains 12.880 images of the river Müglitz (Germany) and is also part of the KIWA project.. The images are acquired using the AXIS Q1645-LE camera, with a 1980x1280 pixel resolution and an approximate focal length of 6 mm. Located at the gauge station in the town of Lauenstein, the camera is also part of Saxony's official measurement network. The dataset contains images taken at 15-minute intervals between December 09th, 2021 and July 07th, 2022.

The UAV dataset consisted of 367 images. The data was captured in Northern Finland to study the river flow and hydromorphology of a river in cold climates (Eltner et al., 2021). The studied river reach has a length of about 1 km and the data was captured in autumn 2020 during low-flow conditions. The images were acquired with a DJI Phantom 4 RTK with a resolution of 5472 x 3648 pixels.

2.2. Deep learning approach for water segmentation

Space-Time correspondence Network - STCN

STCN (Cheng et al., 2021) is a simple and efficient video object segmentation (VOS) network. The key advantage of STCN compared to typical image segmentation networks is that VOS can produce segmentation masks for an entire video or a sequence of images by using only the first annotated frame. Therefore, STCN maintains and updates a “memory” of important features as it processes subsequent frames. Initially, key features are extracted from the query frame and the memory frame using a Siamese key encoder and RGB information. The features are used to compute an affinity matrix. Afterwards, an encoder-decoder network uses the affinity matrix to transform the mask features, stored in the memory, to produce the mask for the frame being queried. The memory is periodically updated every five frames to ensure it aligns with the evolving characteristics of the video. This enables STCN to adapt and provide accurate object segmentation throughout the video sequence.

We label the camera gauges datasets by manually labeling the first frame of the image sequences as can be seen in Figure 2. For the UAV dataset, we label two frames, the first and the frame number 40. We chose the 40th frame after the initial experiment and where we observe a drop in performance around the chosen frame.

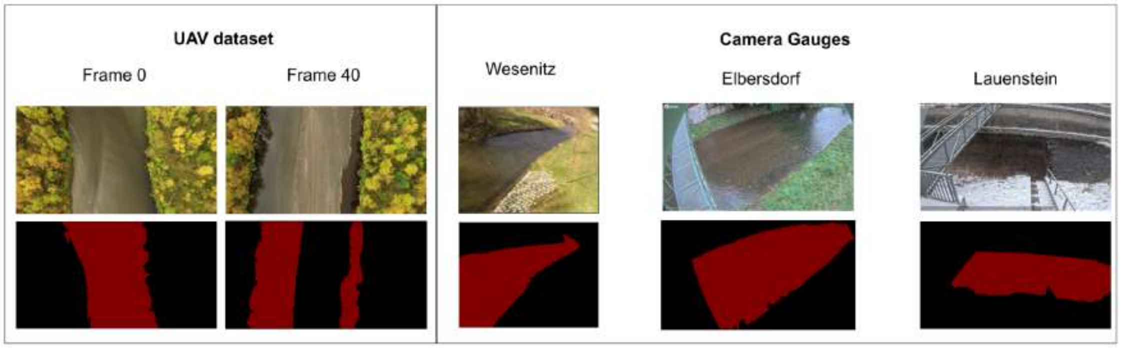


Figure 2: Initial label for each dataset used for STCN. For the UAV dataset we started STCN using two different frames, the first one on the sequence and the 40th frame.

Grounding Dino and Segment Anything model (SAM)

Grounding Dino is a state of the art open-set object detector, able to locate a large range of objects using minimum human inputs, such as text. The model is based on the Dino object detector (Zhang et al., 2022) with grounded pre-training. The key idea to achieve an open-set object detector is the combination of language models with closed-set object detectors. Grounding Dino is composed of a dual-encoder and a single decoder architecture. The dual encoder is composed of an image backbone that extracts image features, a text backbone for text features extraction, a feature enhancer for cross-modality features fusion and a language-guided query selection module that selects cross-modality queries from the image features. Cross modality queries are processed by a cross-modality decoder.

SAM is an innovative approach for promptable image segmentation proposed by Kirillov et al. (2023), inspired by natural language processing models. SAM can receive as prompt input single point, multiple point, bounding box coordinates, masks, and text. It can produce multiple segmentation masks for prompts that are associated with multiple or ambiguous objects. The model is composed of a heavyweight image encoder, a prompt encoder and a mask decoder. SAM image encoder is based on a masked autoencoder pre-trained Vision Transformers that produce an image embedding that will be queried by the input prompt. In the prompt encode, sparse prompts (e.g. points, bounding box, text) are represented by position encoding, while dense prompts (e.g. masks) are embedded by convolutions. The image decoder was employed by a Transformer decoder block alongside a dynamic mask prediction head. SAM was trained using a dataset composed of 11 million images and 1.1 billion masks and trained progressively, known as SA-1B. SA-1B was developed in three stages: a model-assisted manual annotation stage, a second stage semi-automatic stage combining automatically predicted masks and model-assisted annotation, and finally an automatically generated mask by prompting SAM with a set of points.

We use a combination of Grounding Dino and SAM (SAM Dino) for all datasets and SAM for the camera gauges datasets (Figure 1). When using SAM Dino, the query “river” was used as input to Dino, which returned a bounding box with the river location, later used by SAM to create a river mask. On the other hand, for SAM alone, we selected six fixed points for each camera gauge dataset, once the images were static. Using these six points, three for the background class and three for the water class, SAM created the river masks.

Experimental setup

For the camera gauges datasets, the first frame on each sequence was manually annotated and used as initial segmentation mask with STCN model. For SAM, we select three points representing the class water and three points representing the class background. We used three points inside the river once in initial experiment SAM failed to segment the whole water area, therefore, we chose three points along the whole extent of the water area to minimize such issues. For the UAV, we hand annotated two frames, the first frame on the sequence and the 40th frame, once during initial experiments we notice a drop in segmentation performance around the 40th frame. For both camera gauges and UAV datasets, we used the query “river” as input.

STCN, SAM and SAM Dino models were used with pre-trained weights, therefore, we did not re-train any of the models. Inference procedures were conducted using a NVidia RTX 3090 graphic card, with a Ryzen 5700X and 32 Gb of ram memory.

Segmentation models were evaluated using pixel accuracy (ACC) and intersection over union (IoU); two standard metrics for semantic segmentation. ACC (Equation 1) represents the percentage of correctly classified pixels to each class, where an ACC equal to 0 implies that all pixels were wrongly classified, and an ACC of 1 that all the pixels were correctly classified. IoU (Equation 2) quantifies the overlap between the ground-truth and the predicted masks, being the ratio between the intersection and the union of the predicted and the ground-truth area. Therefore, IoU will be equal 1 indicates a perfect match between both masks.

$$ACC \ = \frac{TP + TN}{TP + TN + FP + FN} \#1$$

$$IoU \ = \frac{GT \cap P}{GT \cup P} \#2$$

2.3. Water stage estimation and performance assessment

Water stage was estimated based on methodology proposed by Eltner et al. (2021). For all datasets, 3D georeferenced models were built using Structure from Motion (SfM) strategies. Ground control points were measured using a multi-band GNSS equipment. Following, the 2D coordinates from water masks automatically segmented by the deep learning models were intersected with the 3D model for the respective station. In the 3D model, the Z-coordinate of nearest neighbor points to the ones projected were considered the water height. In order to reduce noise and outliers, we applied LOWESS (Locally Weighted Scatterplot Smoothing). For the Wesenitz station, once the images were not recorded in a constant time series, we used images from 2017-05-15 until 2017-06-23 as a proof of concept for the water level estimation. For the Lauaenstein and Elberdorf station, we used only daylight images to assess the water stage.

To evalured water stade estimation performance we used the mean absolute percentage error (mape), mean absolute error (mae), mean squared error (mse), root mean square error (rmse), pbias, Nash-Sutcliffe efficiency (nse), Kiling-Gupta efficiency (kge), coefficient of determination (r^2), Spearman correlation, mean error and error standard deviation.

3. Results and Discussion

First, we present a quantitative and qualitative analysis of the deep learning approaches for the water segmentation using Wesenitz and the UAV datasets. Then, we present a qualitative analysis of the other two camera gauge datasets. Finally, we assess the water stage for all three camera gauge datasets and compare to reference water stage values.

3.1. Water segmentation performance

Table 1 presents results for the Wesenitz dataset in terms of pixel accuracy and IoU for the background and water classes. SAM Dino achieved the best overall performance, considering all images and the test set, followed by STCN, with models scoring more than 0.9 for all metrics. SAM using six points achieved the worst overall performance. This might be related to the fact that the camera for this dataset had to be changed three times due to system failures, changing camera perspective, therefore, we can not ensure that the selected points were on the river and outside the river for all images. Our results, considering SAM Dino and STCN, were similar to results founded in the literature for the same dataset, being around 0.02% and 0.025% lower for water pixel accuracy and IoU, respectively, for SAM Dino, and, for STCN, 0.03% and 0.06% lower for water pixel accuracy and IoU, respectively. The fact that the camera for this dataset had to be changed several times during image collection could have led to worse results using STCN compared to SAM Dino. Using Dino, for each image or frame, a new river bounding box is assigned, with no interference or relation to previous or future frames, allowing SAM Dino to produce accurate results even when the camera poses changed.

Table 1. Metrics for the Wesenitz dataset. For the Test set, we present metrics only for images presented on the test set, and for FD (Full Dataset), metrics for all images on the dataset. Δ is the difference from our approaches and Eltner et al. (2021) best model.

Model	**Set**	**Pixel Accuracy (Background)**	**Pixel Accuracy (Water)**	Δ	**IoU (Background)**	**IoU (Water)**	Δ
STCN	Test	0.977 ± 0.041	0.949 ± 0.026	-0.031	0.929 ± 0.042	0.923 ± 0.057	-0.057
	FD	0.976 ± 0.043	0.948 ± 0.030	-0.032	0.927 ± 0.044	0.921 ± 0.061	-0.059
SAM Dino	Test	**0.996 ± 0.023**	**0.961 ± 0.022**	-0.019	**0.960 ± 0.030**	**0.957 ± 0.030**	-0.023
	FD	**0.996 ± 0.023**	**0.960 ± 0.022**	-0.020	**0.959 ± 0.031**	**0.956 ± 0.032**	-0.024
SAM 6 points	Test	0.942 ± 0.126	0.673 ± 0.360	-0.307	0.781 ± 0.185	0.672 ± 0.359	-0.308
	FD	0.937 ± 0.135	0.668 ± 0.359	-0.312	0.775 ± 0.191	0.667 ± 0.360	-0.313
Eltner et al. (2021) best model		**0.982 ± 0.006**	**0.980 ± 0.018**	-	**0.982 ± 0.006**	**0.980 ± 0.018**	-

As can be seen on Figure 3, STCN and SAM Dino showed a higher temporal agreement in terms of water pixel accuracy and water IoU, with 75% of the images scoring metrics above 0.9. SAM using six points present a higher desagreement for both metrics, scoring values above 0.9 only in 25% of the time. Considering the Wesenitz dataset, both STCN and SAM Dino were able to produce accurate masks even in challenging lighting conditions and camera position.

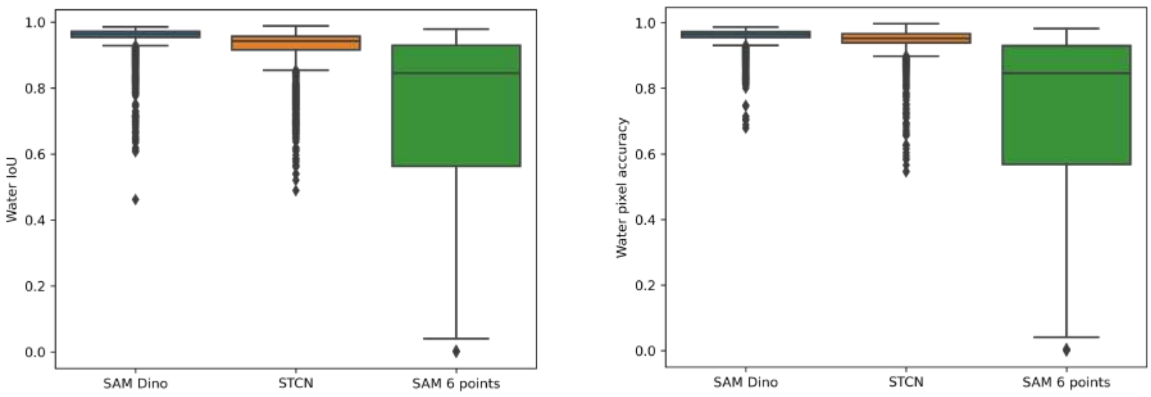


Figure 3. Boxplots for Wesenitz for all three models, considering all the images on the dataset. On the left, water IoU and on the right, water pixel accuracy.

Previous research for water segmentation in camera gauges applied several methodologies using standard image segmentation networks, combining with data augmentation, fine-tuning, and transfer learning strategies. Researchers had found results, in terms of water pixel accuracy and IoU higher than 0.9 on average (Vitry et al., 2019; Vandaele (a) et al , 2021; Muhadi et al., 2021). Wagner et al. (2023), presented results for 32 different neural networks for this task, combining with online and offline data augmentation. When not using data augmentation, results ranged from 0.828 up to 0.928 for water IoU, and when using offline data augmentation, average results ranged from 0.893 to 0.980. Previous approaches relied on large datasets and training of image segmentation neural networks. On the other hand, our approach relies on minimum annotation and our results showed a minimal trade-off between large hand annotated datasets and performance. Moreover, as on Section 3.2, we present water stage results using these approaches in three different camera gauge stations, showing that our approach is robust and can be easily adopted in different environments, regarding static images.

Table 2 shows the results of STCN and SAM Dino models for the UAV dataset. For this dataset, we did not apply SAM with fixed points, once the images were not static. For STCN we tested using two different frames as the initial frame. First we used the first frame on the image sequence (frame 0) and the frame 40, chosen after analysis of the initial results. STCN, using frame 0, and SAM Dino achieved similar results, being SAM Dino slightly better. STCN showed a drop in segmentation performance around the 40th. Therefore, we use the 40th frame as the initial frame to assess if the change on the initial image could affect segmentation performance. Using the 40th frame of the image sequence as the annotated frame for STCN drastically improved model performance. Likewise, standard deviation decreased for this case. Figure 6 shows examples of segmentation masks produced by both variants of STCN. One possible direction can be the use of multiple frames as the “initial frame”. Initially, consecutives RGB frames can be analyzed to check strong differences in the frames, indicating the possible need to add a new annotated frame.

Table 2. Model performance for the UAV dataset considering pixel accuracy and IoU.

Model	**Pixel Accuracy (Background)**	**Pixel Accuracy (Water)**	**IoU (Background)**	**IoU (Water)**
STCN (frame 0)	0.716 ± 0.183	0.962 ± 0.061	0.705 ± 0.181	0.615 ± 0.211
STCN (frame 40)	0.971 ± 0.044	0.971 ± 0.019	0.958 ± 0.044	0.912 ± 0.086
SAM Dino	0.776 ± 0.197	0.934 ± 0.097	0.754 ± 0.197	0.674 ± 0.227

In terms of pixel accuracy, all models reached values higher than 0.9 for water class. For the background class, pixel accuracy values were around 0.7 for STCN and SAM Dino, being STCN using 40th the only modelo with values higher than 0.9. The results indicate that STCN using frame 0 and SAM Dino misclassify background pixels as water, even though both models were able to correctly classify water pixels (Figure 4). On the other hand, in terms of IoU, SAM Dino shows a higher variance in values than STCN frame 0 for both classes (Figure 4), with higher mean value.

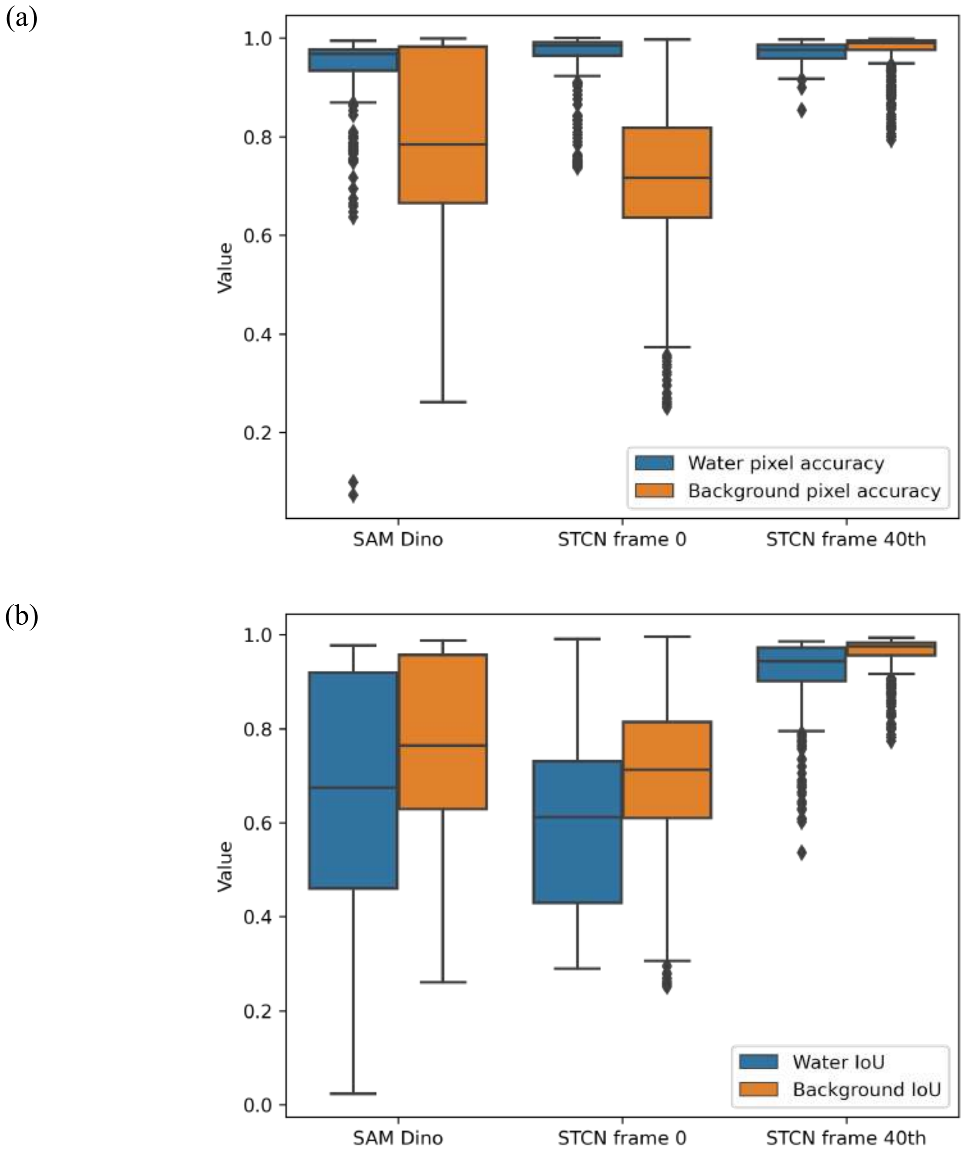
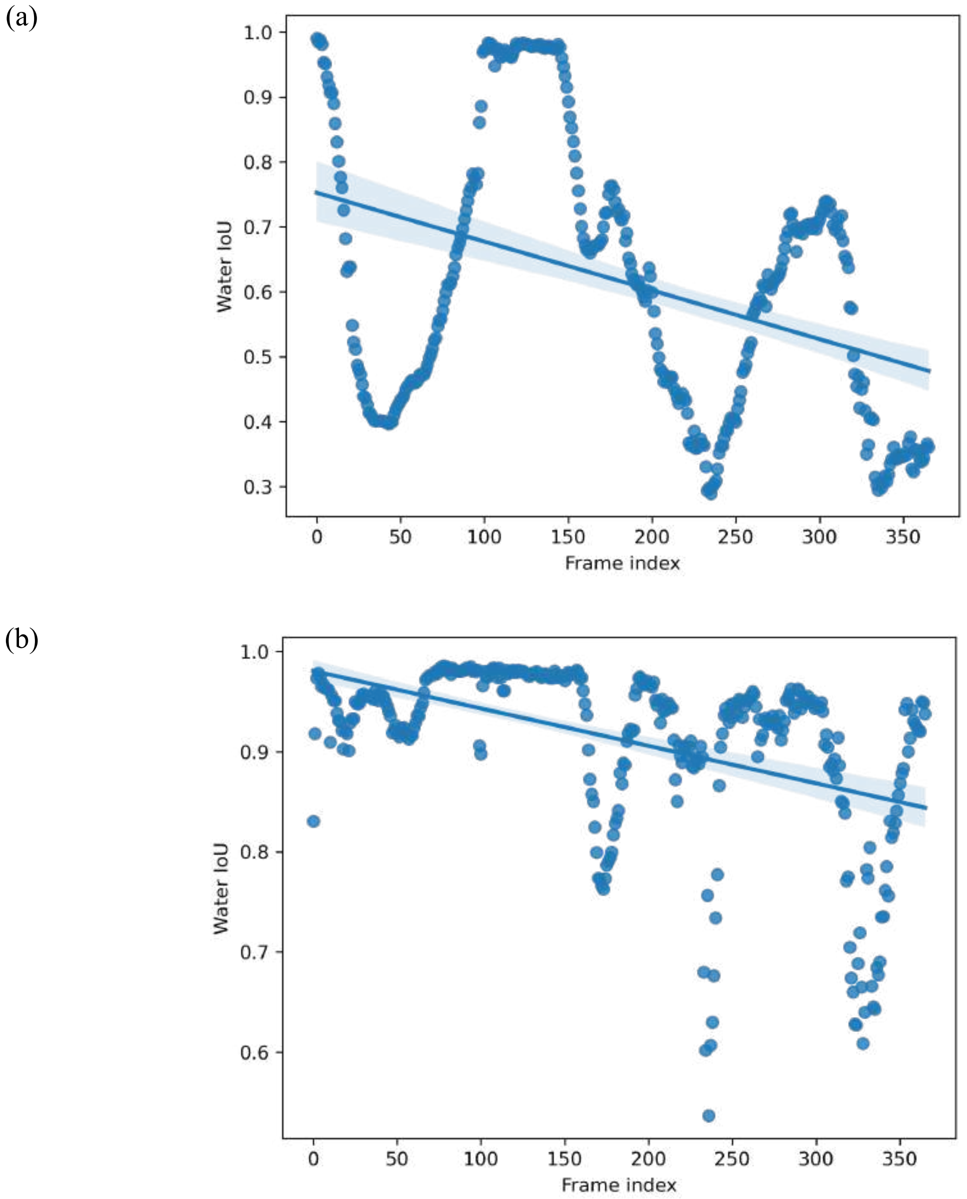


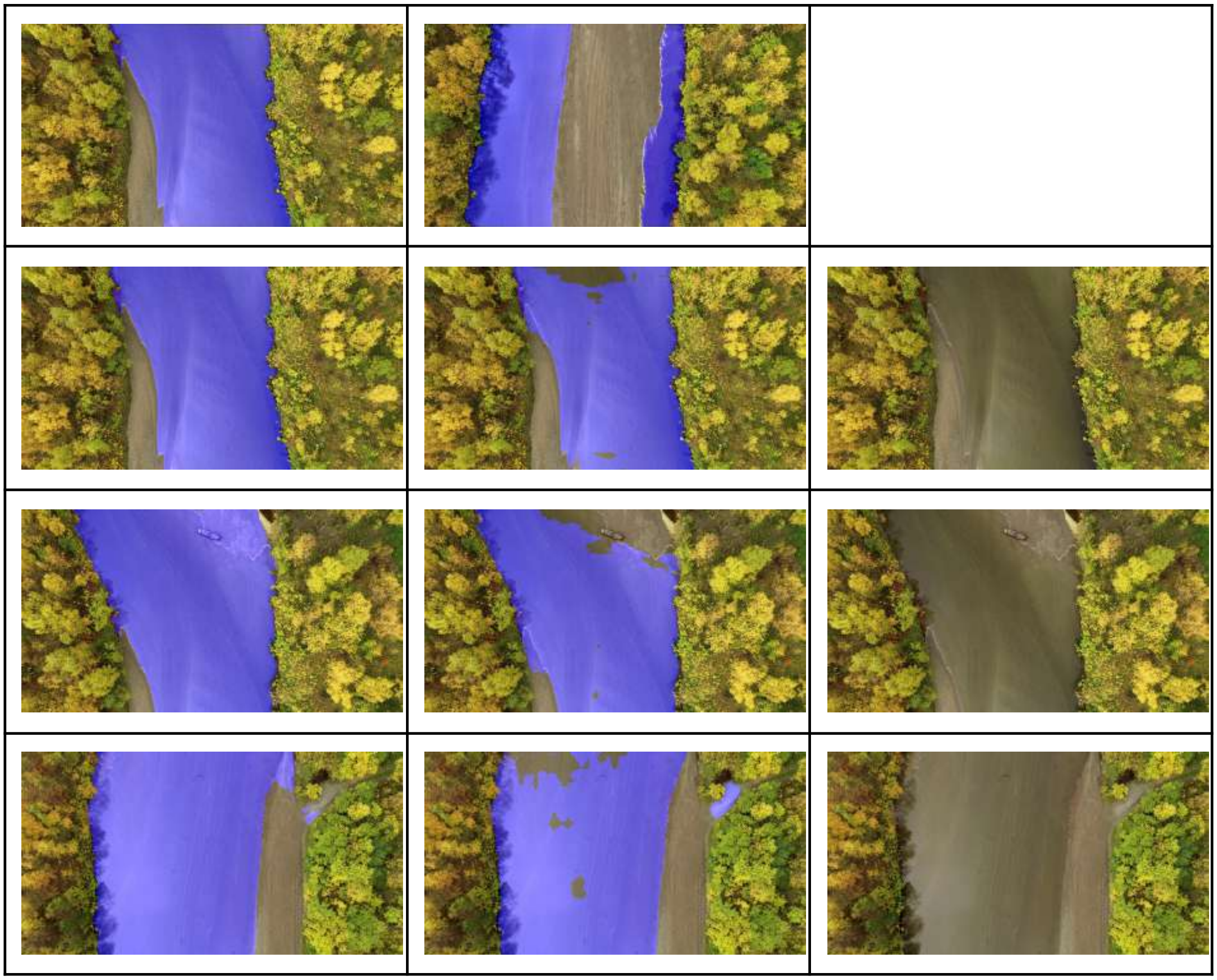
Figure 4: Boxplot for STCN using the first frame, using the 40th frame and SAM Dino regarding pixel accuracy (a), and IoU (b).



,

Figure 5: Temporal evolution of the water IoU for STCN with the first frame (a), and using the 40th (b) frame for the UAV dataset.

Figure 5 shows the temporal evolution of the performance, in terms of water IoU, for STCN frame 0 and STCN 40th frame. Overall, there is a negative performance trend during the evolution of the image sequences. STCN frame 0 presented a higher negative trend, with an overall lower performance, compared to STCN 40th frame. When starting STCN using the first frame, frame 0, the first drop in performance occurs around the 13th frame due to the presence of a water pond on one side of the river. During these frames, STCN classified the pond border as the river border, instead of the sand bank, keeping this information in the memory, propagating the error. Performance increases around the 97th frames, when the sand bank starts to disappear and the river, and decreases again due to the same issue around the frame 154. For STCN 40th frame, frames that were temporarily before the 40th frame, some problems occurred during segmentation, with some parts of the river not being correctly classified as water (Figure 6). Performance issues around the 150th frame were due to misclassification of sand river banks. Around frame 232, some river borders were wrongly assigned as water and around the 330th frame some parts of the river, especially shallow areas, were not classified as water.



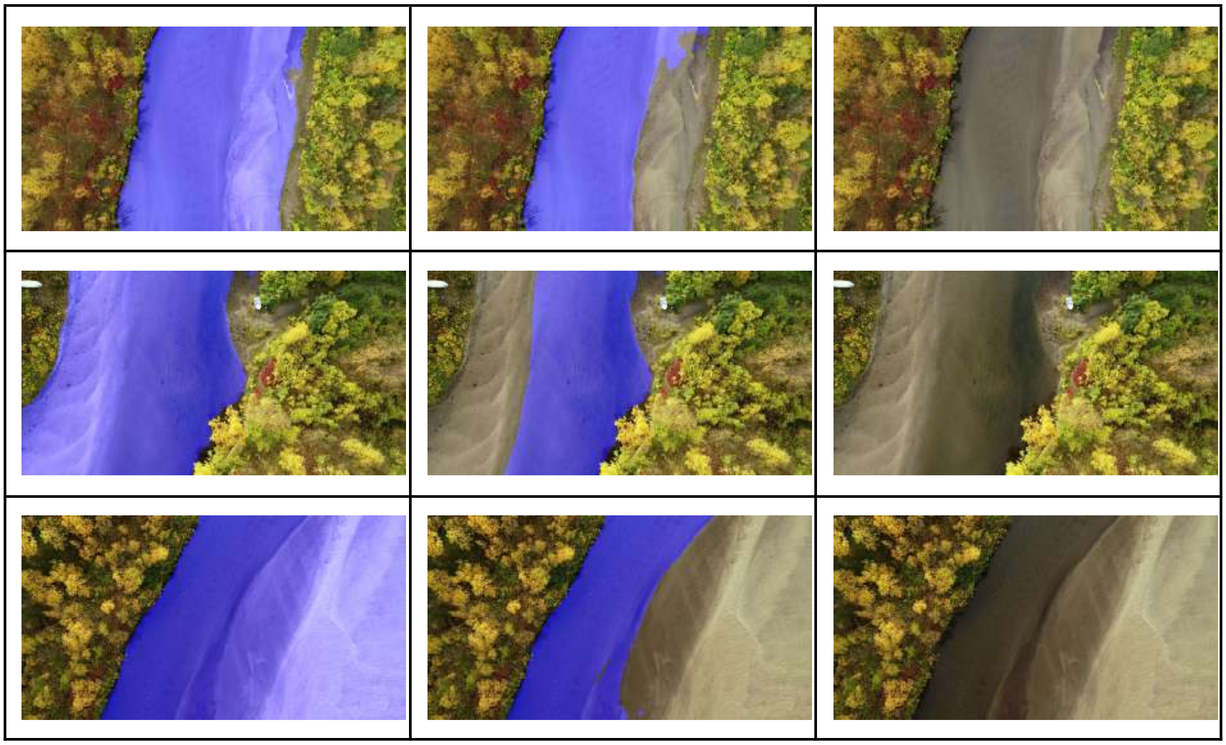


Figure 6: Examples of masks produced by STCN using frame 0 (first column), using the 40th frame (second column), and the respective RGB frame (third column). On the first row, we show the annotated frame used to start STCN, frame 0 on the first column and frame 40th on the second column.

Masks produced by SAM Dino heavily depend on the objective detention made by Grounding Dino. We did not observe correlation between the detection confidence and the water IoU (Figure 7). Nevertheless, we observed that Grounding Dino produces often larger bounding boxes that contain both the river area and parts of the shore banks (Figure 8), leading to an over segmentation of the river area. It can be noticed that the river border close to the edge of the bounding box was accurately segmented. We also observed that shadows posed a challenge on this dataset, as can be seen on Figure 9. Although, when Grounding Dino produces a more adjusted bounding box, SAM were capable of producing correct water masks (Figure 8).

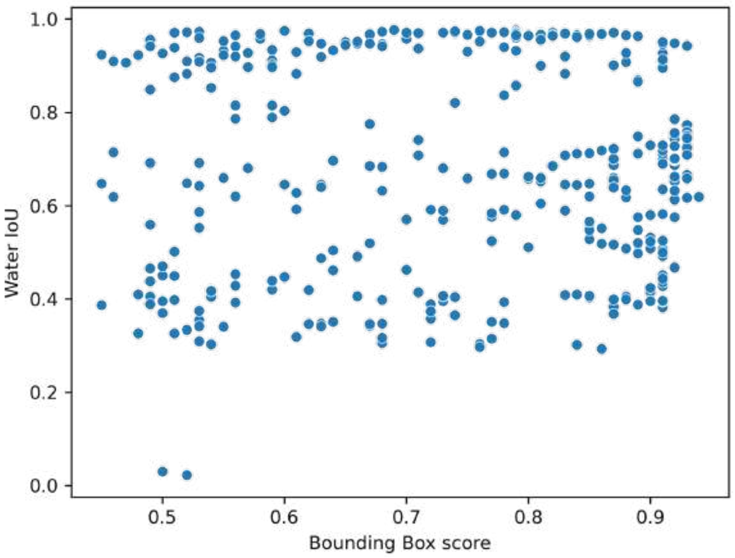


Figure 7: Water IoU for SAM Dino and bounding box score for the UAV dataset.

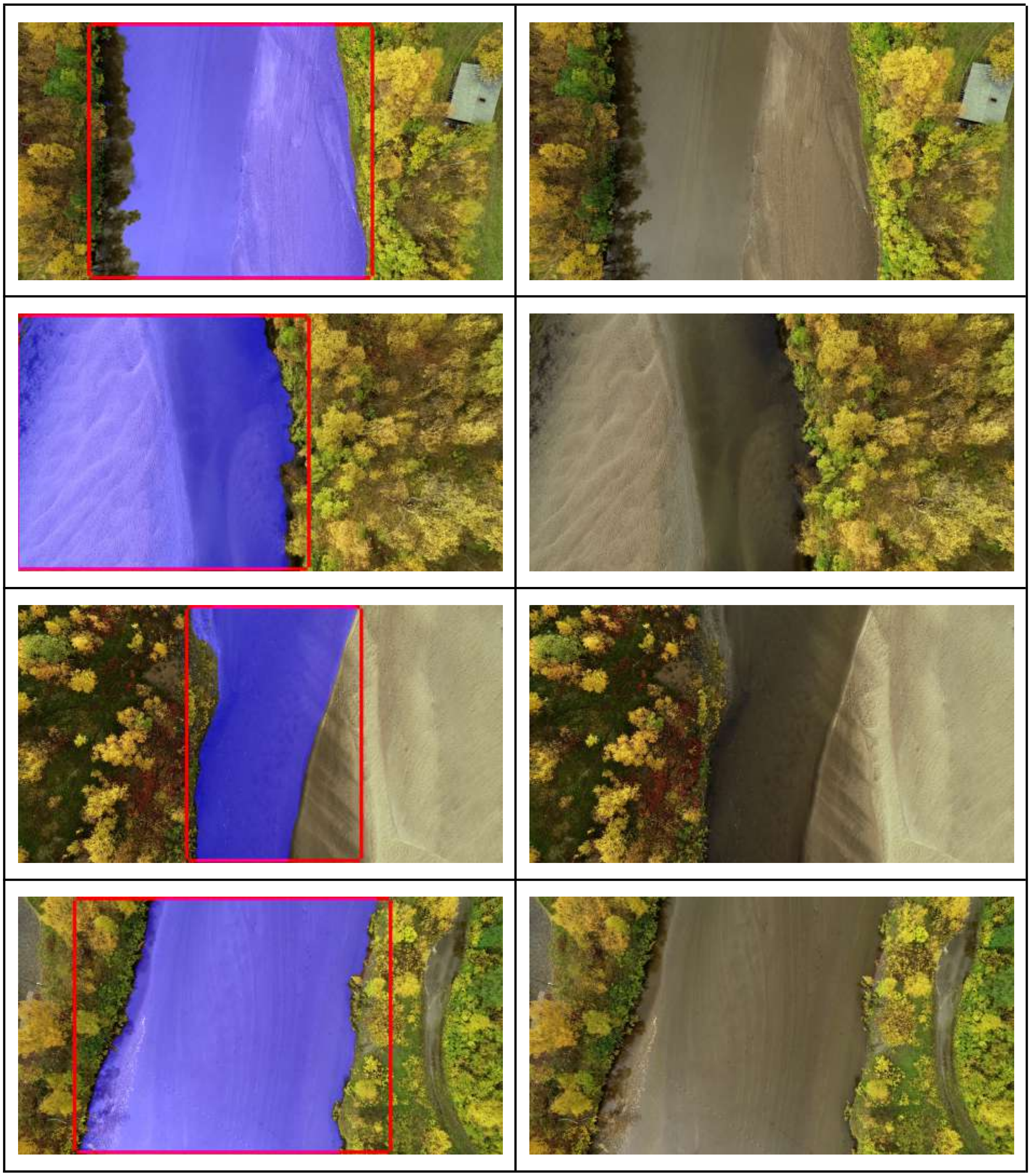
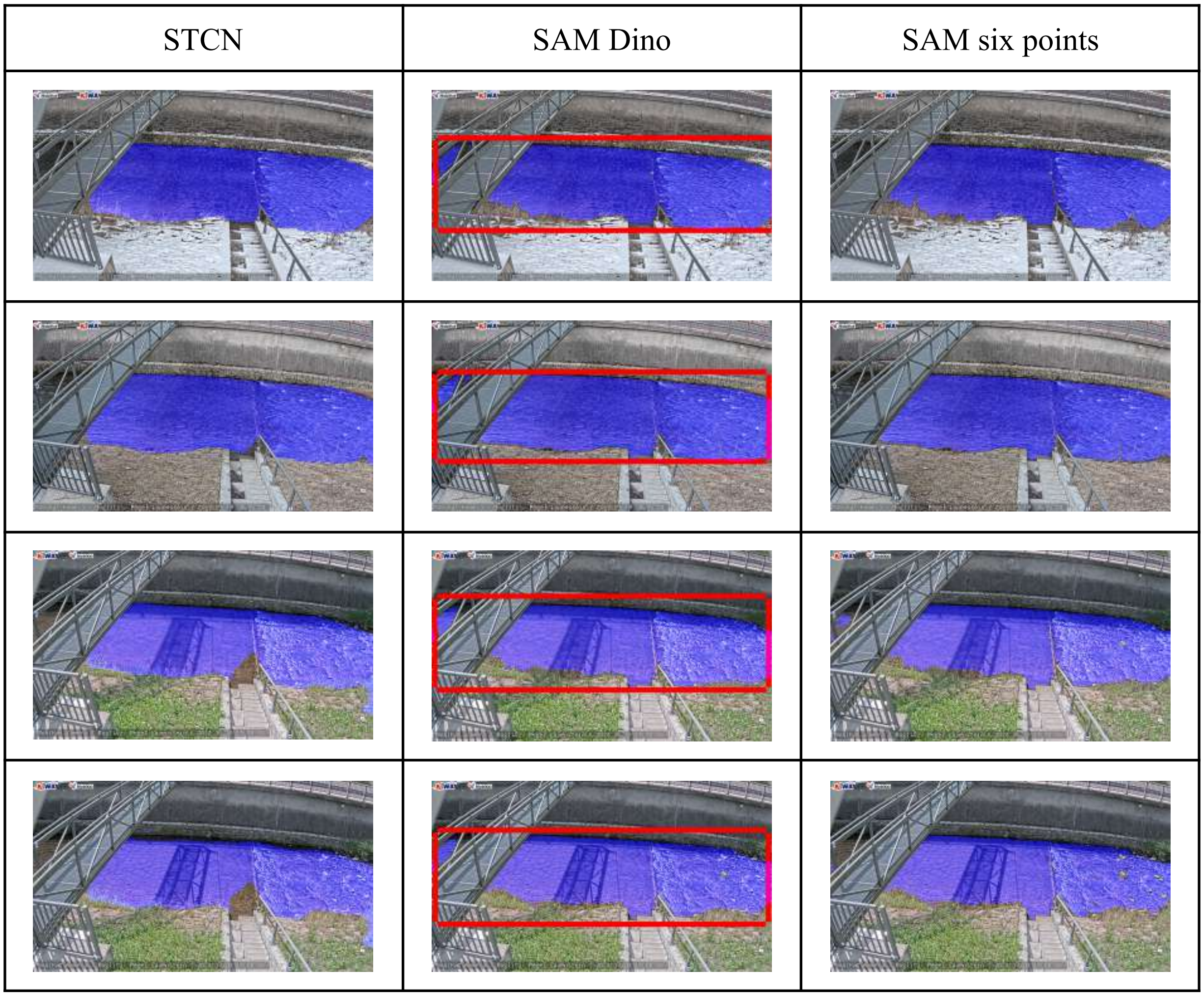


Figure 8: On the first column, examples of bounding boxes and masks generated by SAM Dino, and on the second column the respective RGB image.

Applying these methods to moving images can be considered a more challenging scenario compared to the camera gauges static images. On using these methods in camera gauges river, the object to be detected and segmented (in this case the river), the object is not moving, only expanding and contracting, alongside with changes on the surroundings. That said, on the UAV dataset the object itself is constantly changing position in different images, combining with the movement of the camera, making it a more challenging scenario. Results from the UAV dataset shows that the choice of the initial frame can dramatically impact the model performance, therefore, it needs to be carefully chosen. Nonetheless, even using the STCN with the first frame and SAM Dino, results can still provide an initial assessment of the water area, with a human operator can only correct the mask, reducing the effort needed to produce masks. One possible option for the UAV dataset would be, instead of segmenting the river area, mask the surrounding areas.

Figure A1 (Appendix A) shows results for the Elberdosf dataset. Generally, all models were able to generate accurate masks most part of the time. STCN showed a better overall performance than SAM Dino and SAM using points, with more accurate borders. In most part of the time, as can been seen on section 3.2.1., water level results using STCN for Elberdorsf dataset overestimate reference values, indicating that the used border, in this case the superior one, often includes non river pixels. Moreover, it can be seen on Figure 9, that STCN wrongly classifies snow areas and parts of the upper right portion of the image. SAM Dino and SAM using six points trend to produce masks with edges a few pixels indented towards the center of the mask.

Figure 9 displays results for the Lauenstein dataset. STCN showed better overall performance than SAM Dino and SAM using six points. The STCN model shows problems with shadow during summer, especially between 14 and 17 hours. Changes in light illumination quickly occur between the frames, associated with the time game between the frames (15 minutes), projected shadow for a few frames, and STCN was able to adapt itself. Similar to the Elberdorf dataset, both SAM based models underestimate the border toward the center of the mask. SAM Dino showed some issues during the river detection, leading to misclassification (Figure 10). One solution to overcome this issue would be using Grounding Dino to assess the best bounding box position for all images and fix the coordinates during SAM segmentation inference.



STCN

SAM Dino

SAM six points

Figure 9: Example of masks and bounding box for the Lauenstein dataset. First row: 2021-12-09 11:10:00; second row: 2021-03-18 10:30:00, third row: 2022-06-12 15:30; last row: 2022-07-03 15:00:00

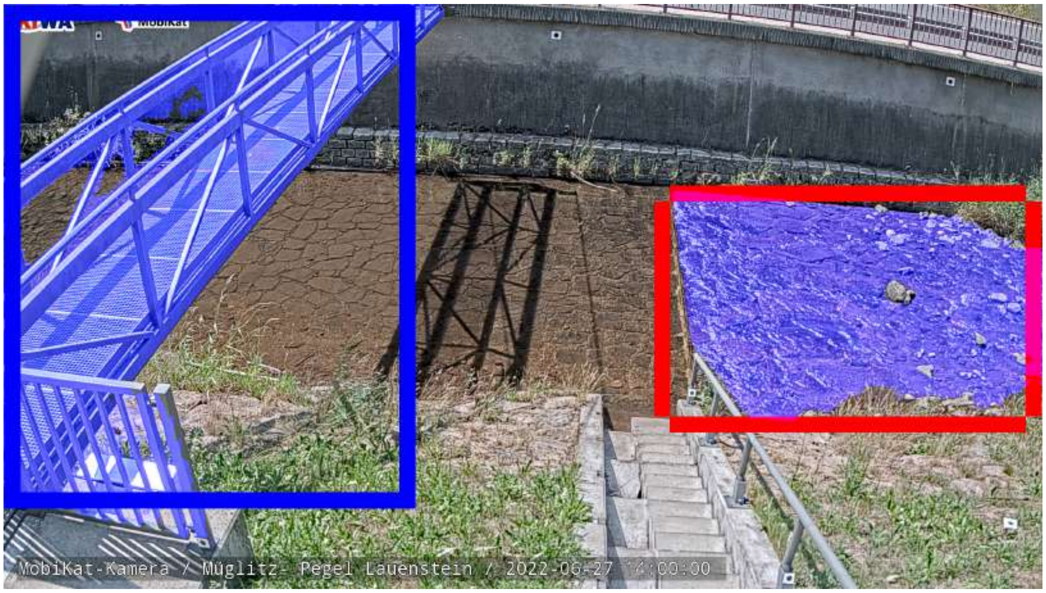

MobiKat-Kamera / Müglitz- Pegel Lauenstein / 2022-06-27 14:00:00

Figure 10:Example of error during river bounding box detection.

3.2. Water level measurement

Table 3 presents the results for the STCN model, using and not LOWESS regression, for Wesenitz, Elbersdorf and Lauentein. Water level results are presented in centimeters. Figure 11 shows results for the tree station, considering and not the LOWESS regression, only for Wesenitz station we display results with LOWESS regression and results from Eltner et al., (2021), for the same station. For Wesenitz station, LOWESS regression did not produce any improvement in terms of metrics. Average deviation for STCN was 0.793 ± 0.414. Average deviation for Ebersdorf station was 1.339 ± 3.066 for STCN and 1.338 ± 2.870 for STCN with LOWESS regression. For Lauenstein, -0.194 ± 3.271 for STCN and -0.195 ± 2.400 for STCN LOWESS. It can be observed that, even though the mean error was not affected when applying the smoothing, standard deviation was reduced for both cases, once LOWESS regression smooths the data using a fraction of nearest points. Generally, LOWESS regression reduces the error on the measured water level. For both stations, STCN achieved a similar performance in terms of RMSE, NSE and Spearman's correlation, with a marginal difference. We observed a reduction on the RMSE from 3.346 to 3.168 and from 3.278 to 2.451 for Elberdorf and Lauenstein, respectively. In the same way, NSE and Spearman's correlation increased, respectively, from 0.843 to 0.859 and 0.946 to 0.959 for Elberdorf, and from 0.881 to 0.933 and 0.936 to 0.954 for Lauenstein, respectively. Further, as can be seen in Figures 13, LOWESS regression was useful to reduce the noise on the measurements, especially for the summer portion of Lauenstein station.

Table 3. Water stage metrics for using STCN model and LOWESS regression for all tree camera gauge stations.

	Wesenitz		**Elbersdorf**		**Lauenstein**	
	STCN	**STCN LOWESS**	**STCN**	**STCN LOWESS**	**STCN**	**STCN LOWESS**
MAPE	0.035	0.035	0.072	0.070	0.194	0.143
MAE	1.375	1.375	2.859	2.775	2.094	1.758
MSE	2.682	2.682	11.193	10.035	10.742	6.006
RMSE	1.638	1.638	3.346	3.168	3.278	2.451
pBias	-3.344	-3.344	-3.149	-3.148	1.165	1.165
NSE	0.557	0.557	0.843	0.859	0.881	0.933
KGE	0.914	0.914	0.826	0.806	0.934	0.917
R^2	0.557	0.557	0.843	0.859	0.881	0.933
Spearman's correlation	0.893	0.893	0.946	0.959	0.936	0.954
Mean error	0.793	1.312	1.339	1.338	-0.194	-0.195
Error std	0.414	0.979	3.066	2.870	3.271	2.400

STCN model overestimated water level for Wesenitz and Elbersdorf stations. In the Wesenitz station, STCN model achieved similar results compared to Eltner et al. (2021), being able to correctly follow trends in the water stage, especially for higher values. In Elbersdorf station, we observed an overestimation especially for water levels below 40 cm. For water level above 40 cm, we observed a better fit, further, the model was able to successfully track the increase and decrease of the water level. For Lauenstein, the model produces a better fit for the most part of the time series, and a good overall performance considering the LOWESS model. During the end of Launstein time series, it can be observed low water stage estimations. This happened during summer, especially between 14 and 16 pm, when the wall close to the river channel projected shadow on the river and the model was not able to adapt itself. LOWESS regression reduces noises and decreases deviations in the measurements (Figure 12 and 13). Figure 13 shows the exceedance curve for all stations. The exceedance curve represents the water stage values and their relative exceedance time. We observed an overall overestimation across all values for Wesenitz station. For Elberdorf, the exceedance curve shows a trend of overestimation for low water stage values that are exceeded by about 60%, and a better agreement for higher values. In Lauenstein, the modeled exceedance curve is almost identical to the reference. In general, this approach was able to detect temporal changes in water stage, especially for high flow.

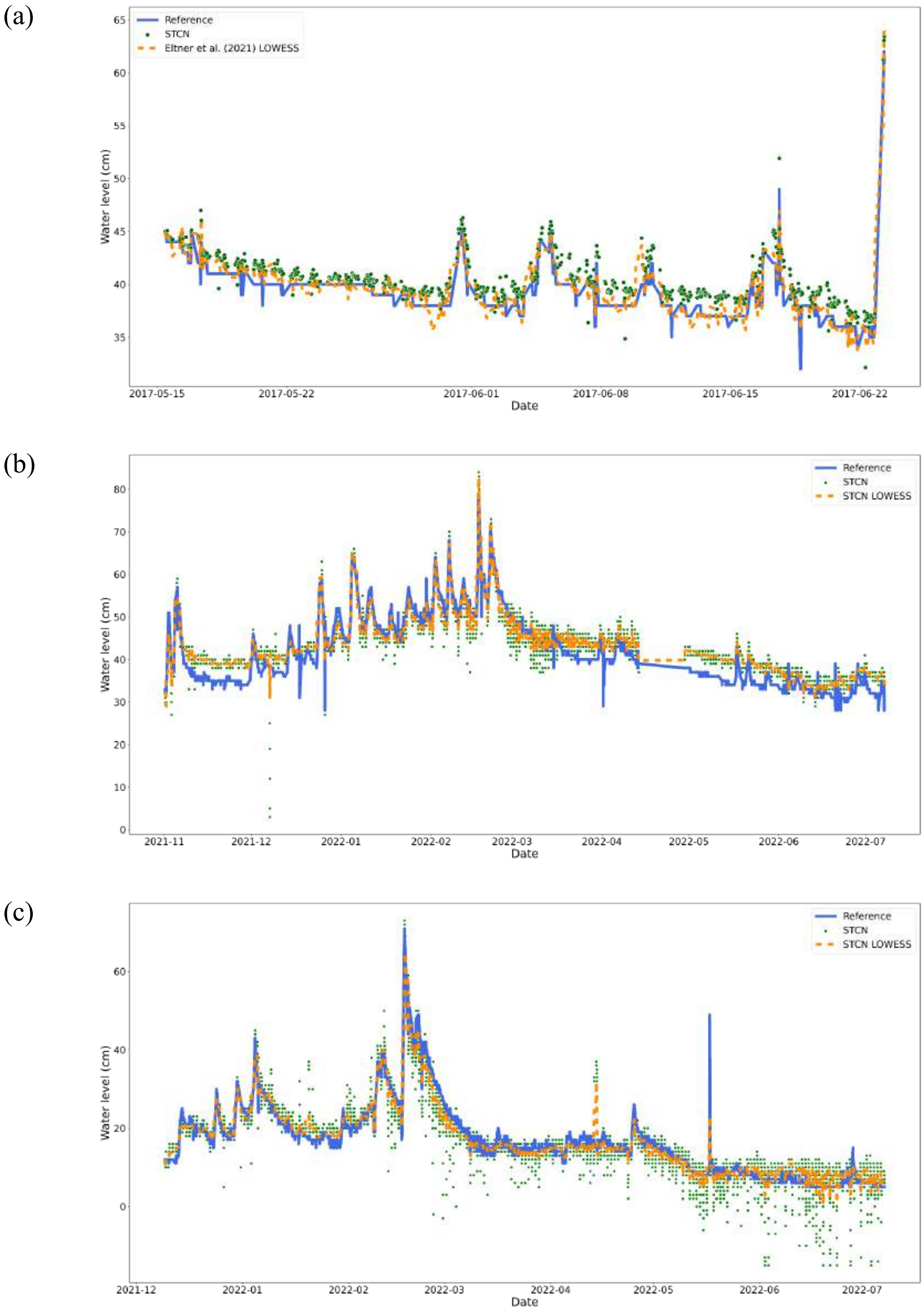


Figure 11: Water stage using STCN and STCN with LOWESS regression compared to reference data. For the Wesenitz, we show the results considering only the LOWESS regression and values from Eltner et al., (2021) for the same station. (a) Wesenitz; (b) Elbersdorf; (c) Lauenstein.

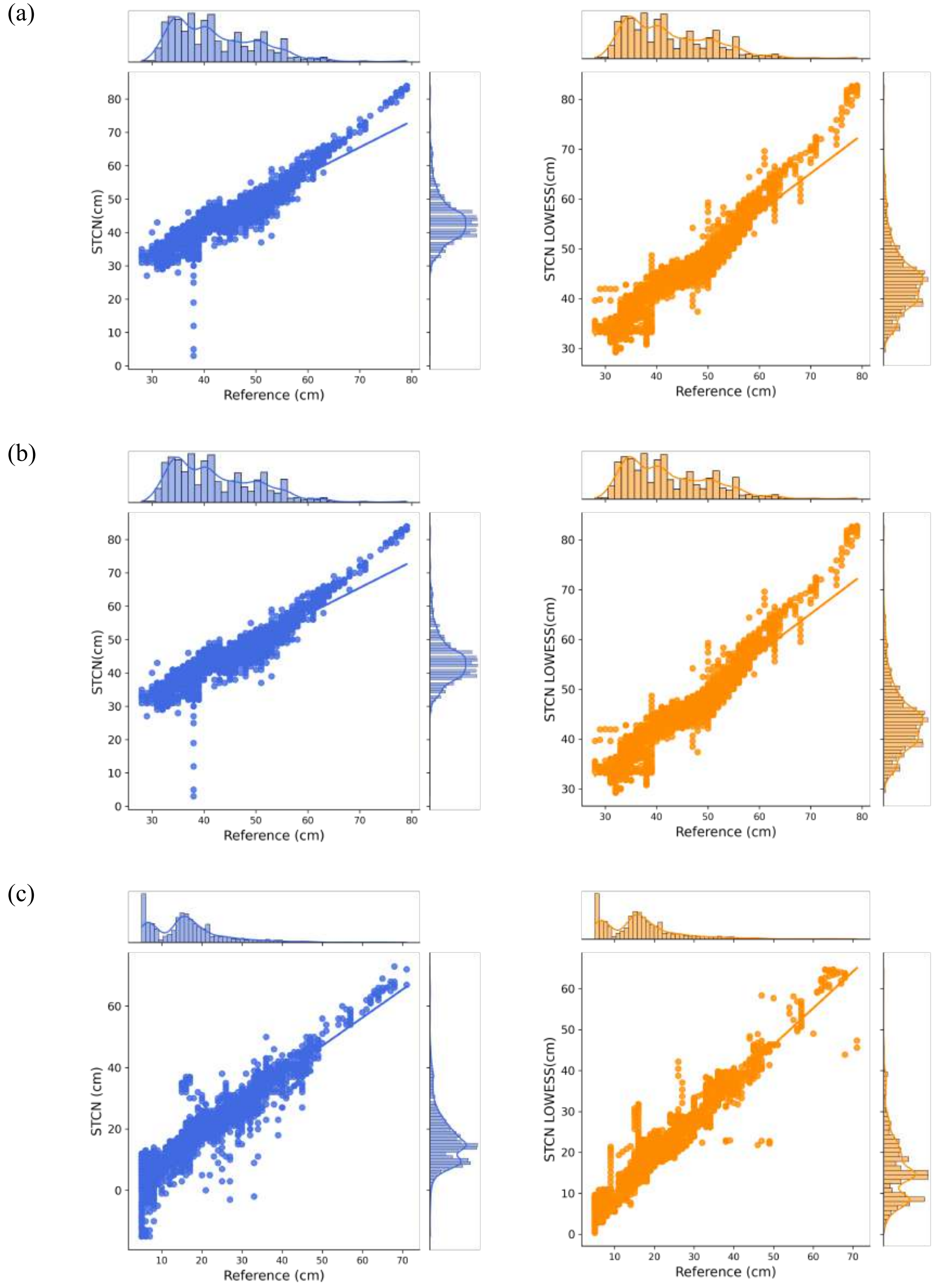


Figure 12: Regression plot with histogram for the STCN model (in blue) and STCN with LOWESS regression (in orange) for (a) Wesenitz; (b) Elbersdorf; and (c) Lauenstein stations.

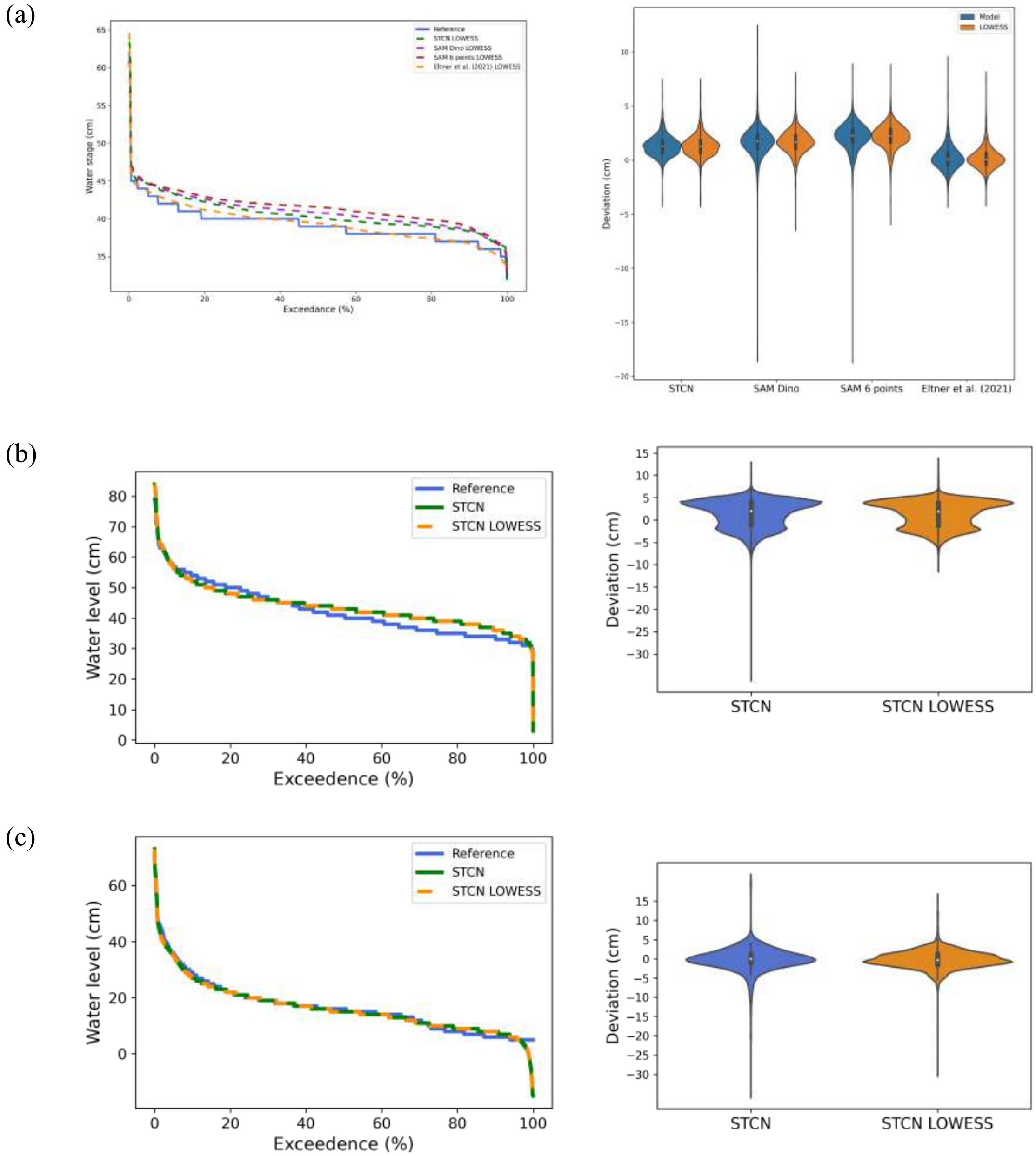


Figure 13: Exceedance curves (left column) and deviation (right column) for STCN and STCN with LOWESS regression for (a) Wesenitz; (b) Elbersdorf; and (c) Lauenstein stations.

Table 4 presents results for all three stations using the two SAM variants, using Grounding Dino and using points. Using these two approaches, we recognized two different behaviors. For Wesenitz station, both approaches overestimated the water stage, with a mean error of 2.133 ± 1.385 and 1.659 ± 1.394, for SAM using points and SAM Dino, respectively. Both models returned similar deviations in water stage, and reduced with the LOWESS regression (Figure A2). Nevertheless, similar to STCN results, for higher water stages both models produce a better fit as can bee seen on Figure A3. This overestimation can be also seen in the exceedance curve (Figure A4). On the other hand, for Elberdorf and Lauenstein both approaches underestimate the water stage, even though they were correct to detect changes in water stage and higher water stages (Figure A3). Using SAM Dino, we detect water stage values higher than 200 centimeters, caused by wrongly assigned bounding boxes, that induce SAM to produce wrong water masks. When excluding these values, SAM Dino produced similar results to SAM using points (Figure A5).

Table 4. Water stage metrics for using SAM using six points and SAM Dino model, with and without using LOWESS regression, for all tree camera gauge stations.

	Wesenitz				**Elbersdorf**				**Lauenstein**			
	6 points		**Dino**		**6 points**		**Dino**		**6 points**		**Dino**	
		Lowess		**Lowess**		**Lowess**		**Lowess**		**Lowess**		**Lowess**
MAPE	0.058	0.058	0.047	0.045	0.138	0.138	0.151	0.151	0.504	0.501	0.913	0.661
MAE	2.261	2.243	1.822	1.760	5.648	5.667	6.160	6.205	6.042	6.036	8.470	6.934
MSE	6.471	5.968	4.696	3.986	48.309	39.155	54.894	45.605	39.797	38.573	614.084	92.243
RMSE	2.544	2.443	2.167	1.996	6.950	6.257	7.409	6.753	6.309	6.211	24.781	9.604
pBias	-5.435	-5.539	-4.226	-4.245	13.245	13.246	14.457	14.461	35.650	35.650	19.583	19.596
NSE	-0.068	0.015	0.225	0.342	0.322	0.451	0.219	0.351	0.558	0.572	-5.858	-0.030
KGE	0.835	0.859	0.835	0.877	0.802	0.856	0.784	0.845	0.641	0.643	-0.841	0.542
R²	-0.068	0.015	0.225	0.342	0.322	0.451	0.219	0.351	0.558	0.572	-5.858	-0.030
Spearman's correlation	0.871	0.876	0.864	0.871	0.953	0.964	0.951	0.960	0.979	0.982	0.923	0.674
Mean error	2.133	2.174	1.659	1.666	-5.631	-5.631	-6.144	-6.145	-5.945	-5.946	-3.320	-3.322
Error std	1.385	1.114	1.394	1.100	4.075	2.728	4.141	2.800	2.109	1.795	24.557	9.011

Camera gauges are an important and promising method to estimate water stage and flow. There are two main sources of errors on this setup, mainly due to the photogrammetry process and during the water line detection. The focus of our study is to assess the potential of methods that can automatically extract water masks from images with minimum input. Challenging environmental conditions, especially light condition direct effect image segmentation, and water stage retrieve. Further, when using Grounding Dino combined with SAM, river bounding box detection was not always accurate, inducing SAM to produce no wrong water masks. By fixing points along the river channel, SAM produced river borders few pixels apart from where they should be, reflecting the overestimation of the water stage in Wesenitz station and underestimation for the other two stations. Not considering any photogrammetry errors, a perfect fit in the water stage could not be achieved once reference values are 15 minutes average values, while water stage calculated using images represent the water stage on the capture moment of a given image. Nonetheless, our approach could consistently track change in water level value, with emphasis in high water stage values, proving to be a valuable tool for flash flood monitoring.

Training of CNN demands powerful hardware and technical knowledge, a fact that can limit the use of these techniques especially for developing countries. Our findings showed that researchers can take advantage of pre-trained models and open-set models to accurately produce water segmentation, being this easily adapted and it can be used in less powerful hardware. Water stage results indicate that these models are notably useful in higher water stages, therefore, can be a powerful tool to increase hydrological networks to monitor floods and flash floods, with minimum to none input. These methods can represent a step forward for hydrologists towards a cheaper, reliable and scalable network of camera gauges that would help to increase resilience to extreme weather due to improved/densified monitoring and to narrow the data gap in hydrology.

4. Conclusions

In this study, we proposed an approach for water stage measurement combining deep learning and photogrammetry minimizing the labeling needed using video object segmentation and pre-trained models for water segmentation. Stationary images were collected in water stage measurement gauges, therefore, we assess the performance of produced masks to retrieve the water stage. We found that for stationary images, models achieved similar results in terms of water pixel accuracy and IoU when compared to traditional image segmentation networks, with difference in water IoU between -0.059 and -0.024 for STCN and SAM Dino, respectively. Using UAV images, STCN and SAM Dino achieved similar results, with water IoU higher than 0.6. Our findings suggest that when using STCN in non-stationary images, the choice of the first frame impacts the model performance. For SAM Dino, issues in segmentation on these images are caused by larger bounding boxes detected by Grounding Dino. Qualitative analysis of images from the camera gauge stations showed that STCN generally achieved a better result than SAM Dino and SAM using points, especially for the river borders. Nevertheless, STCN suffers with shadows, especially during summer, which can be caused by the lower frame rate and the sudden changes in illumination from frame to frame. Performance could be increased by extending STCN memory size to deal with sudden changes in image characteristics.

Regarding water stage, our findings shows that all methods were capable of tracking stage changes, with emphasis for high values. Furthermore, results indicate that STCN is a more suitable option for sub-hour monitoring, achieving best results for all tree camera gauges stations. Our work advances camera gauge technologies, reducing the effort to deploy this system towards large hydrological monitoring networks.

5. Acknowledgments

This research was partially funded by CNPq (308481/2022-4 and 305296/2022-1), and CAPES PrInt (p: 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 research has been partly funded by the BMBF (Federal Ministry of Education and Research; ID: 13N15542). We would like to thank the KIWA (Künstliche Intelligenz für die Hochwasserwarnung) project group for their support. The UAV data were thankfully provided by Eliisa Lotsari and captured within the BMBF project ID 57524996.

6. Appendix

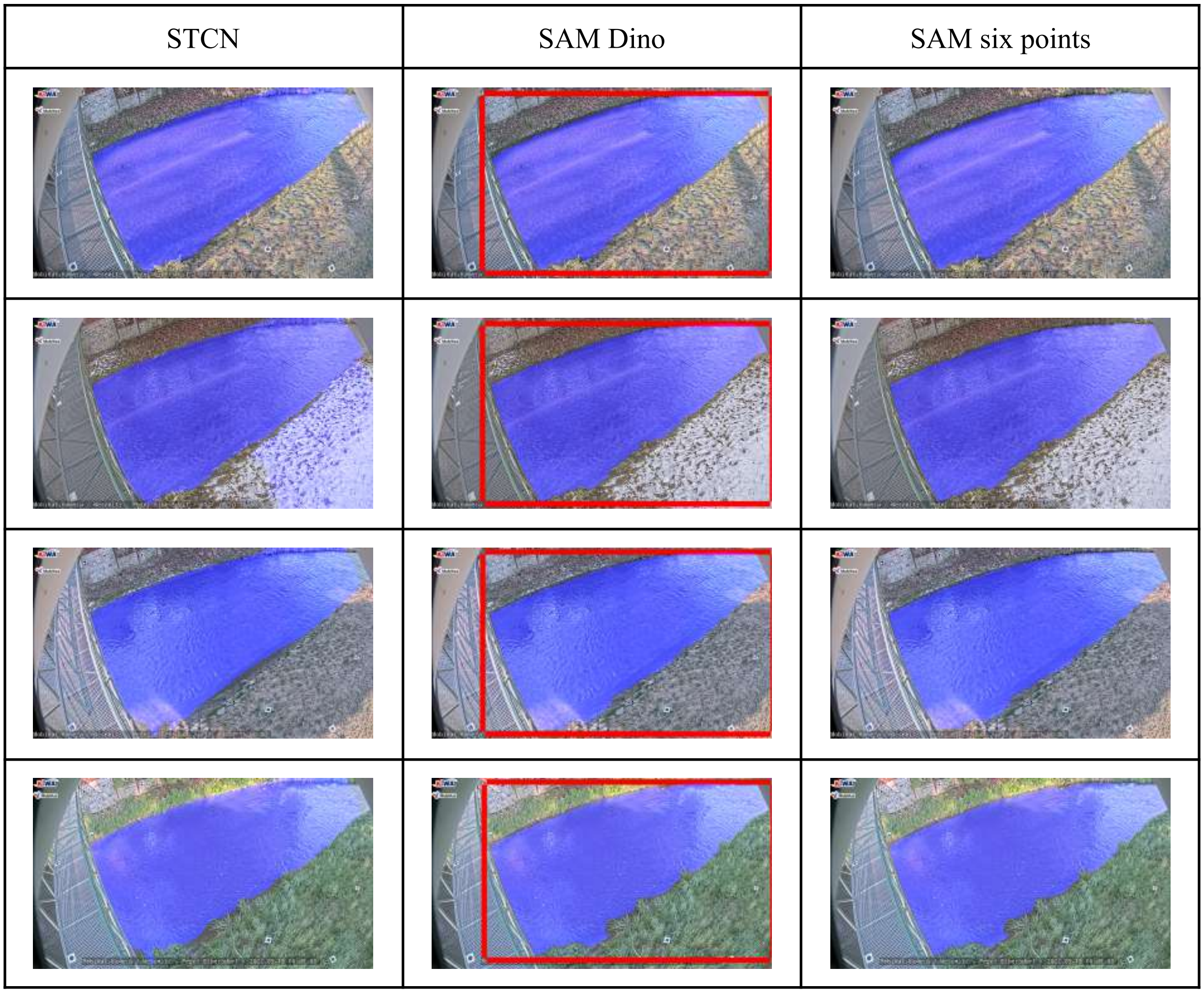


Figure A1: Example of masks and bounding box for the Elberdosf dataset. First row: 2021-12-03 10:45:00; second row: 2021-12-10 12:30:00, third row: 2022-03-12 15:00:00; last row: 2022-05-18 06:45:00

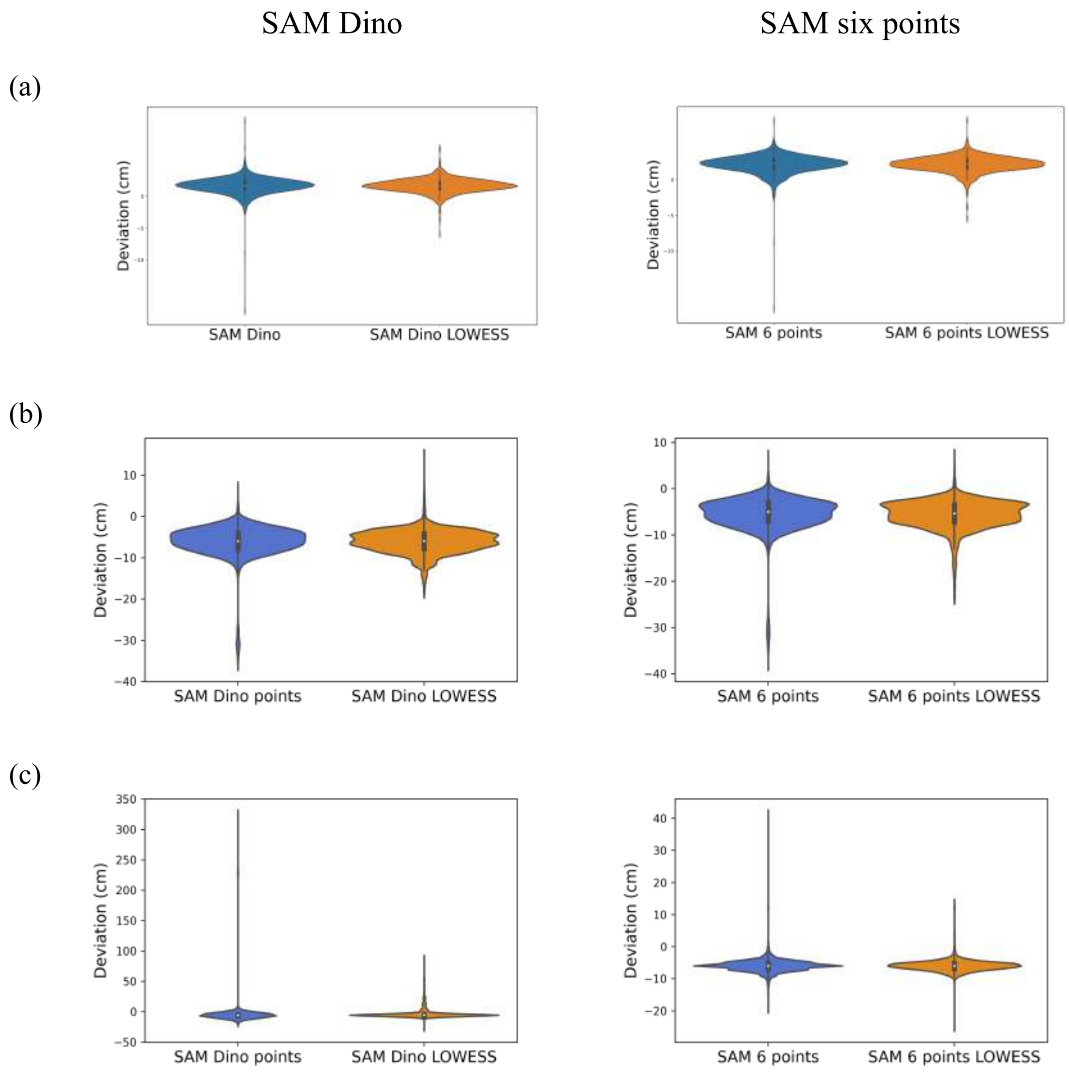


Figure A2: Water stage deviation using SAM Dino and SAM using points, with and without LOWESS regression, compared to reference data. In (a) Wesenitz; (b) Elbersdorf; and (c) Lauenstein stations.

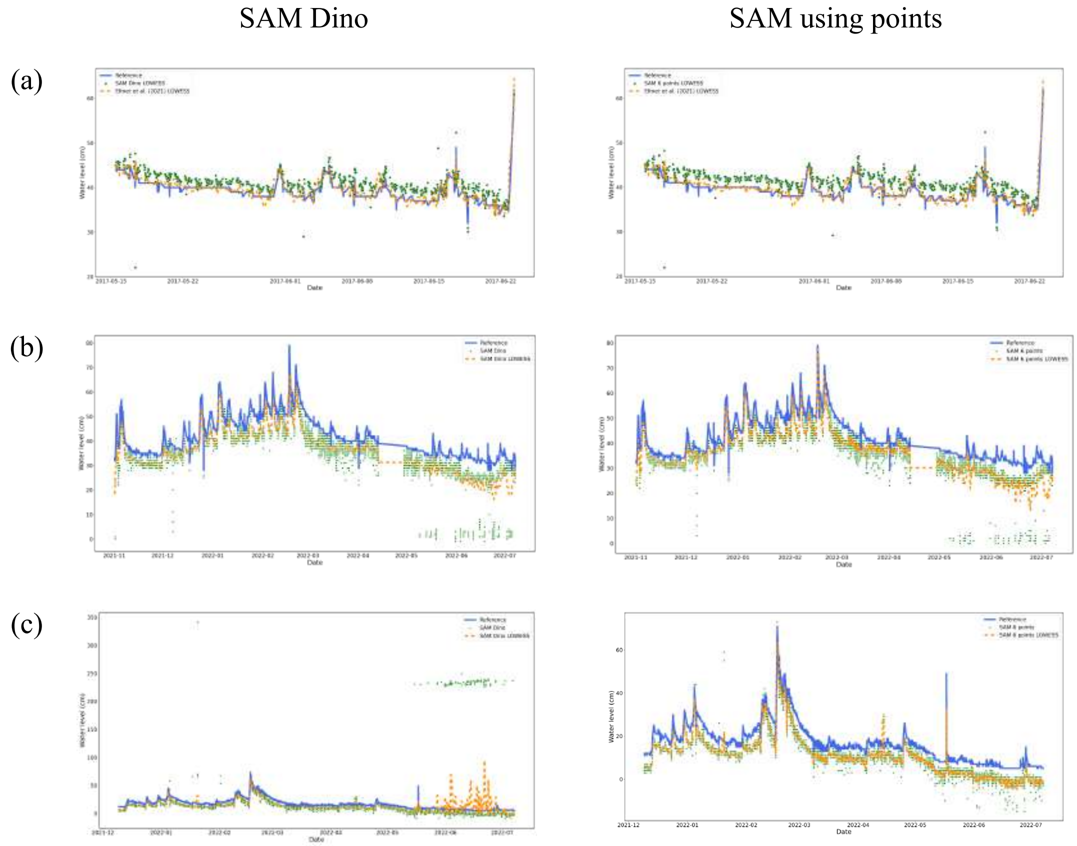


Figure A3: Water stage using SAM Dino and SAM using points, with and without LOWESS regression, compared to reference data. In (a) Wesenitz; (b) Elbersdorf; and (c) Lauenstein stations. For the Wesenitz, we show the results considering only the LOWESS regression and values from Eltner et al., (2021) for the same station.

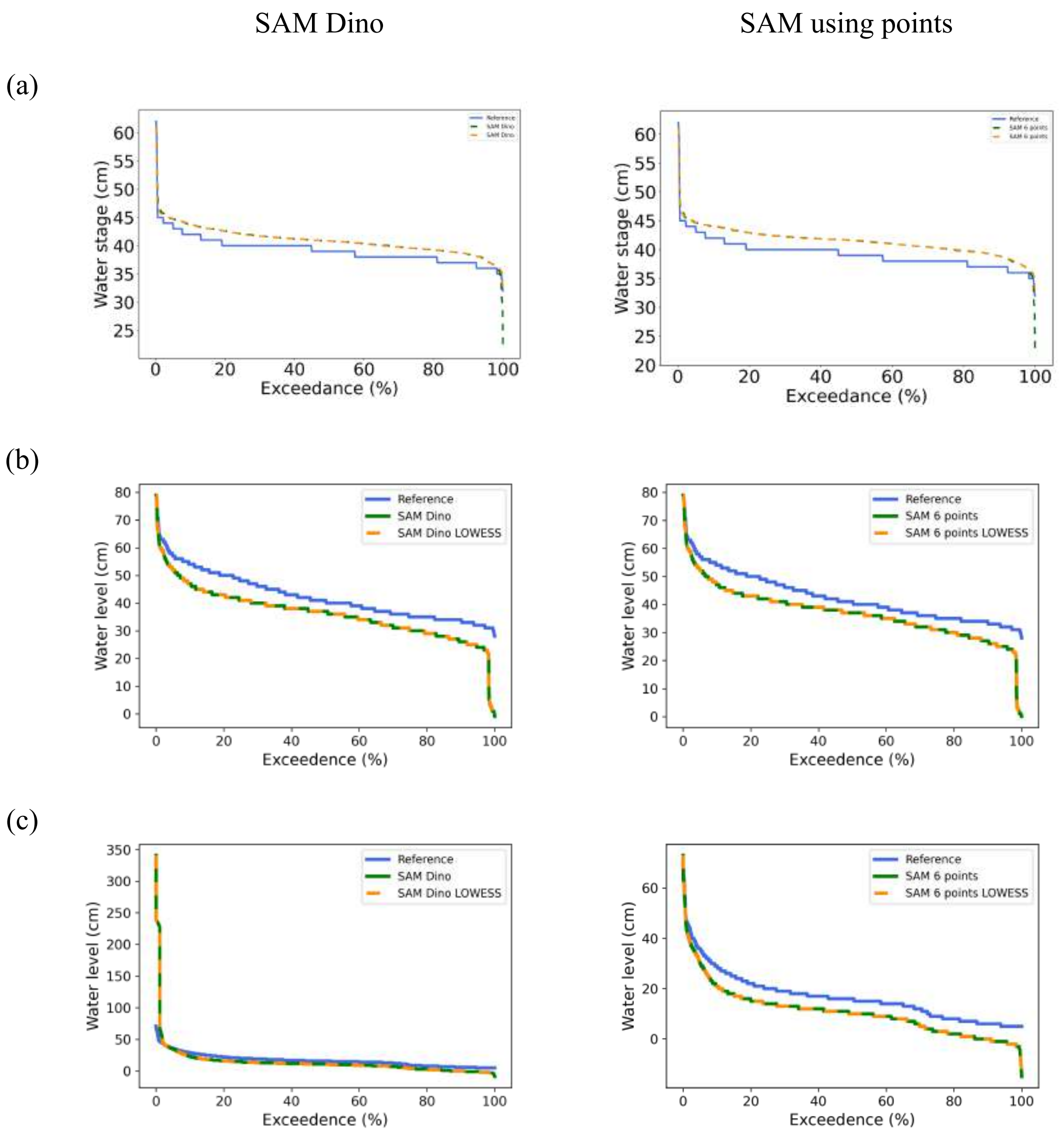


Figure A4: Exceedance curves using SAM Dino and SAM using points, with and without LOWESS regression, compared to reference data. In (a) Wesenitz; (b) Elbersdorf; and (c) Lauenstein stations.

(a)

(b)

(c)

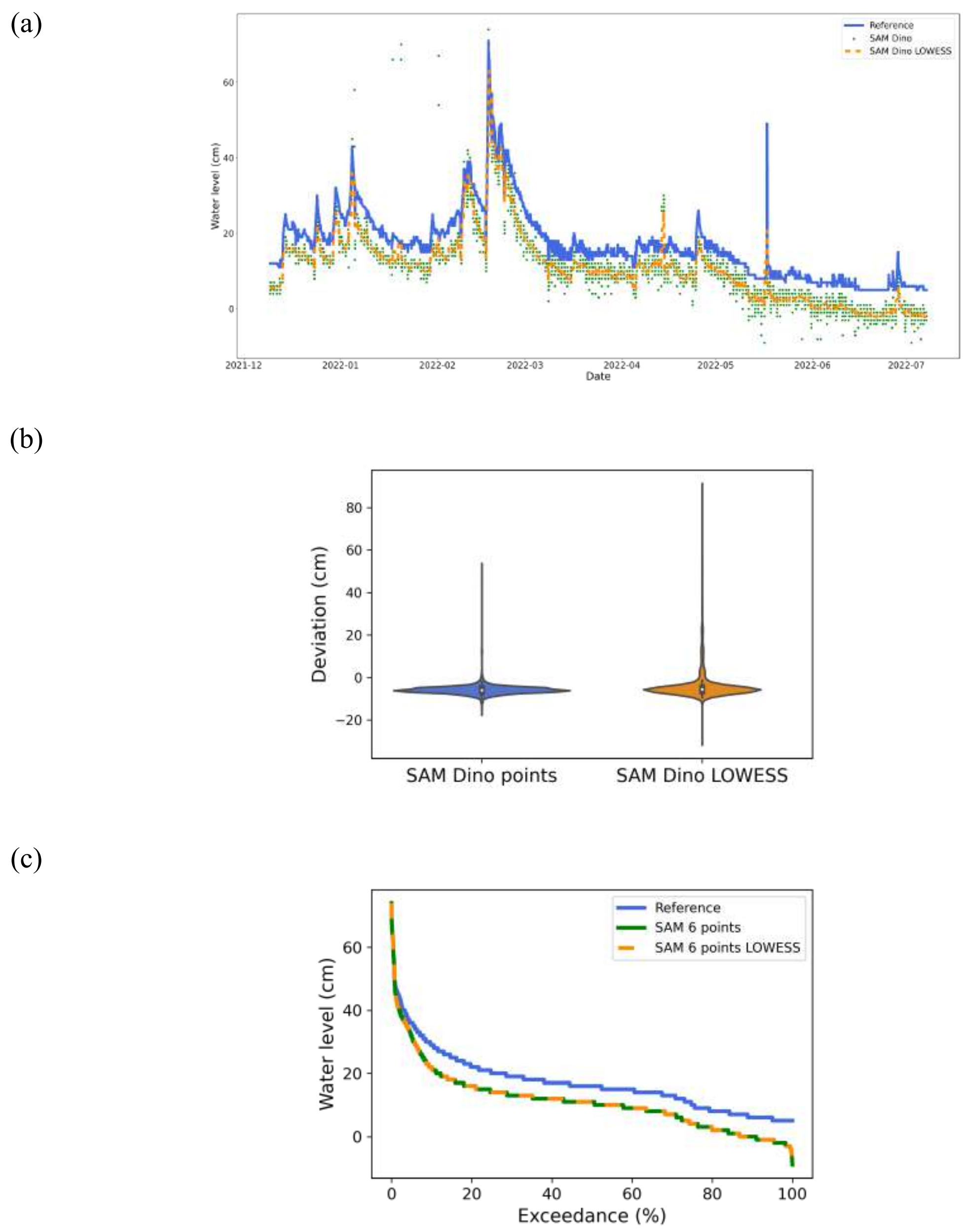


Figure A5: Results for Lauenstein using SAM Dino considering LOWESS regression for water stage lower than 200 cm. In (a) water level, (b) deviation, and (c) exceedance.

7. References

Filonenko, A., Wahyono, D. C., Seo, D., Jo, K. -H. (2015). Real-time flood detection for video surveillance. IECON 2015 - 41st Annual Conference of the IEEE Industrial Electronics Society, Yokohama, Japan, 004082-004085. https://doi.org/10.1109/IECON.2015.7392736.

Ling, F., Boyd, D., Ge, Y., Foody, G. M., Li, X., Wang, L., et al. (2019). Measuring river wetted width from remotely sensed imagery at the subpixel scale with a deep convolutional neural network. Water Resources Research, 55, 5631–5649. https://doi.org/10.1029/2018WR024136

Mullen, A. L., Watts, J. D., Rogers, B. M., Carroll, M. L., Elder, C. D., Noomah, J., et al. (2023). Using high-resolution satellite imagery and deep learning to track dynamic seasonality in small water bodies. Geophysical Research Letters, 50, e2022GL102327. https://doi.org/10.1029/2022GL102327

Vandaele, R., Dance, S. L., and Ojha, V. (2021). Deep learning for automated river-level monitoring through river-camera images: an approach based on water segmentation and transfer learning, Hydrol. Earth Syst. Sci., 25, 4435–4453. https://doi.org/10.5194/hess-25-4435-2021. (a)

Vandaele, R., Dance, S.L., Ojha, V. (2021). Automated Water Segmentation and River Level Detection on Camera Images Using Transfer Learning. In: Akata, Z., Geiger, A., Sattler, T. (eds) Pattern Recognition. DAGM GCPR 2020. Lecture Notes in Computer Science(), vol 12544. Springer, Cham. https://doi.org/10.1007/978-3-030-71278-5_17. (b)

Muhadi NA, Abdullah AF, Bejo SK, Mahadi MR, Mijic A. (2021). Deep Learning Semantic Segmentation for Water Level Estimation Using Surveillance Camera. Applied Sciences. 11(20):9691. https://doi.org/10.3390/app11209691.

Moy de Vitry, M., Kramer, S., Wegner, J. D., and Leitão, J. P. (2019). Scalable flood level trend monitoring with surveillance cameras using a deep convolutional neural network, Hydrol. Earth Syst. Sci. 23, 4621–4634. https://doi.org/10.5194/hess-23-4621-2019.

Chaudhary, P. (2018). Floodwater-estimation through semantic image interpretation, Technical University Munich, Munich, Germany.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2015). The Cityscapes Dataset, in: CVPR Workshop on The Future of Datasets in Vision, 7–12 June 2015, Boston, USA.

Oh, S. W., Lee, J. -Y., Xu, N., Kim, S. J. ((2019). Video Object Segmentation Using

Space-Time Memory Networks," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), pp. 9225-9234. https://doi.org/10.1109/ICCV.2019.00932.

Cheng, H. K., Tai, Y. -W., Tang, C. -K. (2021). Rethinking Space-Time Networks with Improved Memory Coverage for Efficient Video Object Segmentation. In: Ranzato M, Beygelzimer A, Dauphin Y, Liang PS, Vaughan JW, eds. Advances in Neural Information Processing Systems. Vol 34. Curran Associates, Inc. 11781-11794. https://proceedings.neurips.cc/paper_files/paper/2021/file/61b4a64be663682e8cb037d9719ad8cd-Paper.pdf

Peña-Haro, S., Carrel, M., Lüthi, B., Hansen, I., Lukes, R. (2021), Robust Image-Based Streamflow Measurements for Real-Time Continuous Monitoring. Frontiers in Water. 3. doi:10.3389/frwa.2021.766918

Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L. M., Shum, H. -Y. (2022). DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection. arXiv [csCV]. Published online 2022. http://arxiv.org/abs/2203.03605

Mazurowski, M. A., Dong, H., Gu, H., Yang, J., Konz, N., Zhang, Y. (2023). Segment anything model for medical image analysis: An experimental study, Medical Image Analysis. 89. https://doi.org/10.1016/j.media.2023.102918.

Chauveau, B. and Merville, P. (2023). Segment Anything by Meta as a foundation model for image segmentation: a new era for histopathological images,Pathology.https://doi.org/10.1016/j.pathol.2023.09.003.

Giannakis, I., Bhardwaj, A., Sam, L., Leontidis, G. (2024). A flexible deep learning crater detection scheme using Segment Anything Model (SAM), Icarus. 408. https://doi.org/10.1016/j.icarus.2023.115797.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., Zisserman, A. (2010). The pascal visual object classes (voc) challenge, IJCV. 88, no. 2, pp. 303–338, 2010

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/10.1016/j.jag.2023.103305)

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., Torralba. A. (2016). Semantic understanding of scenes through the ade20k dataset. arXiv preprint arXiv:1608.05442.

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

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

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

Lopez-Fuentes, L., Rossi, C., and Skinnemoen, H. (2017). River segmentation for flood monitoring, in: Proceedings of the IEEE International Conference on Big Data (Big Data), IEEE, 3746–3749, https://doi.org/10.1109/BigData.2017.8258373. a, b, c, d

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. https://doi.org/10.1016/j.jag.2023.103305.

Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P. H. S., Hospedales. T. M. (2018). Learning to Compare: Relation Network for Few-Shot Learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 1199-1208, doi: 10.1109/CVPR.2018.00131.

Feyjie, A. R., Azad, R., Pedersoli, M., Kauffman, C., Ayed, I. B., Dolz, J. (2020). Semi-supervised few-shot learning for medical image segmentation. arXiv. https://doi.org/10.48550/arXiv.2003.08462. Disponível em: <https://arxiv.org/abs/2003.08462>.

Liu, W., Zhang, C., Lin, G., Liu, F. (2020). CRNet: Cross-Reference Networks for Few-Shot Segmentation. arXiv. https://doi.org/10.48550/arXiv.2003.10658. Disponível em: https://arxiv.org/abs/2003.10658>.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Zhu, J., Zhang, L. (2023). Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set

Object Detection. arXiv. https://doi.org/10.48550/arXiv.2303.05499.

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. arXiv. https://doi.org/10.48550/arXiv.2304.02643

Banch, X., Wagner, F., Hedel, R., Grundmann, J., and Eltner, A. (2022). Towards automatic real-time water level estimation using surveillance cameras, EGU General Assembly, Vienna, Austria, 23–27 May 2022, EGU22-3225, https://doi.org/10.5194/egusphere-egu22-3225.

GENERAL CONCLUSIONS

The results of **Chapter 1** show the potential of deep learning approaches to increase the understating of the runoff, showing the capability of deep learning to segment water in very challenging scenarios. Results show the importance of considering label imbalance and uncertainty during the training of the deep learning approaches, with a significant impact in models performance. Moreover, presented results show that model ensembling produced better results compared to single models prediction. We also presented results with the impact of the spatial correlation among sample and model transferability to unseen images. By using the best model, we were able to measure, in spatial and time, the water pixel area on each rainfall simulation, directly quantifying the number of the ponds and its connectivity. Further, by comparing water area results with measure discharge we were able to identify different behavior in runoff generation.

In **Chapter 2**, we evaluated deep learning methods that use minimum to non labeled data in water segmentation of camera gauges images for water stage estimation, and uncrewed aerial vehicles (UAV). For the camera gauges, models achieved similar performance during water area segmentation. Qualitative analysis showed that segmentation tends to be worst close to the river shores; further, shadows can influence the performance of these methods. In the case of the UAV dataset, results are promising, although in this case of dynamic images these models should be carefully applied. Regarding the water stage measurement, results indicate that tested approaches can produce a good fit compared to reference water stage data, being able to capture changes in water stage. Our findings further suggest that results are better for high flow, indicating that these methods can be used as an ad hoc solution in areas prone to floods.

Overall, we can conclude that the thesis meets its general objectives. Our results show the potential of deep learning and photogrammetry as tools in soil science and hydrology, unlocking a new frontier in the observation, measurement, and monitoring of environmental systems. The development and results of the chapters provides tools that can lead to a better understanding of some components of the hydrological cycles (runoff generation and streamflow). Limitations of this thesis should be acknowledged to provide direction for future works. Firstly, a greeted dataset should be tested, as well data augmentation strategies and newer deep learning methods (i.e. transformers based models) for the water ponding segmentation on the first chapter. Still in the first chapter, we considered the rainfall simulation plot area as a plane, therefore, only providing a rough estimation of the water area. In the future, plot water areas can be intersected with high resolution terrain models to produce better water area measurements, furthermore, water volume estimation. These terrain models can be also used with RGB frames to train and inference deep learning approaches, possibly leading to better results in terms of water segmentation. Finally, in the second chapter, the most important limitations were on the intersection between 2D water coordinates and the 3D model. Alignment between 2D and 3D spaces requires the use of the ground control points, therefore, these points should be also detected in all images. Misaligned leads to high error in water stage measurment. Regarding the 3D model, changes on the river bed and surrounding lead to the need of updating the 3D model. A further step to deal with these limitations is the use of two or more cameras to utilize stereo-photogrammetry.