Serviço Público Federal
Ministério da Educação
Fundação Universidade Federal de Mato Grosso do Sul

THALES SHOITI AKIYAMA

ARTIFICIAL INTELLIGENCE AND REMOTE SENSING IN URBAN HYDROLOGICAL APPLICATIONS

Campo Grande, MS
Março – 2022

Fundação Universidade Federal de Mato Grosso do Sul
Faculdade de Engenharias, Arquitetura e Urbanismo e Geografia
Programa de Pós Graduação em Tecnologias Ambientais

THALES SHOITI AKIYAMA

ARTIFICIAL INTELLIGENCE AND REMOTE SENSING IN URBAN HYDROLOGICAL APPLICATIONS

Tese apresentada para obtenção do grau de Doutor no Programa de Pós-Graduação em Tecnologias Ambientais da Fundação Universidade Federal de Mato Grosso do Sul, área de concentração: *Saneamento Ambiental e Recursos Hídricos*.

Orientador: Prof. Dr. José Marcato Junior
Coorientador: Prof. Dr. Wesley Nunes Gonçalves e Jun.-Prof.Dr.-Ing. Anette Eltner

Aprovada em: 03/03/2022

Banca Examinadora

José Marcato Junior
Presidente

Ana Paula Marques Ramos
Universidade do Oeste Paulista - UNOESTE

Jefersson Alex dos Santos
Universidade Federal de Minas Gerais - UFMG

Jonathan de Andrade Silva
Universidade Federal de Mato Grosso do Sul - UFMS

Paulo Tarso Sanches de Oliveira
Universidade Federal de Mato Grosso do Sul - UFMS

Campo Grande, MS
Março – 2022

DEDICATÓRIA

Aos meus pais Paulo e Regina, pela compreensão, carinho, amor e todo o apoio que sempre me deram.

Ao meu querido irmão Fábio, pela parceria, amizade e companheirismo.

A Deus, por estar sempre ao meu lado e me guiar nos mais adequados rumos.

AGRADECIMENTOS

Ao professor Dr. José Marcato Junior, por todo apoio, paciência, orientação, ajuda, pelos ensinamentos, ao qual me auxiliaram em todo este caminho e aprendizado.

Aos amigos de faculdade, beisebol, do laboratório de Geomática e da Alemanha pela ajuda, pelas atividades desenvolvidas e momentos de diversão.

À minha namorada Giovana, por me compreender sempre e me apoiar em todas as minhas decisões.

Aos meus coorientadores e supervisores professor Dr. Wesley Gonçalves Nunes e professora Dra. Anette Eltner, pelo auxílio em inúmeras atividades com o intuito de melhorar a nossa pesquisa.

À UFMS, pelo auxílio de equipamentos e locais de trabalho para o desenvolvimento da pesquisa.

À FUNDECT (Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul), ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo financiamento dos projetos (p: 433783/2018-4; 303559/2019-5); e à CAPES pela bolsa de doutorado (Código de financiamento 001) e doutorado sanduíche.

RESUMO

As inundações urbanas despertam uma grande preocupação, pois ocasionam perdas materiais, econômicas, ambientais, e nas piores situações resultam em mortes de seres vivos. Para lidar com essas questões, abordagens preventivas devem ser implementadas para minimizar tais impactos. Embora existam pesquisas procurando solucionar a questão da inundação em áreas urbanas, são escassos os trabalhos relacionados a técnicas de *Deep Learning* (DL) para monitorar recursos hídricos. Devido a esta problemática, este trabalho investiga e propõe métodos baseado em DL para monitoramento hídrico. Primeiramente, analisou-se a performance do modelo em segmentação semântica SegNet na delimitação de corpos da água em imagens RGB, apresentando uma acurácia acima de 97%, mostrando que o modelo é adequado para a segmentação de água. Em seguida, introduziu-se uma abordagem automatizada em medições do nível da água combinando DL e fotogrametria, apresentando correlações entre as medidas de referência e a abordagem proposta de 93%. Também analisou-se diferentes configurações para a SegNet avaliando a performance dos modelos para tarefas de generalização na segmentação de diferentes superfícies aquáticas, mostrando que técnicas como *transfer learning* e *fine-tuning* melhoraram os resultados. Além disso, mostrou-se que há uma redução na acurácia da segmentação quando se reduz a quantidade de imagens rotuladas utilizadas no treinamento da rede. Por fim, analisou-se a performance do modelo *Space-Time Correspondence Network* (STCN) na segmentação da água baseado em estruturas de vídeos, no qual os resultados mostram que o modelo é acurado em delimitar os contornos de um corpo da água em diferentes situações. A maior contribuição deste estudo é a otimização das informações relativas a um corpo de água utilizando técnicas diferentes dos sistemas tradicionais de medição.

Palavras-chave: Aprendizado profundo, Segmentação semântica, Redes neurais convolucionais, Recursos hídricos, Inundações.

ABSTRACT

Urban flooding is a big concern because it causes material, economic, environmental losses, and in the worst situations results in the death of living beings. To deal with these issues, preventive approaches must be implemented to minimize such impacts. Although there are researches seeking to solve the issue of flooding in urban areas, there are few works related to Deep Learning (DL) techniques for monitoring water resources. Due to this problem, this paper investigates and proposes DL-based methods for water monitoring. First, the performance of the SegNet semantic segmentation model in delineating water bodies in RGB images was analyzed, presenting an accuracy above 97%, showing that the model is suitable for water segmentation. Next, an automated approach was introduced in water level measurements combining DL and photogrammetry, showing correlations between reference measurements and the proposed approach of 93%. We also analyzed different configurations for SegNet evaluating the performance of the models for generalization tasks in segmenting different water surfaces, showing that techniques such as transfer learning and fine-tuning improved the results. Furthermore, it was shown that there is a reduction in the segmentation accuracy when the number of labeled images used in the network training is reduced. Finally, the performance of the Space-Time Correspondence Network (STCN) model in the segmentation of water based on video structures was analyzed, which the results show that the model is accurate in delimiting the contours of a body of water in different situations. The major contribution of this study is the optimization of information concerning a body of water using techniques different from traditional measurement systems.

Keywords: Deep learning, Semantic segmentation, Convolutional neural networks, Water resources, Flooding.

SUMÁRIO

1. GENERAL CONTEXT

1.1. INTRODUCTION

Surface runoff is a natural process that is identified as the result of saturated soil being unable to absorb precipitation, thus causing flooding. For instance, surface runoff often reaches areas prone to flood risk, which are those located on lower and flatter sites. This natural event usually occurs due to the dynamism of nature and is the result of the integration of meteorological, hydrological, and human phenomena, and when it is not properly supervised, it can cause serious consequences.

Li et al. (2016) report that the frequent occurrence of flood disasters is caused by: (1) expansion of cities into areas with high flood risk; (2) the drainage development system is slower than the development of cities; (3) the increasing of impermeable areas and (4) the increase of population density results in an increase of flood vulnerability in urban areas.

Urban flooding has aroused great concern due to cities being in constant growth and development, besides being where the population is most concentrated. There are numerous types of losses that such disasters can cause, such as material, economic, environmental, and even human. Yin et al. (2014) mention that catastrophic floods arising from various sources (fluvial, pluvial, coastal) caused, in July 2011, 79 deaths, economic loss of approximately 1.86 billion dollars, and social impacts in urban areas of Chinese cities.

The urbanization process associated with inappropriate planning has several socio-environmental impacts, affecting the environment and society's quality of life. Mirza (2003) reports that poor urban planning associated with insufficient adaptation measures (structural and non-structural) further contribute to urban flooding. Factors such as anthropic intervention, inadequate drainage systems, climate change, inappropriate urban land use and occupation, and countless other conditions not only lead to an intensification of flooding but also an increase in the surface runoff on impervious surfaces, exposure to risks and diseases, occupation of flood-prone areas, and other threats.

Flood risks cannot be totally avoided, hence it is essential that monitoring approaches and prevention measures need to be developed in order to deal with these events and minimize their impacts, such as adequate urban planning, effective engineering projects for water catchment, policies for soil occupation, environmental education, among others. One hydrological observation that is considered important to monitor flooding is the water level. There are different ways to obtain this measurement, such as systems based on pressure gauges and floats, or traditional measuring rulers. Although these are alternatives, extreme events such

as high precipitation or adverse weather conditions often cause the loss of such instruments. In addition, it also makes difficult for people to go on site to take readings and these measurements can be expensive due to maintenance requirements (Morgenschweis, 2010).

In order to find more solutions as alternatives to flooding, many flood forecasting models have been developed and implemented. Li et al. (2016) integrated the Urban Flood Simulation Model (UFSM) and Urban Flood Damage Assessment Model (UFDAM) to propose a framework for analyzing the risks and benefits of flood control measures in urban areas. Chen, Hill, and Urbano (2009) developed and tested the Geographic Information System-based Urban Flood Inundation Model (GUF-IM) to simulate urban flooding.

Although there are many researches seeking to solve the issue of flooding in urban areas, there are few works related to using Deep Learning (DL) techniques to monitor water resources and their behavior. Moreover, few cities have real-time flood alert systems, a fact considered highly relevant because this event is usually caused by high-intensity precipitation in a short period, affecting the population suddenly. Therefore, the main aim is to investigate and propose DL methods for monitoring water bodies. Thus, it is hoped that the results of DL can optimize in time and accuracy the traditional monitoring systems as well as being a supplement to the development of a real-time flood warning system.

1.2. BACKGROUND AND PROBLEM STATEMENT

According to Lecun, Bengio, and Hinton (2015), DL is a sub-branch of artificial intelligence that allows computational models composed of multiple processing layers to learn data representations with various levels of abstraction. It is a method that has improved state-of-the-art in voice recognition, object detection, and its implementation is increasing in other branches, as an example, in the medical field for drug recognition (Gawehn, Hiss and Schneider, 2016), electronic games, internet companies (Google, Baidu, Facebook), among other areas.

In order to evaluate data using computational resources, one alternative is the interaction of DL and remote sensing. The integration of these two areas is a reason to be considered, given that remote sensing seeks to obtain information from the Earth's surface without direct contact with the object of study, and DL makes it possible to obtain information with high accuracy and speed. Zhu et al. (2017) report the interaction of DL and remote sensing, and Osco et al. (2021) present a review on the deep learning field applied in Unmanned Aerial Vehicle (UAV) remote sensing, showing that DL is a powerful tool to automate the acquisition of geoinformation. Figure 1 was generated from word processing returning the most topics linked to "deep learning", "remote sensing" and "water", thus generating a word cloud. Google Scholar, Web of Science and Scopus were the scientific databases used. Analyzing the words

contained within Figure 1, it is possible to note how these areas can relate to each other in order to produce new studies, thus encouraging the development of new methodologies related to environmental monitoring, in our case focused on water resources.

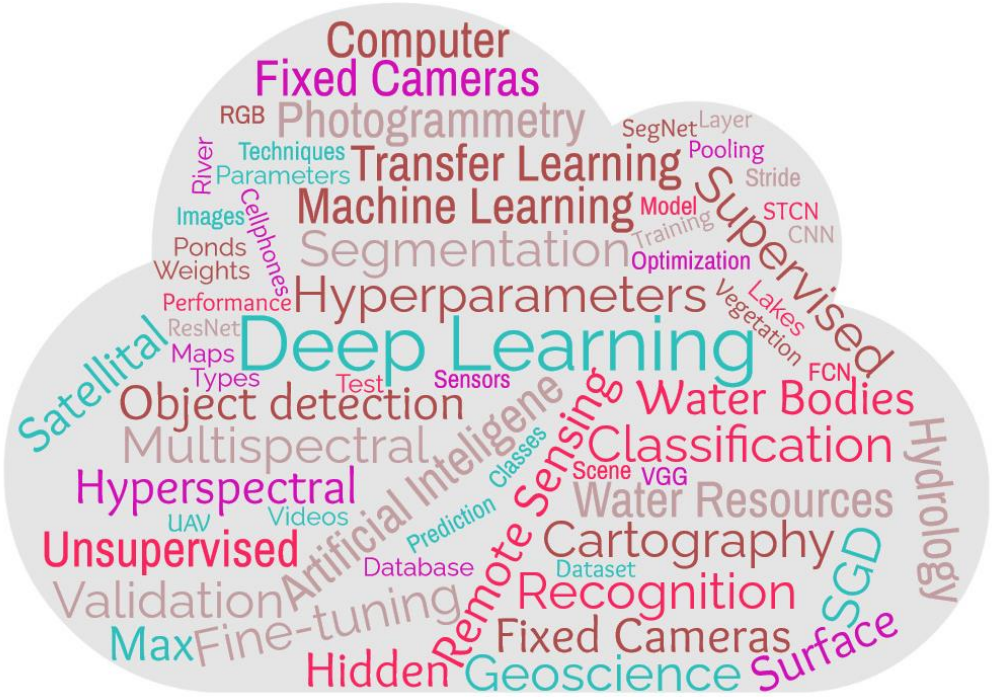


Figure 1. Related themes to deep learning, remote sensing and water.

In the context of monitoring water resources based on remote sensing techniques, whose advantage is that it does not require the installation of equipment directly in the water, the development of image and video-based methods has been growing over the last years. Eltner et al. (2019) proposed a method to measure the flow velocity and flow rate of rivers in an automated way using images obtained by ground sensors and UAV. Ridolfi and Manciola (2018) used UAV to determine water levels in rivers, seeking to improve existing hydraulic models as well as to aid strategies in case of floods or inundations. Detert et al. (2017) combined the reconstruction of an underwater topography based on Structure from Motion (SfM) with tracking algorithms to determine the velocity of the flow based on UAV and thus calculate the flow rate. All these studies used Remote Sensing techniques, but without the use of DL.

There are some studies involving artificial intelligence to monitor water resources. For instance, Acharya, Subedi, and Lee (2019) used machine learning algorithms to identify water surfaces from Landsat 8 images. However, this research did not use DL-based approaches and it only sought to locate water in images, not using the results to find information regarding water resources. Assem et al. (2017) conducted research applying DL algorithms seeking to derive an appropriate long-term methodology to predict flow and water stage parameters. Although this study differs from the first one because it applies DL-based techniques, only

parameters of some water-related information were obtained. Pan, Xiong, and Gui (2018) used CNN techniques to determine the water level in real-time using a surveillance camera system. This last approach is very interesting because it is a new alternative for measuring the water level. Nonetheless, the DL in this approach is used for the readings of the measuring ruler from the surveillance camera system. Therefore, our study differs from this approach precisely because it seeks new ways to use DL-based methods to monitor and obtain information about water bodies.

Classification, segmentation, and object detection are some methods of interpretation that can be generated by combining DL and remote sensing. To achieve the best results for such methods, different approaches must be performed and tested. For instance, Zhang et al. (2017) present the relationship between training dataset sizes, input data size, and depth of a network to memorize the training dataset parameters fully. Çayir and Navruz (2021) mention that dataset size and quality are factors that most affect accuracy in DL research. Complementing this last quote, Linjordet and Balog (2019) report that a very large dataset takes much longer to train and validate the data. These factors are important because there are numerous challenges for building a robust and adequate database, such as obtaining a large amount of data (images, videos), high-quality annotated labels, computational resources, the adequate architecture of the network models, among other configurations. These justifications lead us to develop and apply different methodologies and techniques to target the best approaches for integrating DL and remote sensing.

The use of DL in remote sensing to monitor water resources tend to have several advantages, such as: assisting in the improvement of existing hydraulic models as well as obtaining similar or even superior performance; arise as an alternative to replace measurements made by humans; the use of remote sensors to obtain information from the land surface in inaccessible and dangerous locations; evaluate if economically there is a lower cost compared to traditional methodologies. Such arguments justify the development of this thesis, thus seeking new methods to support a real-time flood warning system and thus helping the cities to become more intelligent in decision-making.

1.3. GOALS

Based on these reasons, this research aims to fill the existing gaps between DL-related studies on remotely sensed data. This section is divided into specific and general goals. First, it is assumed that it is possible to segment rivers contained in images obtained from fixed sensors. From the segmentation results, it is assumed that it is possible to estimate the water level using the river contours. Next, different techniques (number of training images, transfer learning,

fine-tuning, different datasets, among others) related to model performance in segmentation were performed in order to succeed on generalization tasks in delineating water bodies. In this case, it was assumed that it is possible to create a model capable of performing this generalization task. Finally, a video-based approach was investigated because it differs from others that use images. Thus, it was assumed that it is possible to segment water bodies using this structure of videos reducing the need for large labeled datasets.

1.3.1. General Goal

- Investigate and propose DL-based methods for water monitoring in support of a real-time flood and flood warning system.

1.3.2. Specific Goals

- Evaluate the performance of DL techniques for segmenting water from images obtained by ground cameras;
- Analyze different approaches on the performance of DL models to generalization tasks;
- Investigate the integration of DL-based image segmentation with photogrammetric techniques on water stage estimation;
- Investigate the performance of object segmentation DL techniques on video files.

1.4. MAIN CONTRIBUTIONS

This research sought to investigate and propose the integration of DL and remote sensing-based methods for monitoring water resources. The integration of these areas applied to the monitoring of environmental studies has been growing over the years with the advancement and development of technology, mainly due to the high performance that computer resources are offering. Figure 2 presents the number of papers published by year considering Scopus and Web of Science scientific bases. The keywords used to filter these results were “deep learning”, “remote sensing” and “water”.

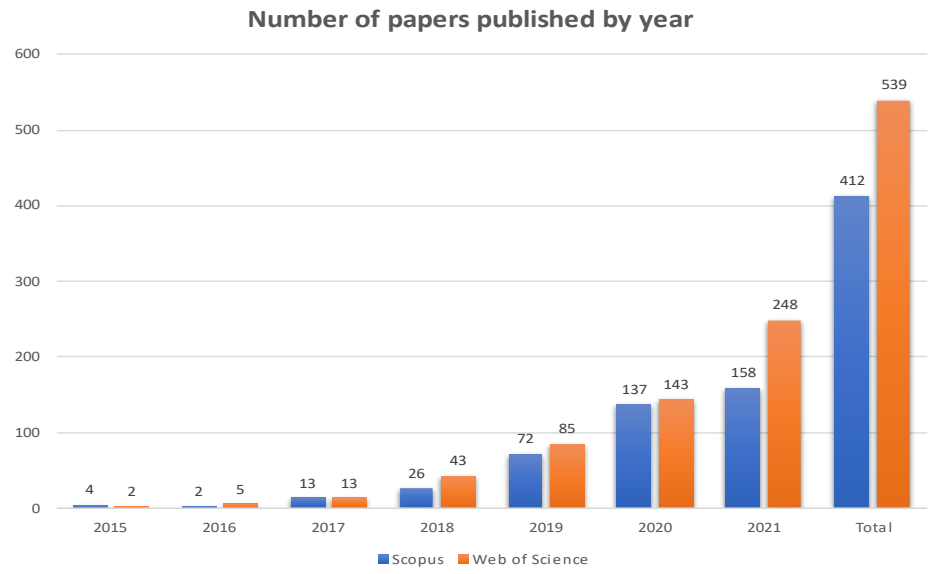


Figure 2. Number of papers related to deep learning, remote sensing and water published over the years.

Analyzing Figure 2, it is possible to check how the number of research related to these areas increased over the years. The development and refinement of the thesis also occurred gradually. The first paper, called "Deep learning applied to water segmentation", was published in 2020, which consisted of studying the DL techniques and methods that were arising that year and applying them to remotely sensed images. The second and third paper, named “Using deep learning for automatic water stage measurements” and “Evaluating different deep learning models for automatic water segmentation”, were published in 2021. It is possible to notice that this year more papers were published, showing the interest which these areas have been generating in several communities. In addition, direct applications to obtain different information in the environmental field were also aimed, such as the automated water level estimation that we have been able to produce. Figure 3 and Figure 4 show the number of papers using the keywords “deep learning”, “remote sensing” and “water level or water stage”, indicating the scarcity of studies relating such areas and how they are emerging recently. In addition, when the keyword “water level” is changed to “water stage”, the numbers of papers decrease, indicating 0 paper in Scopus and 30 papers in Web of Science.

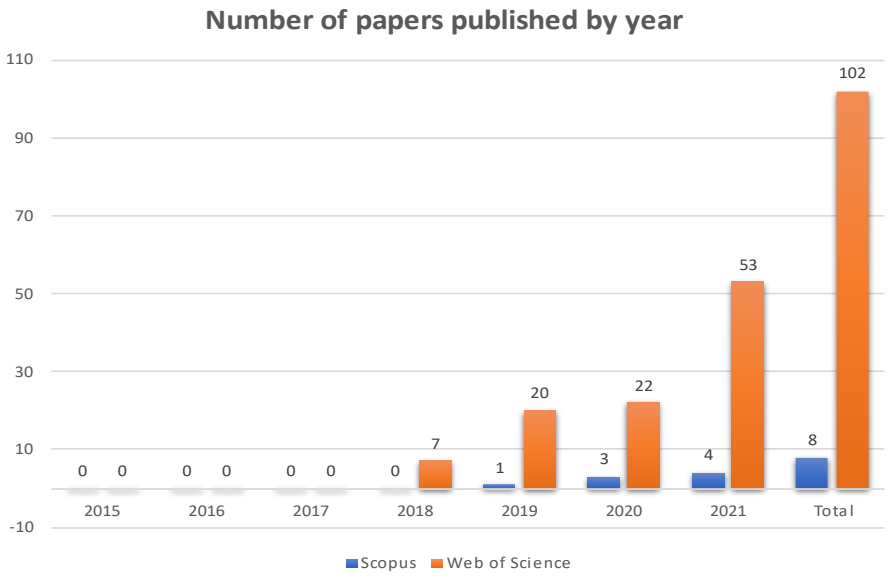


Figure 3. Number of papers related to deep learning, remote sensing and water level published over the years.

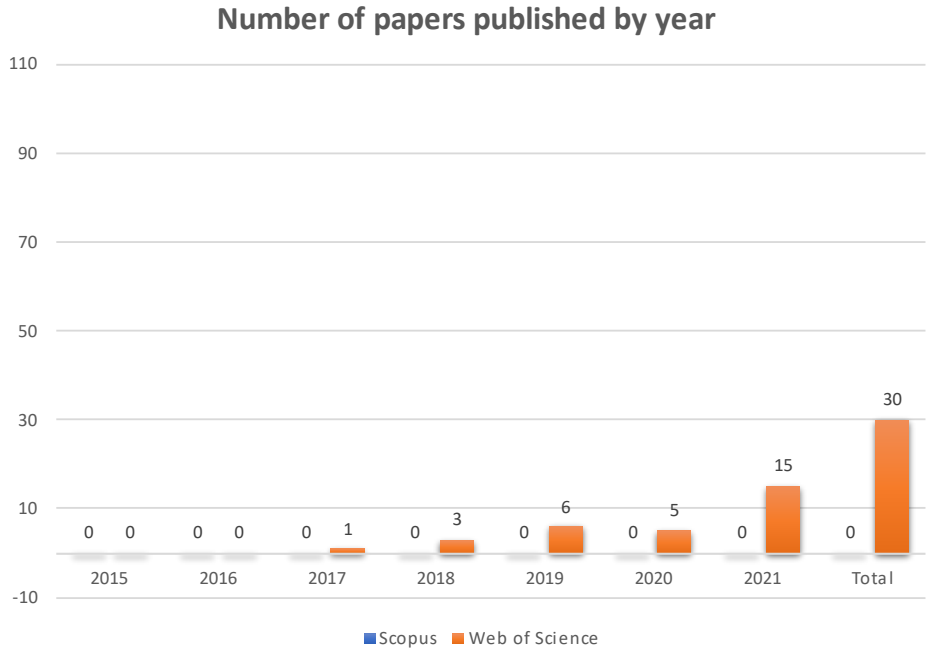


Figure 4. Number of papers related to deep learning, remote sensing and water stage published over the years.

Finally, the last manuscript will be submitted so that it can be accepted and subsequently published. The name is "Semantic segmentation of water considering video structures" and introduces an approach based on Space-Time Correspondence Network (STCN), considered a state-of-the-art in object semantic segmentation for videos. The original paper introducing this approach was published in 2021 (Cheng, Tai and Tang, 2021), emerging as a new alternative for environmental studies relating DL in video structures.

Considering the justifications represented by the advancement of articles integrating the fields of DL and remote sensing, besides the fact that there are few researches related to

water monitoring, our research aimed to use the semantic object segmentation methods considered the state of the art in mid-2020s (SegNet and FCN for images) and 2021 (STCN for videos) aimed applications that include ways to obtain information from a body of water. In this way, we were able to achieve the goals of trying to minimize the gap that exists between such areas.

The data acquisition, such as images and videos, was done with RGB sensors. This is of great importance because this kind of sensor is more cost-effective than other types, such as multispectral and hyperspectral sensors. Therefore, a database was built consisting of images and videos containing different types of rivers, lakes, ponds, streams. The differential of this dataset is the various ways of acquisition, such as fixed and moving sensors, mobile, UAV, surveillance systems, among others. With the distribution and sharing of this dataset, it is hoped that the scientific, computational, and other types of interested communities can benefit and develop new methodologies for monitoring water resources.

Regarding DL methodologies, numerous techniques and approaches have been used and tested. Semantic object segmentation techniques in images and videos to delineate the contours of a body of water, the best settings to obtain the most accurate results, fine-tuning, and transfer learning in order to perform generalization tasks. As the research progressed, it was possible to get good results for identifying and delineating water in images and videos to facilitate other studies that pursue the same goal.

From the results obtained from the segmentation of water bodies, it was possible to integrate them with photogrammetry techniques, thus generating water level automated. This method differs from traditional measurement approaches, in which the installation of in site equipment is required to perform the readings. This method tends to be a revolution for image-based approaches, being an innovative source of information for society in a way that the advancement of technology assists more and more in different daily tasks. Internet of Things (IoT) is an example of the rise of science and technology, emerging as an option for the progress of this research.

The major contribution of this study is the optimization of information compared to traditional methodologies to obtain information related to water. It is expected that other types of information regarding a body of water can be obtained, and new alternatives can be studied in the future in order to integrate and develop new types of systems. This research is not intended to replace standard measurement techniques but rather to seek new solutions and innovations in order always to support society and the progress of science.

1.5. ORGANIZATION OF THE THESIS

The Thesis is organized into several articles, presented as chapters. **Chapter 2** introduces the first published paper presenting the performance of the DL methods to segment water bodies. The results obtained from the segmentation performed for the river were accurate, thus looking for new methodologies to apply these results. In this way, **Chapter 3** provides an approach that combines DL and photogrammetric techniques for automatic water stage measurements. Using the borders obtained from the segmentation of the river from Chapter 2, it was possible to develop a different water stage estimation approach compared to traditional measurement systems. As these previous researches were limited to the same river, we tried to develop new DL models in order to achieve high performance in generalization tasks in segmenting different water bodies. Integrating the segmentation from the previous Chapters with different approaches (transfer learning, fine-tuning, size of training dataset) and image sets of new water bodies, two papers were developed - **Chapter 4** and **Chapter 5**. The first one presents an investigation of different strategies for automatic water segmentation considering images from different areas. The latter aims to analyze the influence of the number of images used for training on model performance. Finally, in order to seek applications for other types of remotely sensed data, a model for object segmentation based on video was investigated. Video-based approaches are emerging as a new alternative in interpreting DL tasks. It is hard to find studies focusing on the segmentation of water bodies consisting of video data. **Chapter 6** presents a DL-based approach for segmentation of water considering video structures. The complete structure of the thesis is shown in Figure 4, presenting a conceptual map of the thesis and how the articles were developed and integrated as the research progressed.

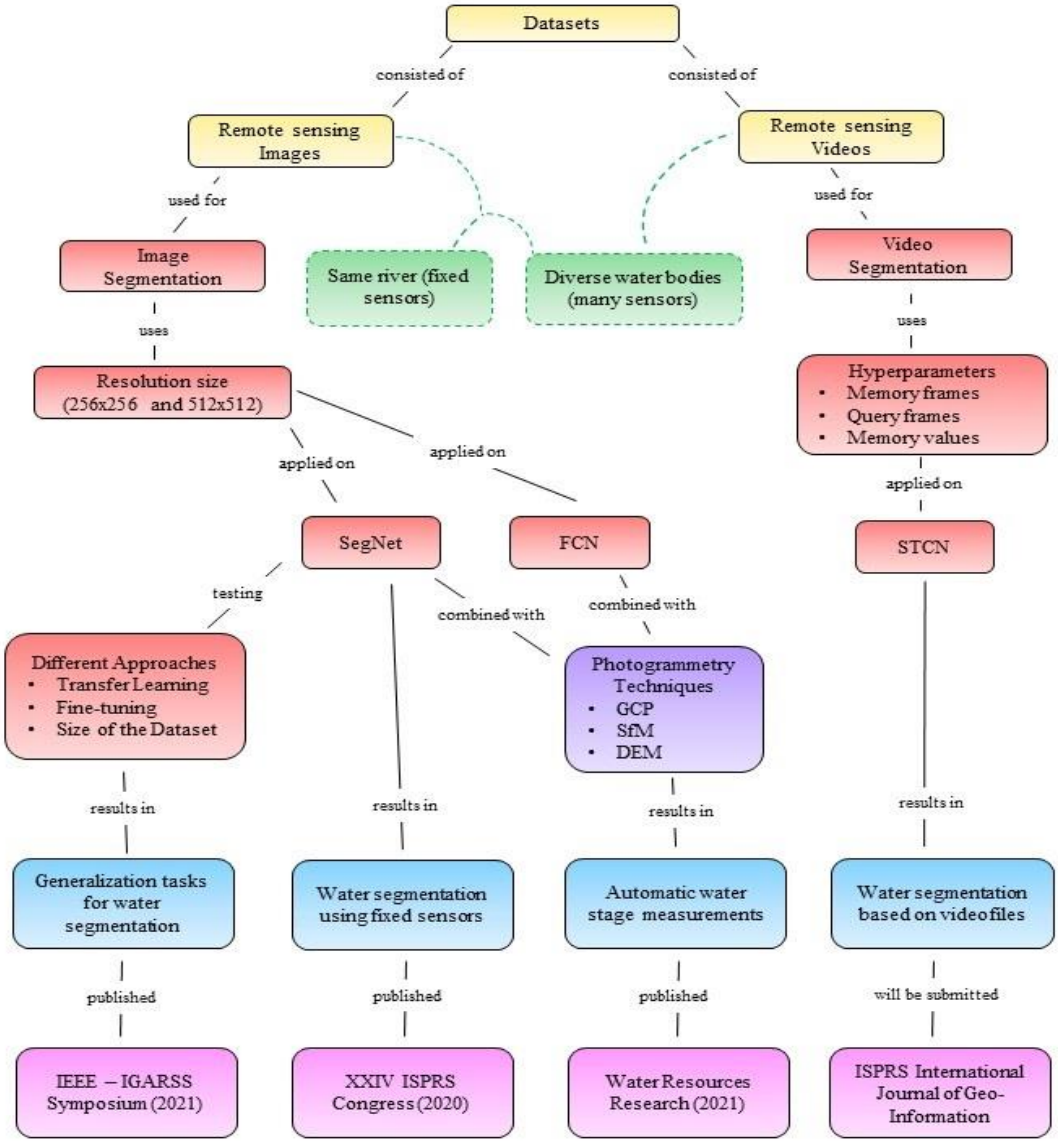


Figure 4. Conceptual map of the thesis. In yellow, the datasets. In green, how the datasets were divided. In red, the DL techniques, methods and approaches. In purple, the photogrammetry techniques. In blue, the main contributions of each research. In pink, the related publications associated with this thesis.

2. DEEP LEARNING APPLIED TO WATER SEGMENTATION

The first paper called “Deep learning applied to water segmentation” was published in the journal “The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B2-2020 - XXIV ISPRS Congress (2020)”. It is referenced as “Akiyama, T. S., Junior, J. M., Gonçalves, W. N., Bressan, P. O., Eltner, A., Binder, F., & Singer, T. (2020). Deep learning applied to water segmentation. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 1189-1193”.

ABSTRACT

The use of deep learning (DL) with convolutional neural networks (CNN) to monitor surface water can be a valuable supplement to costly and labor-intense standard gauging stations. This paper presents the application of a recent CNN semantic segmentation method (SegNet) to segment river water in imagery acquired by RGB sensors automatically. This approach can be used as a new supporting tool because only a few studies are using DL techniques to monitor water resources. The study area is a medium-scale river (Wesenitz) located in the East of Germany. The captured images reflect different periods of the day over approximately 50 days, allowing for the analysis of the river in different environmental conditions and situations. In the experiments, we evaluated the input image resolutions of 256 x 256 and 512 x 512 pixels to assess their influence on the performance of river segmentation. The performance of the CNN was measured with the pixel accuracy and IoU metrics revealing an accuracy of 98% and 97%, respectively, for both resolutions, indicating that our approach is efficient to segment water in RGB imagery.

2.1. INTRODUCTION

It is crucial that measures must be adopted to maintain the safety of the population in growing and developing cities. The process of urbanization associated with inappropriate planning can have consequences affecting the environment and society's quality of life. For instance, urban floods are a concern because they can cause severe effects, such as the death of humans, socio-economic impacts, and material loss. Yin et al. (2015) estimated that catastrophic floods coming from various sources (river, rain, coastal) caused 79 deaths, an economic loss of approximately US$ 1.86 billion and social impacts in urban areas of Chinese cities in July 2011. To cope with these issues, it is essential that preventive approaches, such as improved and densified monitoring systems, should be developed to minimize their impact.

The use of computer systems combined with data information from cities, rivers, weather, and others can contribute to monitoring and controlling urban flood events.

Due to the increased capacity to evaluate data using computational resources, Zhu et al. (2017) report the application of deep learning (DL) in remote sensing, leading to a growth in the number of papers relating to the use of DL in remote sensing. Several review articles were published in the last years regarding the application of DL to remote sensing image analysis. Ma et al. (2019) conducted a comprehensive review of all major sub-areas of the remote sensing field connected to DL; Li et al. (2020) showed the progress of the recent DL based object detection method in both the computer vision and earth observation communities.

Aldebert et al. (2017) mentioned that convolutional neural network (CNN) as one DL method is the most applied in image analysis, and it is able to learn powerful and expressive descriptors from images for a large range of tasks: classification, segmentation, detection, etc. For instance, Santos et al. (2019) applied object detection DL methods to detect tree species in RGB imagery obtained by unmanned aerial vehicle (UAV).

A recent semantic segmentation method and a state-of-the-art CNN structure is SegNet. Yu et al. (2017) state that semantic segmentation makes it easier to understand images because it segments images into semantically significant objects and assigns each part one of the predefined labels. Thereby, different objects from remotely captured images can be extracted simultaneously. Segnet method has been applied in several remote sensing applications. (Du et al., 2018) exploited SegNet technique to classify and extract cropland in high-resolution remote sensing images, showing that the proposed approach efficiently obtained accurate results (98%) for the segmentation task.

The integration of DL and remote sensing in the field of hydrometry is promising, given that remote sensing seeks to obtain information from the Earth's surface without direct contact from the object of study, thus avoiding endangering people and equipment during flood events, and that DL makes automatic measurements possible with high speed and accuracy. For instance, Pan et al. (2018) demonstrated promising results from computer vision systems combined with CNN for river level estimation.

To the knowledge of the authors, there are only a few studies related to the use of DL techniques to densify the monitoring possibilities of (urban) flood events yet. Nogueira et al. (2018) focused on identifying flooding areas from high-resolution imagery using DL approaches. Feng and Sester (2018) described a framework to collect, process, and analyze pluvial flood-relevant information from social media platforms applying DL approaches on user-generated texts and photos.

Such a monitoring tool could also be applied to support real-time flood warning capabilities. The hardware could potentially be simple cameras and thus cost-effective to densify low-cost gauging stations. The use of DL with CNN in remote sensing to monitor surface water can be a valuable supplement to costly and labor-intense standard gauging stations.

The main aim of this paper is to automatically segment river water in RGB imagery using the SegNet semantic segmentation method. We conducted experiments in a river in the East of Germany using RGB imagery collected by a low-cost camera. For the segmentation task, we evaluated different input image resolutions to assess their influence on the river segmentation performance of the SegNet method.

The rest of the paper is organized as follows. Section 2 presents the methodology adopted in this study. Section 3 presents and discusses the results obtained in the experimental analysis. Finally, Section 4 summarizes the main conclusions.

The images were randomly divided into training, validation, and test sets to evaluate the methods and applications. The training set is used to train the methods, while the other sets are used to evaluate the parameters and applications. This procedure guarantees that images used in training are not used in the evaluation and thus ensures correct evaluation of the methods. Pixel Accuracy and Intersection over Union (IoU) were used as validation metrics as they are the most widely used in the literature.

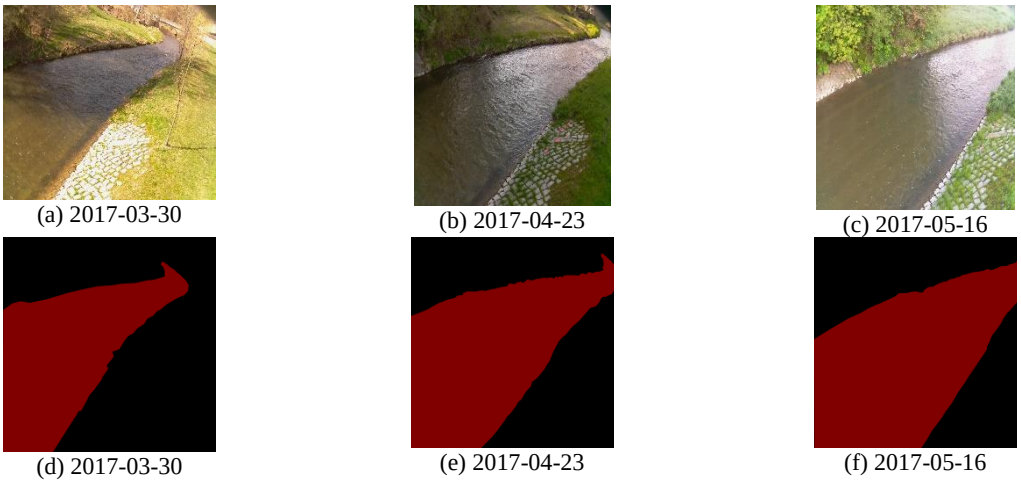
2.2. METHODOLOGY

2.2.1 Image Dataset

The observed river is the Wesenitz featuring a medium scale-catchment located in the East of Germany. A low-cost Raspberry Pi camera sensor was installed 4 m above the ground at a lantern to monitor the river from an oblique perspective (Figure 1).

Figure 1. Position of the camera used at the Wesenitz river.

The dataset was acquired with the 5-megapixel sensor Raspberry Pi Camera Module v2.1 connected to the corresponding single-board computer Raspberry Pi Zero. The image resolution is 2592 x 1444 pixels, and the pixel pitch amounts 1.4 µm. The camera is equipped with a fixed lens with a focal length of 2.9 mm, resulting in a wide field of view at the investigated river section. The Pi camera was calibrated prior to the installation using a scaled temporary calibration field. Image sequences of 5 images are captured every half hour during daylight (Eltner et al., 2018). The captured images reflect different periods of the day over a period of about 50 days allowing for the analysis of the river in different conditions and ambient situations. A total of 3,407 images have been annotated from 2017-03-30 to 2017-05-16 using the LabelMe Software. Figure 2 shows examples of original and labeled images.



(a) 2017-03-30 (b) 2017-04-23 (c) 2017-05-16

(d) 2017-03-30 (e) 2017-04-23 (f) 2017-05-16

Figure 2. Examples of original and labeled images from the dataset.

2.2.2 Semantic Segmentation Method

The CNN SegNet by Badrinarayanan et al. (2017) was used to segment the pixels into water area and background in imagery acquired by the Raspberry Pi RGB sensors. SegNet consists of a symmetrical encoder-decoder followed by a pixel-wise classifier, as shown in Figure 3. The encoder network is similar to the convolutional layers in VGG16 (Simonyan and Zisserman, 2014). These convolutional layers are designed for image classification, and the SegNet encoder network is significantly smaller and easier to train than many other architectures because the fully connected layers of VGG16 are removed. The higher resolution feature maps at the deepest encoder output are acquired when the fully connected layers are discarded. Therefore, the number of parameters in the SegNet encoder network reduces significantly. In the encoder network, convolutions are performed, and a set of feature maps are produced. In other words, this step consists of one or more convolutional layers, which then are batch normalized, and an element-wise rectified-linear non-linearity (ReLu) is applied. Then,

max-pooling is used to achieve translation invariance over small spatial shifts in the input image.

The decoder network is composed of convolutional and a set of upsampling layers, and the memorized max-pooling indices from the encoder feature map(s) are used for upsampling the low-resolution feature map(s). Since the upsampled maps are sparse, convolution layers are applied, producing dense feature maps. In each of these maps, batch normalization is used. The detail preservation can be valuable for delineating the water area and background with good accuracy. In the end, the decoder output has the same resolution as the input image, and a multiclass softmax classifier is applied (Garcia-Garcia et al., 2017). The multi-class softmax classifier activation function produces a probabilistic value for each pixel-wise classification, where the predicted segmentation matches the most likely class at each pixel.

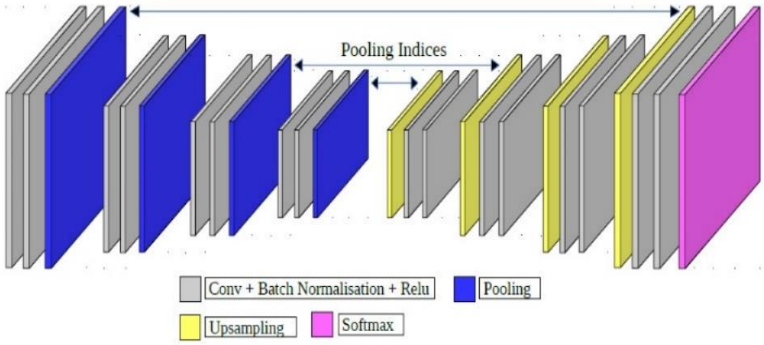


Figure 3. SegNet architecture is composed of an encoder and decoder. The encoder extracts a low-resolution feature map, and the decoder upsamples it to obtain a pixel-wise classification. Source: adapted from Badrinarayanan et al. (2017).

2.2.3 Experimental Setup

During the experiments, we evaluated the input image resolutions of 256 x 256 and 512 x 512 pixels to assess their influence on the river segmentation performance. Garcia-Garcia et al. (2017) mentioned that the integration of information from various spatial scales are required to deal with semantic segmentation. Finding the most suitable image resolution is necessary to balance local and global information. When these steps are done properly, it is possible to achieve good pixel-level accuracy and to deal with local ambiguities

The image dataset was randomly divided into training (60%), validation (20%), and test datasets (20%). The training dataset is used to train the SegNet. The validation dataset was used to determine the learning rate, defining how the weights are adjusted in the CNN and estimating the best suitable number of epochs during training to reduce the risk of overfitting. Finally, the test dataset is used to report the success of the trained network. ImageNet (Deng et al., 2009) was used to determine the pre-trained weights of the SegNet encoder. This procedure is known as transfer learning. The stochastic gradient descent optimizer was used for training with a

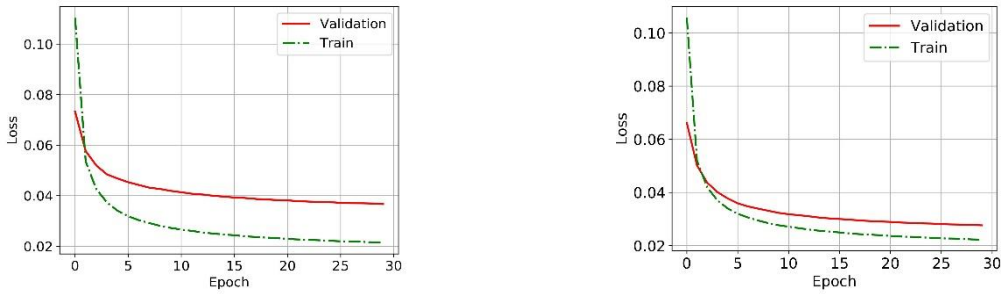
learning rate of 0.001. The number of epochs at which the loss function stabilized in training and validation datasets was 30.

The performance of the river segmentation was measured with the pixel accuracy and Intersection over Union (IoU) metrics. The pixel accuracy shows in percentage the pixels that were correctly classified, while the IoU calculates the ratio between the number of intersecting pixels of ground truth and predicted mask and the number of unified pixels of both masks.

Data processing was performed with a desktop computer on the Ubuntu 18.04 operating system (Intel(R) Xeon(R) Central Processing Unit (CPU) E3-1270@3.8Ghz, Random Access Memory (RAM) 64 GB, NVIDIA Titan V Graphic Processing Unit (GPU) 5120 Compute Unified Device Architecture (CUDA) cores, 12 GB main memory). The algorithms were coded with Keras-Tensorflow, an open-source neural network library written in Python.

2.3. RESULTS AND DISCUSSION

The loss function of SegNet showed indications of overfitting for the resolution of 256 x 256 pixels (Figure 4.a). However, using a resolution of 512 x 512 pixels (Figure 8.b) indicated that overfitting was mitigated because the loss values in training and validation were similar. Generally, the loss function stabilized with the chosen number of 30 epochs, and increasing the resolution from 256 x 256 to 512 x 512 further improved the segmentation. These results show that low-resolution input images make the learning of CNN more difficult. Furthermore, it has to be noted that the higher the resolution of the image is, up to a certain limit to consider memory constraints, the more important details can be learned.



(a) 256 x 256 pixels (b) 512 x 512 pixels

Figure 4. Loss function for SegNet using resolutions of 256 x 256 and 512 x 512 pixels.

Assessing the performance of the DL classification with pixel accuracy reveals an accuracy of 99% considering 256 x 256 pixels resolved images (Table 1). The accuracy improves even further when the resolution is increased to 512 x 512 pixels.

Table 1. Results using pixel accuracy in the subset of images.

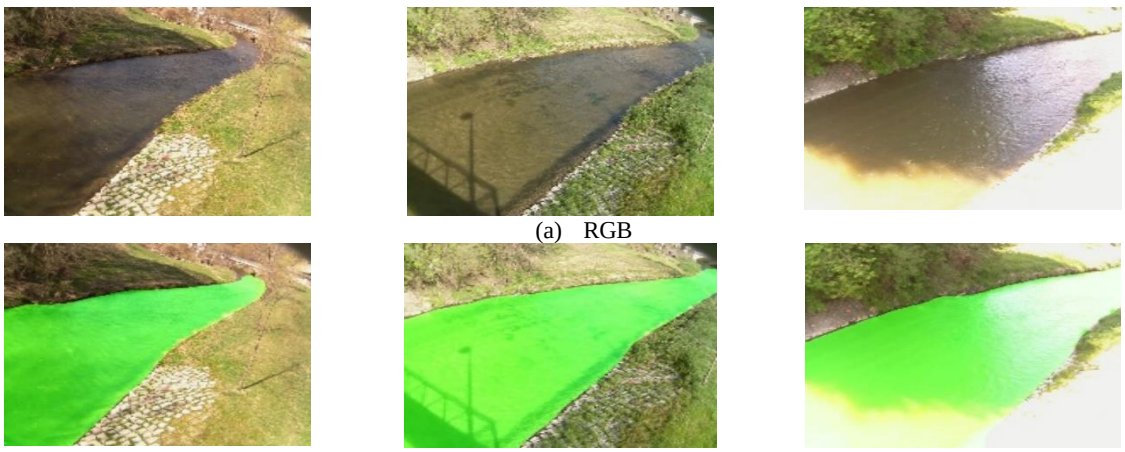
Resolution (pixel)	Pixel Accuracy	
	Background	River
256 x 256	0.9880 (±0.006)	0.9890 (±0.004)
512 x 512	0.9920 (±0.005)	0.9916 (±0.004)

The IoU reveals for both resolutions an accuracy of about 98%. However, the results are slightly better for the higher resolution images (Table 2). Both accuracy estimates, pixel accuracy and IoU, indicate that the method used is efficient to segment water in low-cost camera images.

Table 2. Results using IoU in the subset of images.

Resolution (pixel)	IoU	
	Background	River
256 x 256	0.9750 (±0.005)	0.9795 (±0.005)
512 x 512	0.9821 (±0.005)	0.9852 (±0.005)

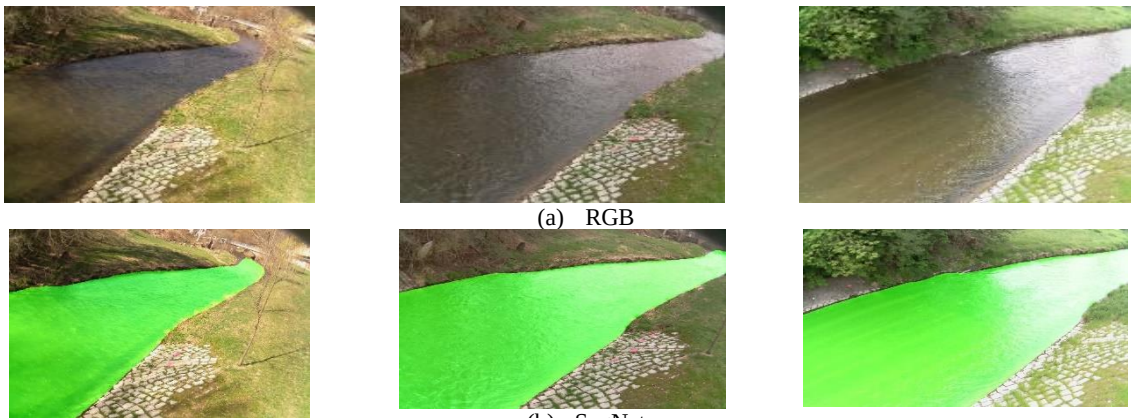
Due to numerous adversities (changes of weather, lighting conditions, and camera position), it is required to assess the learning generalization. Figures 5 and 6 show the segmentation of test images in different circumstances, displaying that river pixels were classified accurately.



(a) RGB

(b) SegNet

Figure 5. Examples of different illumination conditions in (a) original images; (b) SegNet segmented images.



(a) RGB

(b) SegNet

Figure 6. Examples of different point of views in (a) original images; (b) SegNet segmented images.

2.4. CONCLUSIONS

In this study, we present the application of a semantic segmentation method (SegNet) to automatically segment river water in imagery acquired by RGB sensors. The results for pixel accuracy and IoU indicated that the SegNet method is useful to segment the water in imagery, also considering different image resolutions. In addition, although there was a high number of adversities, the segmentation of test images in different circumstances was performed accurately with errors below 2.5%. In future works, it should be evaluated how well it is possible to replicate this segmentation at different rivers. In addition, water segmentation could be applied to obtain various information from a body of water, such as level, speed, and discharge. For instance, there are already works related to image-based approaches applied successfully to camera gauges, and thus being possible to extract water level information automatically. Consequently, it could improve traditional methodologies and become a new source of information.

ACKNOWLEDGMENTS

This research was partially funded by CNPq (p: 433783/2018-4, 303559/2019-5) 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). The authors would like to acknowledge NVidia© for the donation of the Titan X graphics card used in the experiments.

2.5. REFERENCES

Audebert, N., Le Saux, B., Lefèvre, S., 2017. Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images. Remote Sensing 9, 368. https://doi.org/10.3390/rs9040368.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39, 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615.

Borga, M., Stoffel, M., Marchi, L., Marra, F., Jakob, M., 2014. Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows. Journal of Hydrology 518, 194–205. https://doi.org/10.1016/j.jhydrol.2014.05.022.

Da Xu, L., He, W., Li, S., 2014. Internet of things in industries: A survey. IEEE Transactions on industrial informatics 10, 2233-2243. https://doi.org/10.1109/TII.2014.2300753.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database. IEEE Computer Vision and Pattern Recognition, 248-255. https://doi.org/10.1109/CVPR.2009.5206848.

Du, Z., Yang, J., Huang, W., Ou, C., 2018. Training SegNet for cropland classification of high resolution remote sensing images. AGILE Conference.

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, 10-362. https://doi.org/10.1029/2018WR023913.

Feng, Y., Sester, M., 2018. Extraction of pluvial flood relevant volunteered geographic information (VGI) by deep learning from user generated texts and photos. ISPRS International Journal of Geo-Information, 7, 39. https://doi.org/10.3390/ijgi7020039.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Garcia-Rodriguez, J., 2017. A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857.

Li, K., Wan, G., Cheng, G., Meng, L., Han, J., 2020. Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing 159, 296-307. https://doi.org/10.1016/j.isprsjprs.2019.11.023.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B. A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS journal of photogrammetry and remote sensing 152, 166-177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Nogueira, K., Fadel, S. G., Dourado, Í. C., Werneck, R. D. O., Muñoz, J. A., Penatti, O. A.B., Calumby, R.T., Li, L.T., dos Santos, J. A., Torres, R. D. S., 2018. Exploiting ConvNet diversity for flooding identification. IEEE Geoscience and Remote Sensing Letters 15, 1446-1450. https://doi.org/10.1109/LGRS.2018.2845549

Pan, J., Yin, Y., Xiong, J., Luo, W., Gui, G., Sari, H., 2018. Deep learning-based unmanned surveillance systems for observing water levels. IEEE Access 6, 73561-73571. https://doi.org/10.1109/ACCESS.2018.2883702.

Santos, A. A. D., Marcato Junior, J., Araújo, M.S., Di Martini, D.R., Tetila, E.C., Siqueira, H.L., Aoki, C., Eltner, A., Matsubara, E.T., Pistori, H., Feitosa, R.Q., Liesenberg, V., Gonçalves, W.N., 2019. Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. Sensors 19, 3595. https://doi.org/10.3390/s19163595.

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

Yin, J., Ye, M., Yin, Z., Xu, S., 2015. A review of advances in urban flood risk analysis over China. Stochastic environmental research and risk assessment, 29, 1063-1070. https://doi.org/10.1007/s00477-014-0939-7.

Yu, B., Yang, L., Chen, F., 2018. Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 11, 3252-3261. https://doi.org/10.1109/JSTARS.2018.2860989.

Zhu, X. X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine 5, 8-36. https://doi.org/10.1109/MGRS.2017.2762307.

3. USING DEEP LEARNING FOR AUTOMATIC WATER STAGE MEASUREMENTS

The second paper, named “Using deep learning for automatic water stage measurements” was published in the journal “Water Resources Research”. Thales Shoiti Akiyama contributed to the writing, review and development of the deep learning codes. It is referenced as “Eltner, A., Bressan, P. O., Akiyama, T., Gonçalves, W. N., & Marcato Junior, J. (2021). Using deep learning for automatic water stage measurements. Water Resources Research, 57(3), e2020WR027608”.

ABSTRACT

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

3.1. INTRODUCTION

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

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

Observing the river reach with an optical device enables capturing qualitative and quantitative information beyond the water stage or flow velocities. For instance, ephemeral rivers and extended flooded areas become observable, and snow coverage and ice growth can be assessed. Furthermore, the vegetation evolution might be monitored, or the river cross-section changes determined to ensure a continuous and reliable water stage discharge curve. Thus, camera observations can practically support hydrological monitoring. Of course, quality constraints have to be considered when using camera-based approaches (e.g., Elias et al., 2020).

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

superiority of a CNN-based approach to another conventional image processing approach to measuring snow depth with errors below 5 cm and with high flexibility using a measuring rod.

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

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

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

3.2. METHODS

In this study, two CNNs, fully convolutional network (FCN; Long et al., 2015) and SegNet (Badrinarayanan et al., 2017), are tested to evaluate their suitability to segment, and

thus classify, water surfaces in imagery automatically. The boundary of the segmented water area (i.e., waterline) is intersected with a digital elevation model (DEM) of the observed river reach to acquire water stage measurements. The workflow for preparing the study area, camera setup, CNN approach, and water stage estimation is displayed in Figure 1a. A list of the terminology, originating from computer vision and photogrammetry, used in this manuscript to describe the methods is provided in supplement A to facilitate the understanding of the introduced approaches.

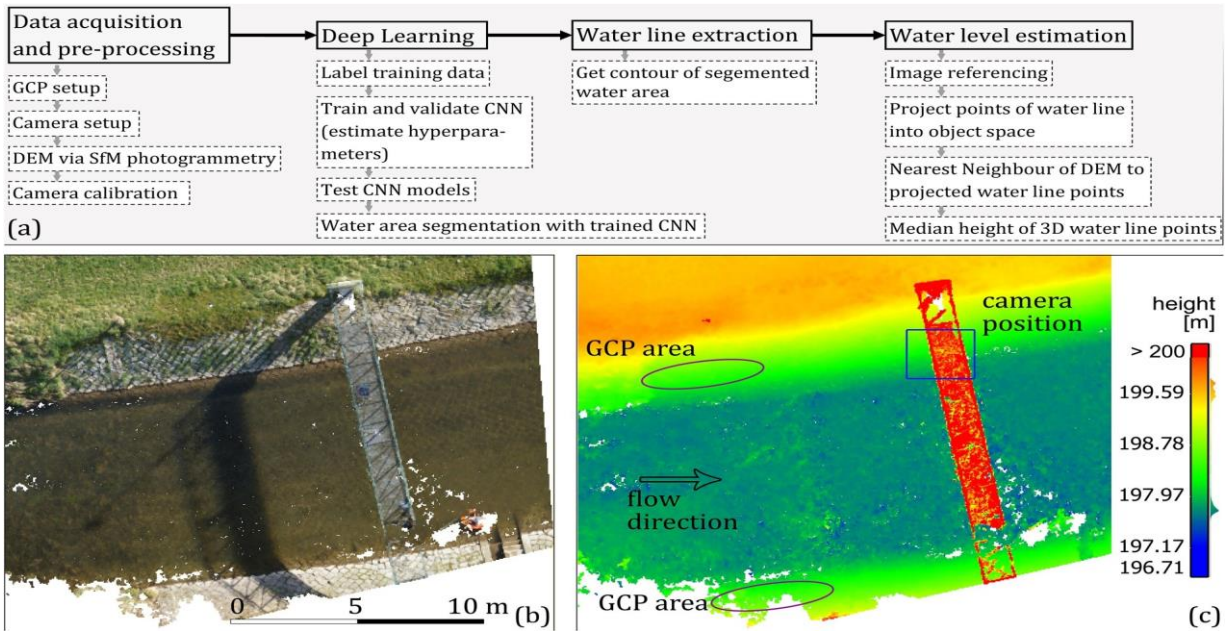


Figure 1. (a) Workflow to automatically extract the water stage from images. (b) Orthophoto and (c) DEM of the area of interest, including an indication of the camera position and ground control points (GCPs) locations.

3.2.1. Study Area

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

Recent advancements in UAV (unmanned aerial vehicle) and 3D reconstruction from images enable easy, flexible, and affordable calculation of high-resolution topography data, which has led to a significant increase in their combined application in environmental sciences (e.g., Eltner et al., 2016). A DEM and orthophoto of the area of interest have been reconstructed (Figures 1b and 1c) using the computer vision techniques structure-from-motion (SfM; Ullman, 1979) and multiview stereo (MVS; Seitz et al., 2006), summarized here as SfM

photogrammetry, from a field campaign in March 2017 (Eltner et al., 2018). Thereby, 20 overlapping images were captured with the UAV Asctec Falcon 8 equipped with a Sony Nex 5N (with a 6-mm fixed lens) at a flight altitude of about 25 m. Afterward, the DEM, comprised of a 3D point cloud, was generated with Agisoft PhotoScan software with a 3D error below 14 mm compared to independent terrestrial light detection and ranging (LiDAR) data. If no UAV is available, terrestrially captured images can also be used to calculate the DEM of the area of interest due to the platform independence of SfM photogrammetry. In total, 17 GCPs were surveyed with a total station and located around the area of interest with mm-accuracy to scale the image measurements. The resulting dense 3D point cloud was processed to correct the underwater areas for the refraction effect, considering the multimedia photogrammetry tool developed by Dietrich (2017). However, if this approach is used to calculate the 3D model beneath the water surface, it has to be considered that the water needs to be calm and that the river bed has to be visible in the images. The resulting precision of the underwater area was 2.7 cm compared to total station point measurements (Eltner et al., 2018).

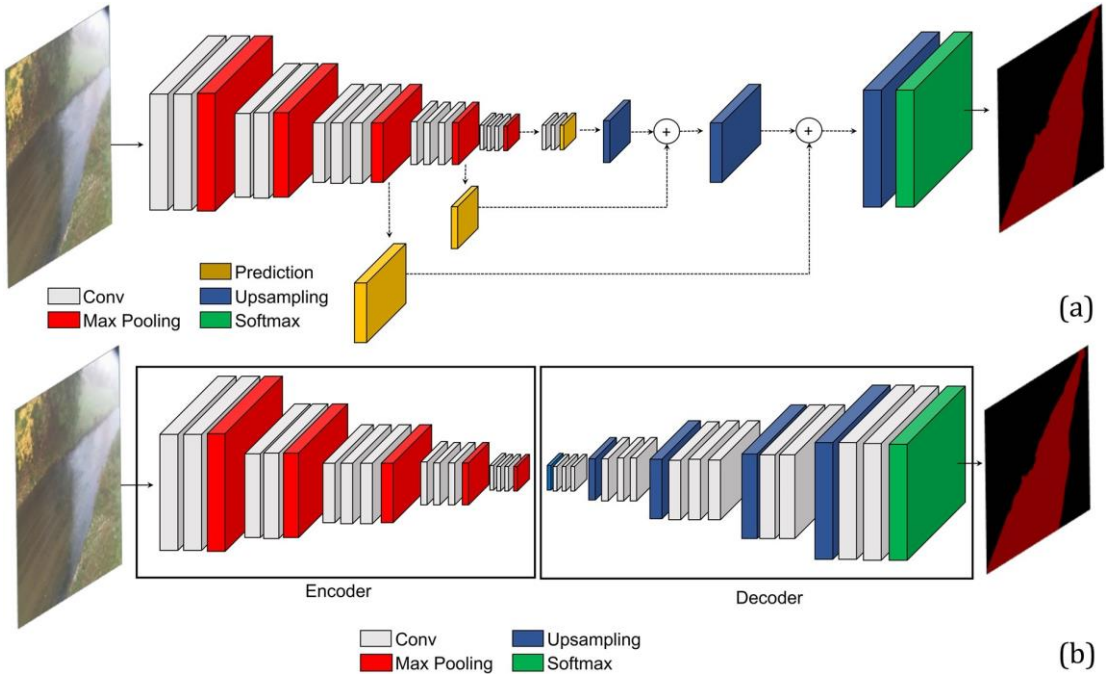


Figure 2. Fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks like semantic segmentation. (a) fully convolutional network (FCN) architecture; the figure is adapted from Long et al. (2015). (b) SegNet architecture is composed of two parts: encoder and decoder.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

GCPs, represented by white circles on black background, were distributed at the left and right shore at the river region observed by the camera (Figure 2). They were measured with a total station with mm accuracy in the same local coordinate system of the 3D model. Due to

strong vegetation growth in the later observation period, additional natural GCPs were extracted at stone corners, and corresponding 3D coordinates were retrieved from the referenced UAV data (Eltner et al., 2018).

3.2.2. Image Acqusition

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

Due to various system failures at this observation spot, cameras had to be changed three times during the observation period. A new calibration of the interior geometry of the camera was necessary for each change to project the image measurements into object space correctly, allowing the metric water stage retrieval (Section 3.2.5). The camera calibration was performed with an in-house built calibration board. Images were captured from different perspectives to estimate the interior orientation parameters (IOPs), minimizing correlations (Luhmann et al., 2014). In the end, the focal length, principal point, and radial distortion parameters were calculated within a bundle adjustment in the software Aicon 3D Studio. The radial distortion parameters are especially important for the potential erroneous projection of image measurements in the image corners.

3.2.3. Water Area Segmentation

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

3.2.3.1. Application of CNN to Water Segmentation

Deep learning allows automatic methods to learn patterns from raw data through multiple layers of processing (LeCun et al., 2015) in artificial neural networks (ANN). Each

layer transforms the input representation into a higher-level representation. In this way, the deeper layers learn aspects of the raw data that are most important to the task while discarding irrelevant variations (higher-level representation).

Deep learning has offered great advances in many fields. CNN is a type of ANN, which is mainly used in image analysis. CNN is composed of convolution, pooling, and fully connected layers. A convolution layer receives an input volume (e.g., an image), which convolves with a set of learnable filters to produce an output volume, also called a feature map. Filters are trained to highlight relevant features on the activation map during learning. After the convolution layer, it is common to apply an activation function such as ReLU (Rectified linear unit) that utilizes a nonlinear function f(x0) = max(x, 0). Pooling layers (e.g., max or average) are applied to reduce the computational cost by reducing the resolution while maintaining the relevant features. The max-pooling layer reduces the resolution maintaining only the maximum value in a given region (usually 2 × 2). Similarly, the average-pooling layer maintains only the average value for a region. Fully connected layers are applied after several convolution and pooling layers to classify the input volume in predefined categories. The last layer of CNN uses a softmax activation function to produce classification probabilities. The learning of layer filters is performed via stochastic gradient descent (SGD; LeCun et al., 2015). The main idea of SGD is to change the filters to minimize the loss function, which measures the discrepancy between the predicted category and the ground-truth (LeCun et al., 2015). The learning rate controls the magnitude of updating the filters.

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

FCN converts CNNs used for classification tasks (composed of convolutional and fully connected layers) into fully convolutional networks that produce coarse activation maps. In this way, FCN can produce a class for each pixel instead of a class for the entire image. In the FCN method, the CNN structure of the model VGG16 (Simonyan et al., 2017) is remodeled by discarding the last layer (final classifier) and by converting all fully connected layers into convolutions. Thus, the original decision-making layer is replaced by (learnable) filters allowing for the input of different sized images. In FCN, convolutional layers with filters of size 1 × 1 are appended to the last layer to predict scores for all classes (yellow layers in Figure 2a), that is, in this study, water and nonwater, at each coarse output locations (scales). Finally, an upsampling layer is used to bilinearly upsample the coarse outputs. This layer increases the resolution by replicating the values of the neighbors. To refine the spatial precision, FCN fuses

the prediction layer with shallower layers of the network by summing predictions and applying a softmax function at the end (as shown in Figure 2a) (Long et al., 2015; Torres et al., 2020).

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

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

FCN and SegNet are segmentation methods already used in several applications, e.g., for scene understanding (Badrinarayanan et al., 2017) and tree segmentation (Torres et al.,

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

3.2.4. Image Dataset

We manually annotated 20,309 images from March 30, 2017 to April 30, 2018 using the software LabelMe (Wada, 2018; see examples in Figure 3). The processed images reflect different periods of the day in different seasons for about 1 year allowing the analysis of the river in different situations. The high number of images during different environmental scene representations is important to reduce the overfitting probability of the model. This data set is referred to as a full data set (FD) in the experiments. To assess the best resolution for resizing the images, which is necessary when adopting CNN methods, we built a subset consisting of 3,407 images from March 30, 2017 to May 16, 2017 (first months).

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

3.2.5. Experimental Setup to Train the CNNs

In this study, each image is downsampled to a fixed resolution due to memory consumption during the CNN training. We evaluated the best image resolution with a subset of the image data set composed of 3,407 images from March 30, 2017 to May 16, 2017. During the experiments, we evaluated resolutions of 256 × 256 and 512 × 512 pixels. The image data set was randomly divided into training (60%), validation (20%), and test sets (20%). The

training set is used to train segmentation methods, while the validation set was used to tune hyperparameters (learning rate and the number of epochs). We refer to Goodfellow et al. (2016) for more information on CNN training. Finally, the test set is used to report the results of the proposed approach. After identifying the more suitable image resolution of 512 × 512 (Section 3.3.1), the FD was used for training following the same workflow as performed for the subset. For each day, the images were randomly divided into training (60%), validation (20%), and test sets (20%). As the test set has images of different dates and seasons, this set is suitable for evaluating the methods with respect to their accuracy and generability. Finally, the classified water pixels are converted into a single water line by extracting the boundary of the water segment based on a traditional image processing technique proposed by Suzuki and Abe (1985).

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

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

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

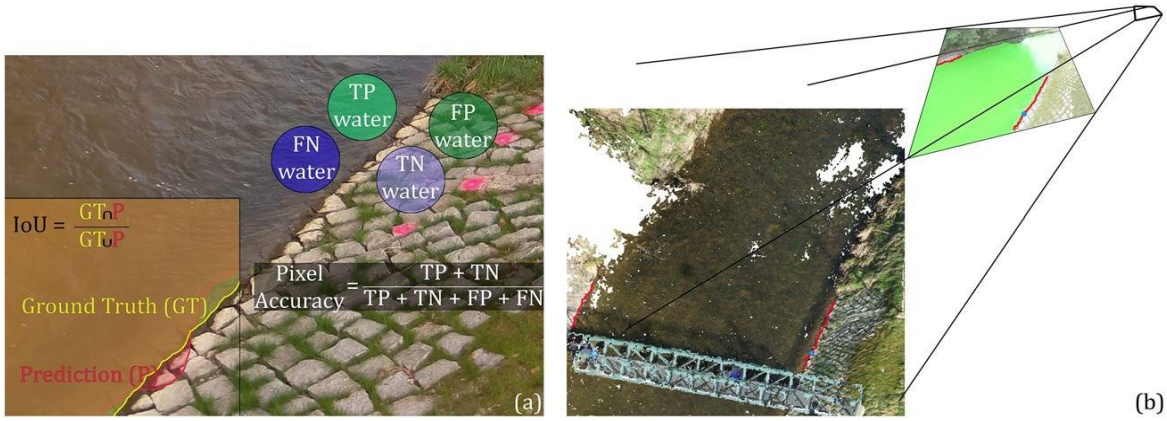


Figure 4. (a) Error metrics (IoU and pixel accuracy) and corresponding equations used to estimate the performance of image segmentation with CNNs. (b) Oriented camera in the same reference system (object space) as 3D point cloud displaying area of interest to transform image measurements (red water line)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

3.2.6. Image Measurements Referencing

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

After retrieving all the necessary parameters, the image points of the water line are projected into object space to intersect these points with the 3D point cloud describing the river reach (Figure 4b). In the 3D point cloud, the nearest neighbor point to the projected image water line point is chosen as a valid point. The Z-coordinate of these water line points corresponds to the water height. The water lines are solely intersected at the location of the paved river reach and thus mostly vegetation-free areas. Vegetated areas are not valid for the intersection approach because plant growth changes the 3D appearance of the river reach. However, for the measurements, a stable 3D object is assumed. If this is not the case, the 3D model of the area of interest would need to be updated for each image-based water stage retrieval. We intersect the image information, i.e. the water contour, with the left and the right river shore. The results

are intersected water line points for the left and right shore, separately. The differentiation between both river sides is chosen to evaluate the potential influence of object to camera distance on the water stage calculation error. To estimate for each image a single water stage value from all intersected water line points, we calculate the median of all height values. Furthermore, a local robust weighted regression filter (Cleveland, 1979) is applied to each of the time series to smooth the temporal water stage change detection to mitigate the impact of strong outliers during the data comparison to the gauge measurement.

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

3.3. RESULTS

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

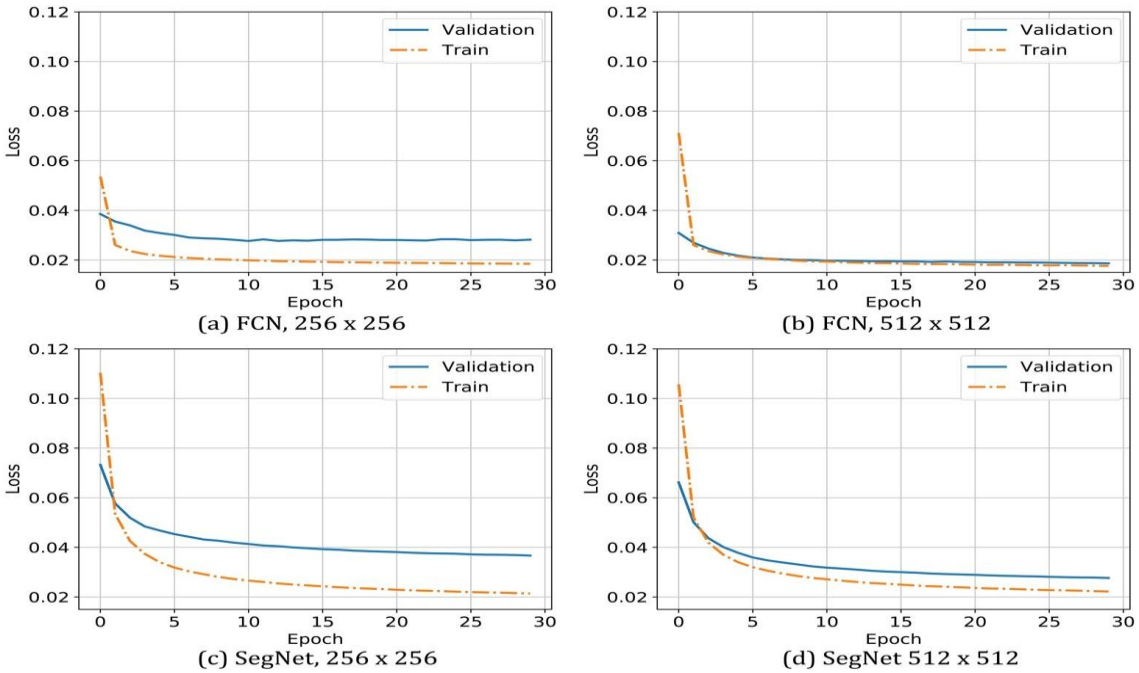


Figure 5. (a) The loss function for two segmentation methods using resolutions of 256 × 256 and 512 × 512 pixels training with the subset image dataset.

3.3.1. Water Segmentation in the images

To assess water segmentation, we varied the resolution of the input image from the subset of the image data set composed of 3,407 images. Figure 5 presents the loss function for both segmentation methods using resolutions of 256 × 256 and 512 × 512 pixels. As can be

seen, the loss functions of SegNet and FCN using the resolution of 256 × 256 pixels showed indications of overfitting because the loss during validation with unseen images remains higher than during training. In contrast, the loss functions of both methods for resolution of 512 × 512 pixels indicate low overfitting (especially for FCN), as the loss values in training and validation were similar. In any case, the loss function for both methods stabilized with the chosen number of 30 epochs.

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

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

3.3.2. Water Stage Estimation

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

measurements at the left shore is also in the range of the automatic approaches (1.7 ± 2.3 cm; supplement B).

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

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

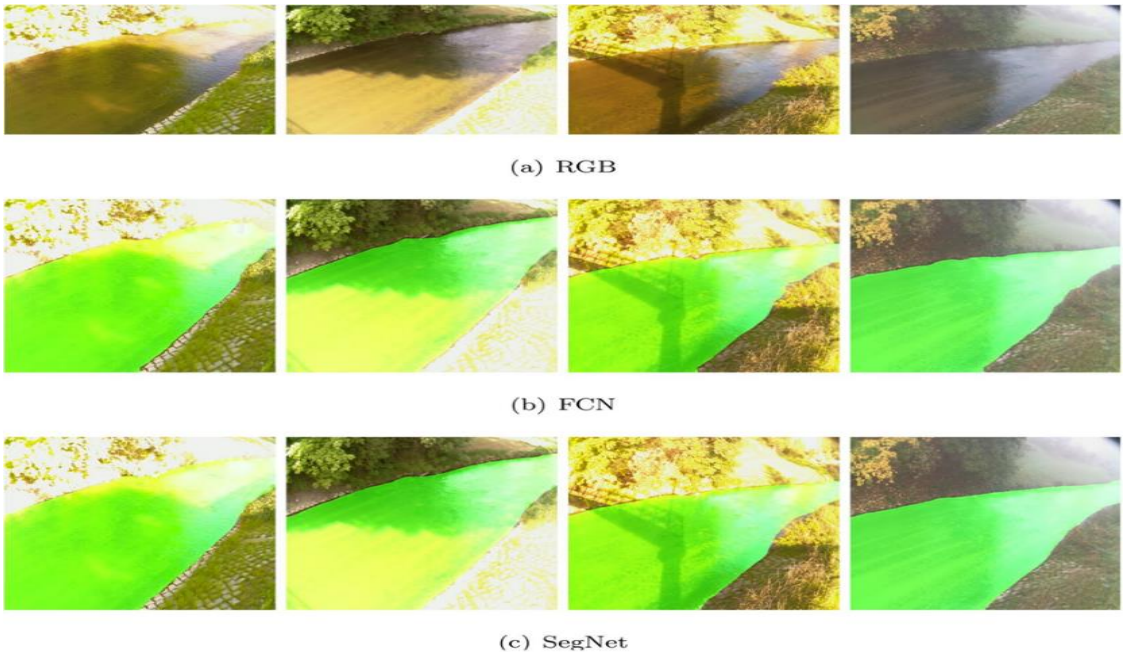


Figure 6. Examples of different illuminations and view in (a) test images segmented by (b) FCN and (c) SegNet.

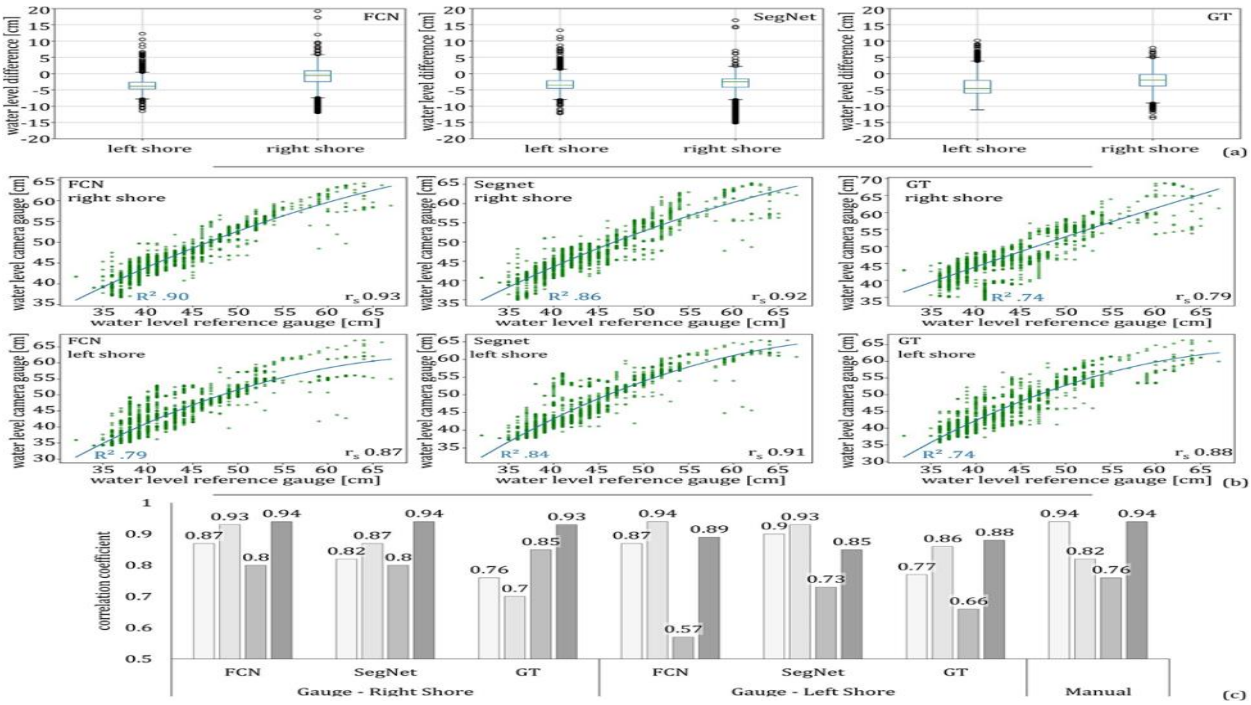


Figure 7. (a) The error of image-based water stage estimation using FCN, SegNet, and GT considering all four time series; the water stage differences between the reference gauge and the camera gauge are displayed. (b) Spearman correlation coefficient rS and scatterplot for each image-based approach compared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

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

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

Besides the absolute comparison of water stages between the reference gauge and the image-based approaches (i.e., camera gauge), we also considered how well water stage changes were captured by the camera gauge (Figure 7b). Therefore, the spearman correlation coefficient was calculated, revealing a very good performance of FCN and SegNet regarding the measurement of water stage variations. The lowest correlation value was 0.87 (left shore, FCN),

and the highest correlation coefficient amounts of 0.93 (right shore, FCN). Interestingly, the GT performs lower than both prediction methods. Assessing the absolute deviations and capturing of water stage change reveals that neither FCN nor SegNet can be considered as outperforming the other. The single point-based manual measurement of the water-shore-border depicts a correlation (0.89); in the range of CNN approaches. The similar performance of the point-based and CNN approaches indicates an overall strong impact of the accuracy of intersecting the image measurements with the scaled real-world point cloud at the error of water stage estimation regardless of the water delineation approach.

3.3.3 Seasonal Performance

In the next step, we took a closer look at different intervals (explained in Section 3.2.4) separately to check if different factors, such as lighting or vegetation growth, on the performance become obvious and to investigate if one CNN is preferable during different environmental conditions. During the first two seasons (spring and early summer), the water stage estimation showed, in most cases, lower accuracies if the right shore is used to intersect the image measurements (Table 2). In the third and fourth observation periods (summer and early autumn), measurements performed with the left shore reveal lower accuracies, probably due to strong vegetation growths between the cobblestone joints, which did not occur at the more strongly shaded right shore.

Considering the correlations between the CNN-based water stage estimation and the reference gauge for each period separately (Figure 7c) revealed that the best results are achieved at the left shore during early summer and at the right shore during autumn, which partly confirms the descriptive statistical findings (Table 4). The lowest correlation coefficients were calculated for the summer period at both shores. Also, the manual measurements using single points show the lowest accordance to the reference gauge in that period.

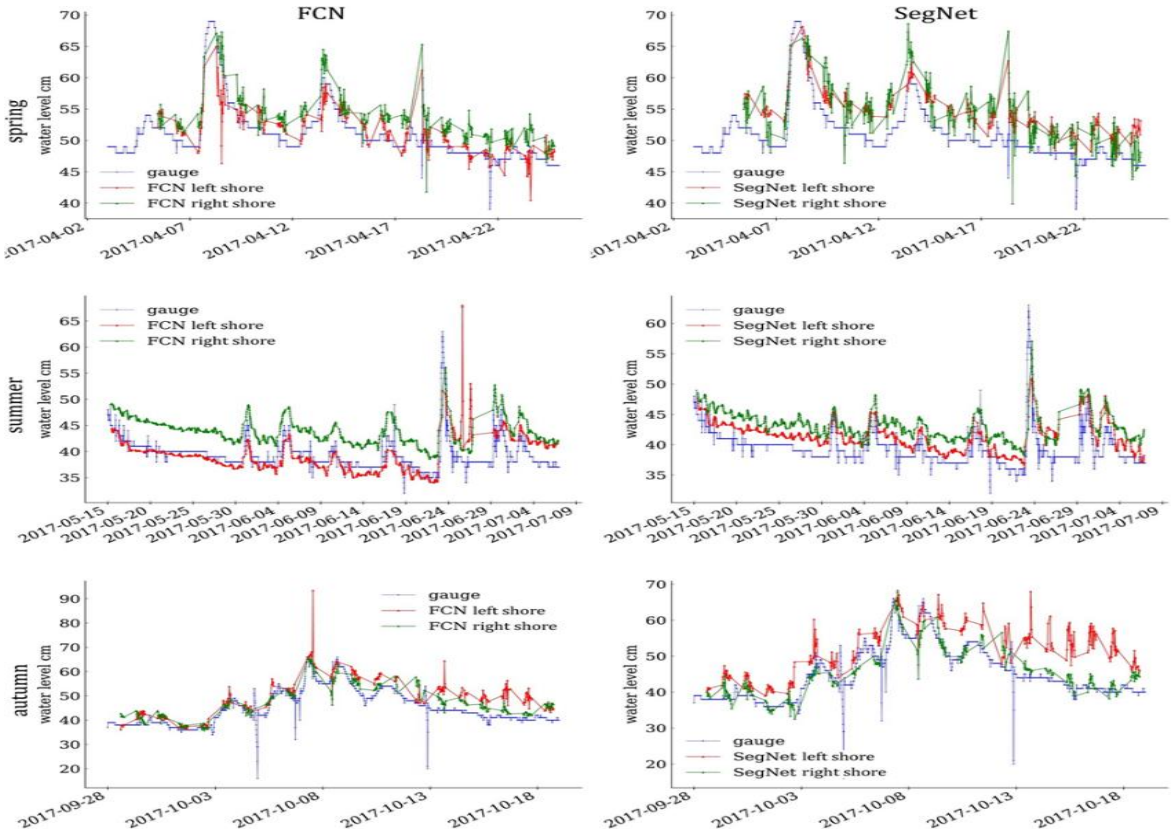


Figure 8. Time series of water stage detection with FCN and SegNet at left and right shore compared to gauge.

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

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

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

especially large overestimation of the water stage occurred on October 17 for the FCN approach. Higher errors were also observed during autumn for the manual measurement approach (Table 2, supplement D).

3.4. DISCUSSION

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

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

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

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

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

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

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

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

Previous camera gauge research (Eltner et al., 2018) at this location achieved a correlation of 0.87 for the summer period, revealing that our novel approach allows for similar or even better water stage accordance. Eltner et al. (2018) used an image-based classification attempt relying on the spatiotemporal water texture, and also required image sequences with further processing steps. The previous approach relied on carefully preselected areas for the water line intersection in object space and failed during very challenging environmental conditions. In contrast, the CNN-based methods do not require such preselection, making them more transferrable and more robust on different lighting conditions, facilitating their operationalization. However, if the water is not visually distinguishable in the images, also the CNN methods will fail as they still rely on the image content. Furthermore, our approach has been developed for images captured during the day. To allow for a continuous river observation the image-based technique needs to be extended to night time photos.

Although the training of the initial CNN has high demands regarding knowledge and hardware, when the CNN has been trained, its application is simple, low-level, and can be used for inference in nonpowerful processors. The aim of this study was to demonstrate the potential of CNNs for robust water segmentation, and we anticipate that future application, and thus an

extension of training data, will increase the transferability to many more river reaches to measure water stages. For future work, it is suggested to investigate unsupervised domain adaption and few-shot learning techniques to increase the generalization capacity of the models considering less or even no additional labeled data set from other rivers.

3.5. CONCLUSIONS

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

DATA AVAILABILITY STATEMENT

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

3.6. REFERENCES

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Chen, Y., Fan, R., Yang, X., Wang, J., & Latif, A. (2018). Extraction of urban water bodies from high-resolution remote-sensing imagery

using deep learning. Water, 10(5), 585. https://doi.org/10.3390/w10050585

Chollet, F. (2015). Keras. GitHub Repository. Retrieved from https://github.com/fchollet/keras

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association, 74(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. IEEE Transactions on Industrial Informatics, 10(4), 2233–2243. https://doi.org/10.1109/TII.2014.2300753

Dietrich, J. T. (2017). Bathymetric structure-from-motion: Extracting shallow stream bathymetry from multi-view stereo photogrammetry.Earth Surface Processes and Landforms, 42(2), 355–364. https://doi.org/10.1002/esp.4060

Elias, M., Eltner, A., Liebold, F., & Maas, H. G. (2020). Assessing the influence of temperature changes on the geometric stability of smartphone-and Raspberry Pi cameras. Sensors, 20(3), 643. https://doi.org/10.3390/s20030643

Eltner, A., Elias, M., Sardemann, H., & Spieler, D. (2018). Automatic image-based water stage measurement for long-term observations in ungauged catchments. Water Resources Research, 54(12), 10–362. https://doi.org/10.1029/2018WR023913

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

Eltner, A., Sardemann, H., & Grundmann, J. (2020). Technical Note: Flow velocity and discharge measurement in rivers using terrestrial and UAV imagery. Hydrology and Earth System Sciences, 24, 1429–1445. https://doi.org/10.5194/hess-24-1429-2020

Fang, W., Wang, C., Chen, X., Wan, W., Li, H., Zhu, S., et al. (2019). Recognizing global reservoirs from Landsat 8 images: A deep learning approach. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(9), 3168–3177. https://doi.org/10.1109/JSTARS.2019.2929601

Feng, W., Sui, H., Huang, W., Xu, C., & An, K. (2018). Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model. IEEE Geoscience and Remote Sensing Letters, 16(4), 618–622. https://doi.org/10.1109/LGRS.2018.2879492

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

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

Herschy, R. W. (2008). Streamflow measurement (3rd ed., p. 510). CRC Press.

Isikdogan, F., Bovik, A. C., & Passalacqua, P. (2017). Surface water mapping by deep learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(11), 4909–4918. https://doi.org/10.1109/JSTARS.2017.2735443

Jiang, W., He, G., Long, T., Ni, Y., Liu, H., Peng, Y., et al. (2018). Multilayer perceptron neural network for surface water extraction in Landsat 8 OLI satellite images. Remote Sensing, 10(5), 755. https://doi.org/10.3390/rs10050755

Kopp, M., Tuo, Y., & Disse, M. (2019). Fully automated snow depth measurements from time-lapse images applying a convolutional neural network. Science of The Total Environment, 697, 134213. https://doi.org/10.1016/j.scitotenv.2019.134213

Kraus, K. (2007). Photogrammetry: Geometry from images and laser scans (2nd ed., p. 459). Berlin, Germany: Walter de Gruyter.

Kröhnert, M., & Eltner, A. (2018). Versatile mobile and stationary low-cost approaches for hydrological measurements. ISPRS Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2, 543–550. https://doi.org/10.5194/ isprs-archives-XLII-2-543-2018

LAWA Bund/Länder-Arbeitsgemeinschaft Wasser (2018). Leitfaden zur Hydrometrie des Bundes und der Länder – Pegelhandbuch. Stuttgart, Germany: Kulturbuch-Verlag GmbH.

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

Leduc, P., Ashmore, P., & Sjogren, D. (2018). Stage and water width measurement of a mountain stream using a simple time-lapse camera.
Hydrology and Earth System Sciences, 22(1), 1–11. https://doi.org/10.5194/hess-22-1-2018

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Paper presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA: CVPR.

Luhmann, T., Robson, S., Kyle, S., & Boehm, J. (2014). Close-range photogrammetry and 3D imaging (2nd ed., p. 683). Berlin, Germany: Walter de Gruyter.

Morgenschweis, G. (2010). Hydrometrie (p. 582). Berlin Heidelberg: Springer-Verlag.
Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In Paper presented at the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: ICCV.

Pan, J., Yin, Y., Xiong, J., Luo, W., Gui, G., & Sari, H. (2018). Deep learning-based unmanned surveillance systems for observing water levels. IEEE Access, 6, 73561–73571. https://doi.org/10.1109/ACCESS.2018.2883702.

Ran, Q. H., Li, W., Liao, Q., Tang, H. L., & Wang, M. Y. (2016). Application of an automated LSPIV system in a mountainous stream for continuous flood flow measurements. Hydrological Processes, 30(17), 3014–3029. https://doi.org/10.1002/hyp.10836

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods, 9(7), 671–675.

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., & Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In Paper presented at the 2006 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY.

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

Stumpf, A., Augereau, E., Delacourt, C., & Bonnier, J. (2016). Photogrammetric discharge monitoring of small tropical mountain rivers: A case study at Rivière des Pluies, Réunion Island. Water Resources Research, 52(6), 4550–4570. https://doi.org/10.1002/2015WR018292.

Suzuki, S., & Abe, K. (1985). Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, 30(1), 32–46.

Torres, D. L., Feitosa, R. Q., Happ, P. N., La Rosa, L. E. C., Marcato Junior, J., Martins, J., et al. (2020). Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. Sensors, 20(2), 563. https://doi.org/10.3390/s20020563

Ullman, S. (1979). The interpretation of structure from motion. Proceedings of the Royal Society of London. Series B. Biological Sciences, 203(1153), 405–426. https://doi.org/10.1098/rspb.1979.0006

Wada, K. (2018). Labelme: Image Polygonal Annotation with Python. GitHub Repository. Retrieved from https://github.com/wkentaro/ labelme

Young, D. S., Hart, J. K., & Martinez, K. (2015). Image analysis techniques to estimate river discharge using time-lapse cameras in remote locations. Computers & Geosciences, 76, 1–10. https://doi.org/10.1016/j.cageo.2014.11.008

APPENDIX A

Supplemental material for article:
Using deep learning for automatic water stage measurements

Table of contents:

Supplement A:
Used terminology in this study:

Activation function: Function determining the activation of neurons in a neural network, which includes nonlinearity.

Activation map: Feature map obtained after a layer of the neural network.

Camera calibration: Approach to estimate the interior camera parameters, for instance considering a specific image acquisition scheme (different perspectives and rotations of the camera) and calibration field (coded targets with approximated known coordinates to enable automatic measurement/identification of these targets in the image).

Camera pose: Exterior camera geometry comprising camera position and orientation.

Convolutional neural network (CNN): A class of neural network composed of convolutional layers.

Encoder: First part of CNN responsible for extracting relevant features from the image.

Decoder: Second part of CNN responsible for up-sampling to predict pixel-wise class labels.

Epoch: A training cycle through the complete training data set.

Exterior camera geometry: Extrinsic camera parameters describing the position (three translations) and orientation (three angles) of the projection centre of the camera in a superior or arbitrary coordinate system.

Fully connected layer: Connects all neurons in one layer with neurons in the next layer.

Fully convolutional network: CNN composed only of convolutional layers.

Ground control point (GCP): Well-visible and distinguishable point in the image to which the 3D coordinates are known to georeference the image(s) and/or 3D model.

Ground Truth (GT): Manually labelled area to enable accuracy assessment of the segmentation in the images.

Image segmentation: Process of partitioning an image into multiple regions.

Interior camera geometry: Intrinsic parameters enabling the mathematically description of the deviation of the camera geometry from the central perspective. Parameters comprise principal distance, which is the distance between the image sensor and the projection centre (usually equal focal length), principal point, which is the perpendicular of the projection centre onto the image plane, and camera distortion parameters such as radial distortion, which describe the deviation of straight rays of light passing through the projection centre and intersecting the image plane.

Interior orientation parameters (IOR): Intrinsic parameters to describe the interior camera geometry.

Learning rate: Hyperparameter in the learning algorithm determining the rate of change in CNN in response to the estimated error.

Loss function: Function estimating the loss by comparing the prediction of CNN and target value.

Point cloud: 3D points, mostly with very high density, that describe the surface of an object or area of interest.

Pooling layer: Layers that reduce the spatial size of the activation map.

Softmax: Function used to obtain a probability distribution of the classes.

Spatial resection: Estimation of the exterior (and if required also interior) camera parameters by relating 2D image coordinates with their corresponding 3D coordinates (usually GCPs).

Stochastic gradient descent (SGD): Optimization algorithm used in learning to find the best CNN parameters.

Test: Set of images used to evaluate the results after training.

Training: Set of images used to train CNN.

Validation: Set of images used to find hyperparameters (e.g., learning rate) during training.

Used software in this study:

Agisoft Photoscan (now Metashape): SfM photogrammetry software that allows for fully automatic, scaled 3D reconstruction based on overlapping 2D imagery and some georeferencing information.

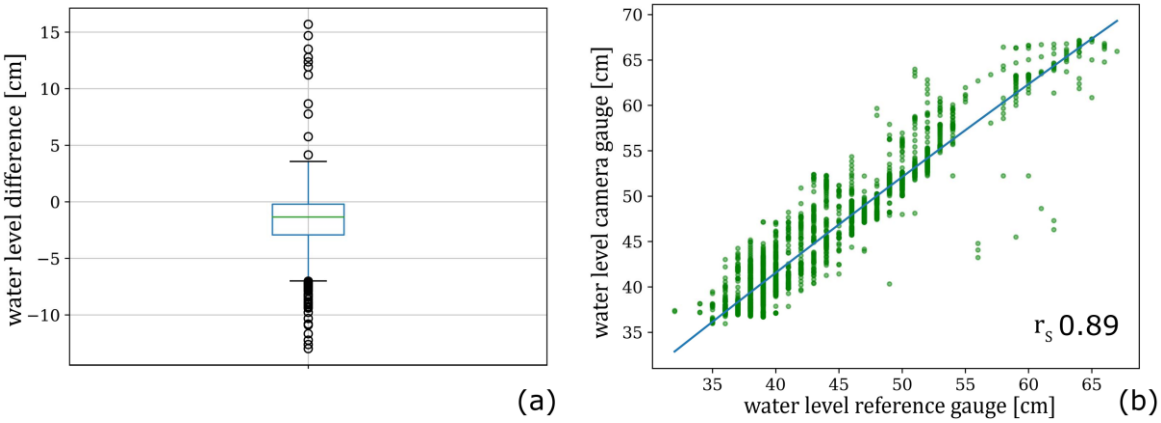
Aicon 3D Studio: Software to enable highly accurate 3D measurements with cameras.

LabelMe: Software that facilitates the generation of ground truth and training data for large image series.

Keras-Tensorflow: Python library enabling easy and user-friendly setup of models for machine learning.

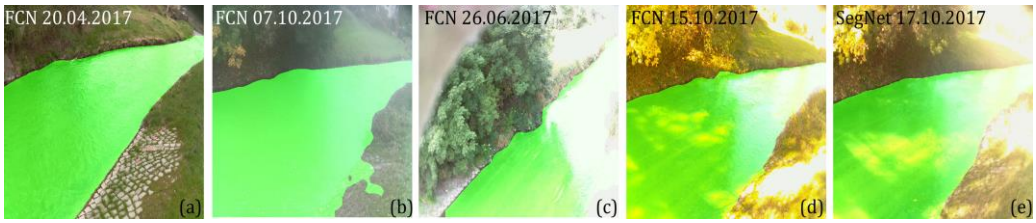
Supplement B:

(a) The error of image-based water stage estimation (considering all four time series) using the water level measured manually in the images at one point along the left shore; the water stage differences between the reference gauge and the image-based gauge are displayed. (b) Spearman correlation coefficient rS and scatterplot of the image-based approach compared to the gauge reference.



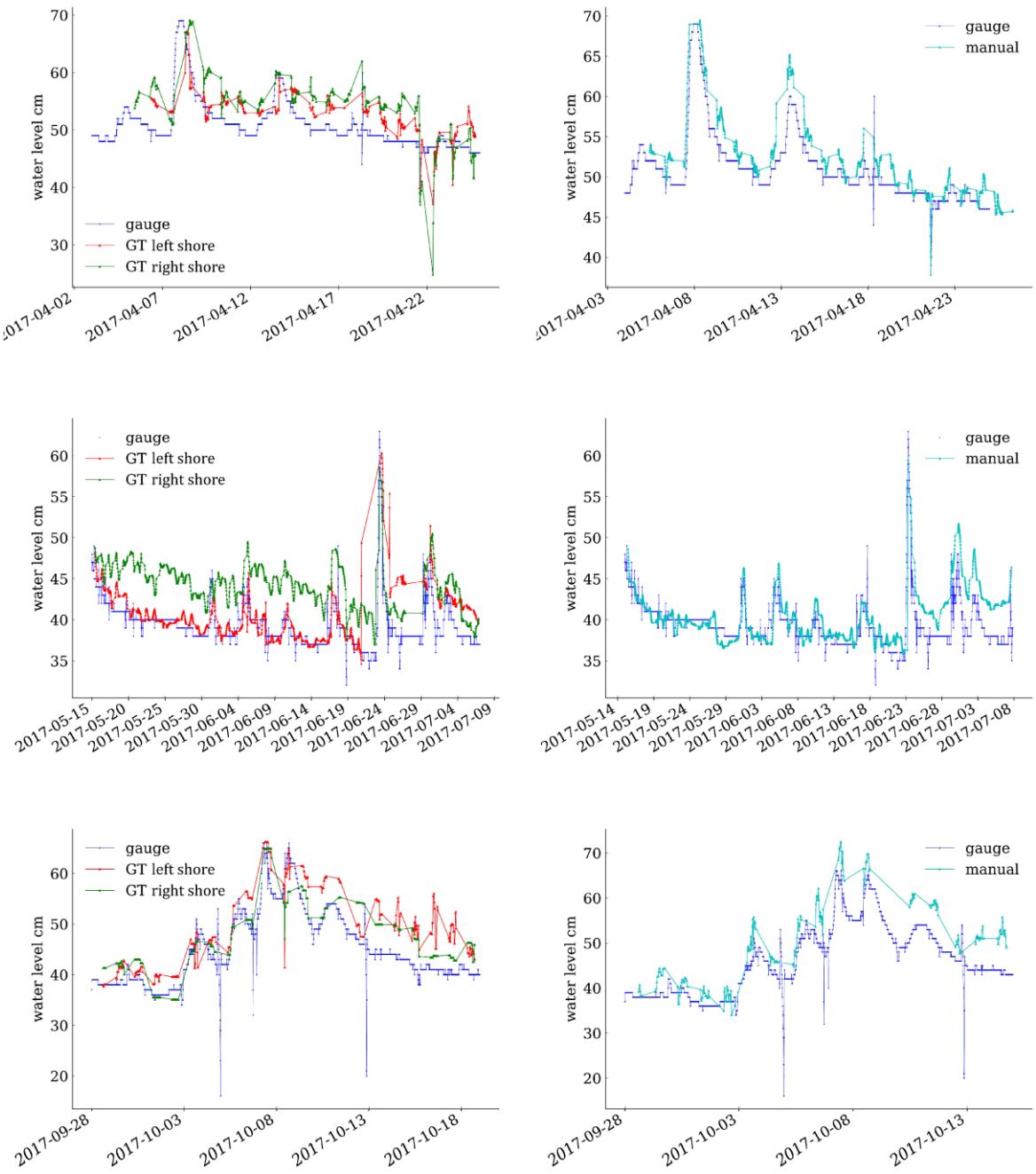
Supplement C:

Examples for different challenging environmental conditions leading to difficulties in water segmentation using the trained CNN FCN and SegNet. (a) favourable lighting conditions, (b) foggy conditions, (c) overexposure, (d and e) strong shadows and overexposed areas in the same scenario.



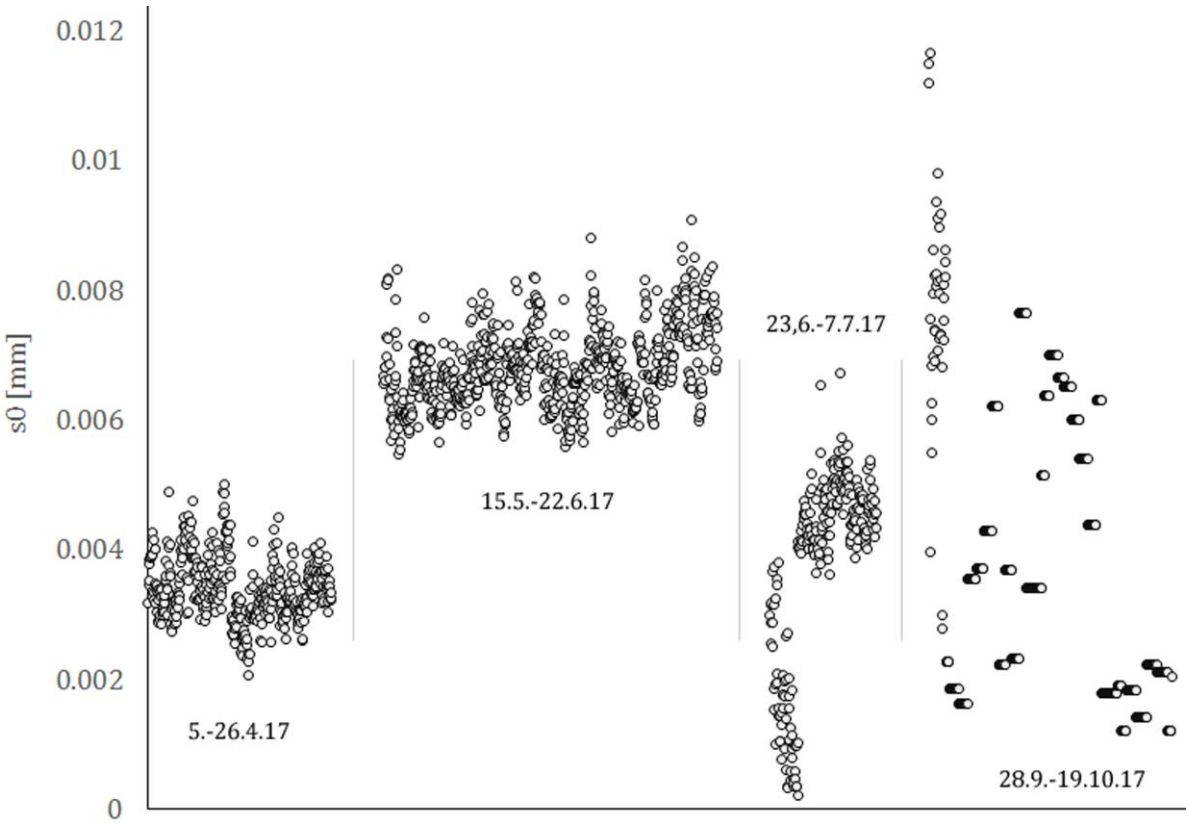
Supplement D:

Time series of water stage detection with ground truth (GT) data measured at left and right shore and considering point-based (manual) measurements measured at left shore only compared to reference gauge values.



Supplement E:

Accuracies of camera pose estimation for each image-based water stage measurement.



4. EVALUATING DIFFERENT DEEP LEARNING MODELS FOR AUTOMATIC WATER SEGMENTATION

The third paper was published in the journal "IEEE International Geoscience and Remote Sensing Symposium IGARSS". It is named "Evaluating different deep learning models for automatic water segmentation". It is referenced as "Akiyama, T. S., Junior, J. M., Gonçalves, W. N., de Araújo Carvalho, M., & Eltner, A. (2021, July). Evaluating different deep learning models for automatic water segmentation. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 4716-4719). IEEE".

ABSTRACT

Deep learning (DL) methods, integrated with imagery obtained by remote sensors, are considered a novel source of information in the field of hydrometry. Their results can be a support for standard gauging systems. In this research, three different model definitions based on the SegNet architecture were developed: training a new model, using a pre-trained model, and performing transfer learning. The main goal of this approach is the performance assessment of convolutional neural network (CNN) generalization to segment images containing different water bodies automatically. Diverse sensors were used to obtain RGB images from different areas of the world. The effectiveness of the CNN was estimated using pixel accuracy and IoU metrics. Training a new model and using transfer learning revealed similar high accuracies that were at least twice as accurate compared to the pre-trained model. However, the transfer learned model is preferred due to significantly lower training expenses.

4.1. INTRODUCTION

Flood effects are considered one of the most expressive natural disasters in the world. When floods impact urban areas, significant consequences are faced, such as socio-economic and infrastructure losses, damaged properties, environmental impacts, affecting people's lifestyle, and, in the worst case, causing the death of living beings [1][2].

Since flood risks cannot be avoided entirely, measures need to be implemented to minimize catastrophes caused by this event, aiming at the best measures to manage it. Besides methods to prevent flood risks, such as urban planning, effective engineering projects for water harvesting, land-use policies, and environmental education, there are numerous models for flood simulation and forecasting, which are being developed and implemented. For instance, Li et al. [3] integrated the Urban Flood Simulation Model (UFSM) and Urban Flood Damage Assessment Model (UFDAM) to propose a structure for risk analysis and the benefits of flood control measures in urban areas. Chen et al. [4] developed and tested the GIS-based Urban

Flood Inundation Model (GUF-IM) to simulate urban flooding. Flood risk management, as well as simulation and forecasting, relies on water resource monitoring. A novel approach in that regard is the usage of deep learning (DL), although the number of studies related to this method is scarce. It further entails the potential for real-time flood assessment, which few cities utilize yet. This is especially relevant in regions where flash floods, caused by high-intensity precipitation in a short period, occur and affect the population suddenly.

DL is a sub-branch of artificial intelligence that allows computational models composed of several processing layers to learn data representations with various levels of abstraction. It is a type of machine learning that enables computers to learn from experience and recognize the world in terms of hierarchical order [5][6]. Recently, DL has been introduced into remote sensing (RS) and the field of geoscience to evaluate data obtained by remote sensors. It is increasingly applied to develop machine learning methods aiming to analyze environmental changes from imagery. For instance, Santos et al. [7] detected specific tree species using object detection methods in RGB imagery obtained by unmanned aerial vehicle (UAV). Using DL in combination with RS can offer a range of benefits to monitoring water resources and urban flood events.

The convolutional neural network (CNN) is a particular kind of multi-layer neural network primarily used in the field of recognizing visual patterns within images [8][9]. Its main tasks are object detection, classification, and segmentation, among others. One state-of-the-art CNN is SegNet, a segmentation method, which is designed to be efficient for pixel-wise semantic segmentation to enhance the process of understanding images [10][11]. Weng et al. [12] used a residual SegNet network to segment water areas using RS images, whereas Du et al. [13] applied a SegNet technique to classify and extract cropland in high-resolution RS images. Our previous work [14][15] showed the potential of SegNet to segment water in terrestrial imagery; however, only images from one geographical area were considered.

The purpose of this paper is to develop three models based on the SegNet architecture. The main goal of these models is to assess the efficiency of CNN generalization and the ability to automatically segment different water bodies from RGB images acquired from different sensors and platforms (aerial and terrestrial). It is expected that the obtained results can be a promising source to complement standard gauging devices as well as to improve the development of real-time flood warning systems.

4.2. METHODOLOGY

4.2.1 Dataset

For the evaluation of this study, two image datasets were considered. A requirement of the datasets was that any water body needed to appear in the collected images. In addition, different image acquirement conditions were considered so that it was possible to implement various perspectives at the water bodies.

Dataset 1 is the same used by [14][15]. It consists of 3,407 images collected by a low-cost Raspberry Pi camera sensor, which was deployed above the ground to monitor a medium-scale river in the East of Germany. The Pi camera stayed at the same position and registered images during different periods of the day to allow for an analysis of the river in different situations.

Dataset 2 is composed of 5,169 images from different locations around the world and consists of diverse photos of rivers, lakes, and ponds. Smartphones, fixed cameras, and UAVs were used to collect Dataset 2. Unlike Dataset 1, it contains images of water bodies with different shapes, sizes and captured from various perspectives. Considering both datasets, a total of 8,576 images have been labeled using LabelMe Software. Fig. 1 shows examples of original and labeled images.

Figure 1. Examples of original and labeled images.

4.2.2 SegNet Architecture for Image Segmentation

Badrinarayanan et al. [11] developed SegNet, which consists of an encoder-decoder network followed by a pixel-wise classifier layer. In this work, SegNet was applied to segment the pixels into the water area and background. The encoder network is based on the VGG16 network. The encoder layers apply a series of convolutional and max-pooling layers to extract a low-resolution feature map that describes the characteristics in the image.

The decoder network upsamples its input feature map applying the memorized max-pooling indices from the corresponding encoder feature map(s). Since the upsampled maps are sparse, convolution layers are applied, producing dense feature maps. The detail preservation can be valuable to delineate the border between the water area and background with good accuracy. Finally, the decoder output has the same resolution as the input image, and a multiclass softmax classifier is applied [16].

4.2.3. Experimental Setup

In order to evaluate the performance of the SegNet architecture in different situations, three approaches were defined with the main objective to assess the power of generalization and the ability to segment images containing different water bodies. The first approach consists of training a SegNet from scratch with a robust set of images capturing water bodies in different situations (Dataset 2). The second and third approaches aim to assess whether the use of prior information (images of a single river - Dataset 1) can improve the results of the first approach. Thus, the second approach consists of evaluating how SegNet, trained only with Dataset 1, is able to segment images containing different water bodies (test set - Dataset 2). As Dataset 1 is composed of images of only one river, the objective is to evaluate the generalizability. Finally, the third approach is to load SegNet with pre-trained weights from Dataset 1 and retrain it with Dataset 2 to assess whether transfer learning and fine-tuning help to improve learning.

4.2.3.1. Training a New Model

To train SegNet from scratch, Dataset 2 is applied and using encoder weights obtained from VGG16. An image resolution of 512 x 512 pixels was considered for all images. Dataset 2 was randomly divided into training (60%), validation (10%), and test datasets (30%). The number of epochs was 50, and the applied optimizer was stochastic gradient descent (SGD) with a learning rate of 0.001. The new model developed in this study is named “model 1”. We use the training set to update the parameters of SegNet. The validation set was needed to determine hyperparameters such as the learning rate and the number of epochs during training to reduce the risk of overfitting. The test set is used to report how well CNN was trained.

4.2.3.2. Testing a pre-trained Model

A pre-trained model is a CNN that has been previously trained using a large dataset. This model can be used for similar segmentation tasks. The pre-trained model used in this study was the CNN developed by Akiyama et al. [14] and Eltner et al. [15]. It was trained using Dataset 1 with images of one river only but with a large number of images. It is referred to as

"model 2". The aim is to assess the ability to generalize in images of different scenarios (Dataset 2).

4.2.3.3. Transfer Learning and Fine-Tuning

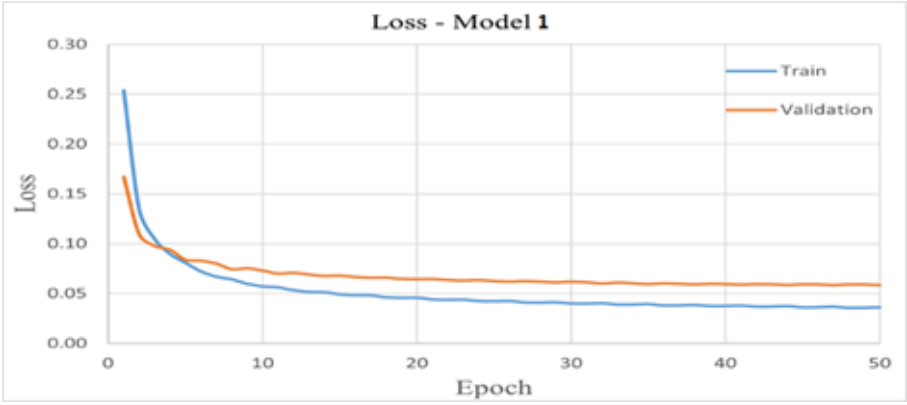
In the third approach, we again use the (with Dataset 1) pre-trained SegNet. However, this time fine-tuning is performed [17]. Therefore, we retrain the parameters of the pre-trained SegNet on Dataset 2 with a low learning rate. The parameters are adjusted at the same time that Dataset 2 is being validated. Therefore, improvements can be achieved by adapting the features of a pre-trained model to the new data instead of training from scratch. The resulting model is referred to as "model 3".

4.2.4. Assessment of the model performance

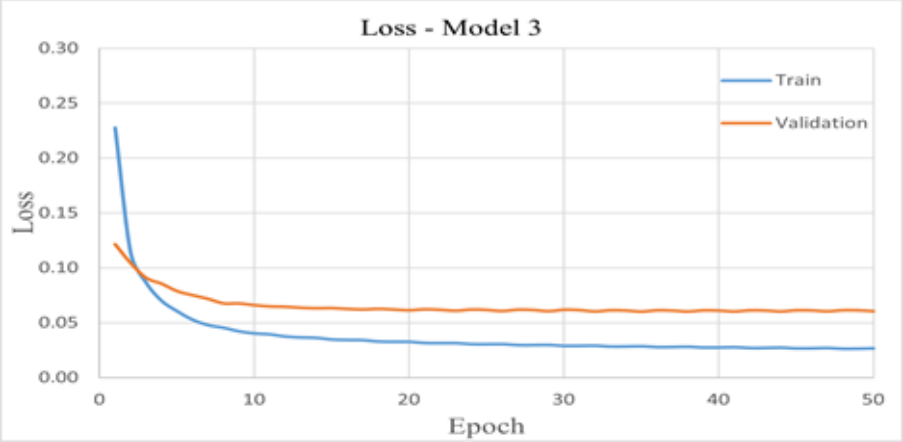
The metrics to assess the segmentation success by the generated model were pixel accuracy and Intersection over Union (IoU). The former presents the number of pixels accurately classified (by percentage). The latter estimates the intersection rate of pixels considered as ground truth and the respective predicted mask. This metric is calculated as unified pixels (by area).

4.3. RESULTS AND DISCUSSION

The following figures and tables show the training and test results. The training of the models was adequate with the number of epochs because the loss curve stabilized (Fig. 2). We can also see that the loss of the training and validation set is close, indicating low overfitting.



(a)



(b)

Figure 2. Loss-Function for SegNet using models 1.(a) and 3 (b).

The results of pixel accuracy and IoU metrics indicate that using the pre-trained model (model 2) on new images (Dataset 2 test set) to segment water areas leads to insufficient results (Table 1). The main reason is that the pre-trained model was trained using only images obtained from the same river and the same camera perspective. In contrast, developing a new model from scratch (model 1) revealed good results in pixel accuracies and IoU. However, for that approach, during training, all parameters had to be trained from the ground up.

Table 1. Results using pixel accuracy and IoU.

	Pixel Accuracy		IoU	
	Background	Water	Background	Water
Model 1	0.991 ± 0.031	0.944 ± 0.091	0.977 ± 0.046	0.916 ± 0.011
Model 2	0.765 ± 0.199	0.525 ± 0.269	0.688 ± 0.179	0.283 ± 0.198
Model 3	0.989 ± 0.036	0.946 ± 0.089	0.975 ± 0.050	0.914 ± 0.124

The results for model 3, which applied transfer learning and fine-tuning, are as good as the results of model 1. Compared to model 1, model 3 converged faster using pre-trained weights. This can be observed in epoch 10, where the loss value is less than 0.05 for model 3 and higher for model 1, presenting better results using fewer data. Thus, this approach of adjusting the parameters to the new dataset can be considered the most useful because it does not require as much effort during training as model 1 does and achieves similar accuracies.

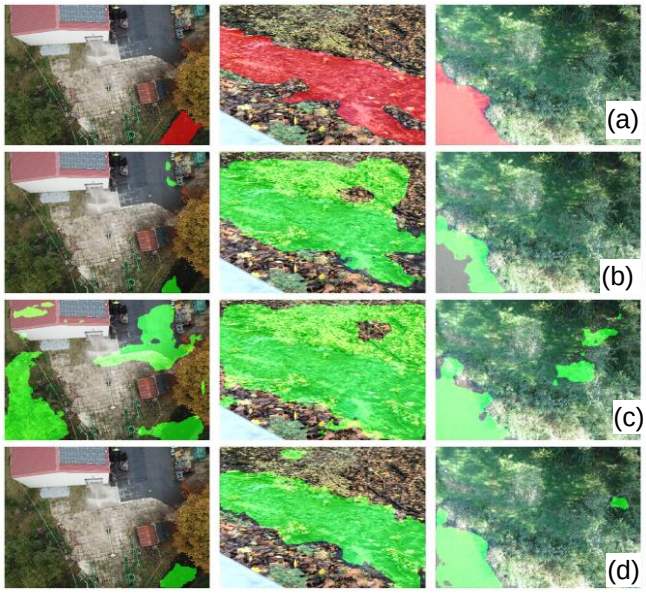


Figure 3. Segmentation of the test images applying the developed models: ground truth (a), model 1 (b), model 2 (c), and model 3 (d).

Model 2 failed to distinguish between the two trained classes because it was only trained using images from the same river and same camera perspective, thus presenting the worst segmentation for the generalization attempt (Fig. 3). Evaluating the results in Table 1 for models 1 and 3, both models show similar results. However, when analyzing their segmentations in Fig. 3, model 3 presents the best results distinguishing what is water and background compared to model 1.

4.4. CONCLUSIONS

Three models based on SegNet architecture were introduced to segment water bodies in images obtained with different sensors and from various perspectives. Generalization does not perform well when utilizing a model that was pre-trained using images only from one river and perspective. However, using the same pre-trained model but applying transfer learning and fine-tuning with additional diverse imagery, better segmentation results could be achieved. Furthermore, the new model converges faster with the pre-trained weights. Therefore, the next step is to improve the model further to increase the generalization capacity in different images to be integrated into camera-based water resource monitoring devices.

ACKNOWLEDGMENTS

This research was partially funded by CAPES Print (p: 88881.311850/2018-01), CNPq (p: 433783/2018-4, 303559/2019-5), and Fundect (p: 59/300.066/2015). The authors acknowledge the support of CAPES (Finance Code 001) and UFMS.

4.5. REFERENCES

[1] T. Tingsanchali, “Urban flood disaster management” Procedia engineering, vol. 2, pp. 25-37, 2012.

[2] J. Yin, M. Ye, Z. Yin and S. Xu, “A review of advances in urban flood risk analysis over China” Stochastic environmental research and risk assessment, vol. 29, no. 3, pp. 1063-1070, 2015.

[3] C. Li, X. Cheng, N. Li, X. Du, Q. Yu and G. Kan, “A framework for flood risk analysis and benefit assessment of flood control measures in urban areas” International journal of environmental research and public health, vol. 13, no. 8, pp. 787, 2016.

[4] J. Chen, A.A. Hill and L.D. Urbano, “A GIS-based model for urban flood inundation,” Journal of Hydrology, vol. 373, no. 1-2, pp. 184-192, 2009.

[5] Y. LeCun, Y. Bengio and G. Hinton, “Deep learning” nature, vol. 521, no. 7553, pp. 436-444, 2015.

[6] I. Goodfellow, Y. Bengio, A. Courville and Y. Bengio, “Deep learning” Cambridge: MIT press, vol. 1, no. 2, 2016.

[7] A.A.D. Santos, J.M. Junior, M.S. Araújo, D.R. Di Martini, E.C. Tetila, H.L. Siqueira, C. Aoki, A. Eltner, E.T. Matsubara, H. Pistori, R.Q Feitosa, V. Liesenberg, W.N. Gonçalves, “Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs” Sensors, vol. 19, no. 16, pp. 3595, 2019.

[8] K. O’shea and R. Nash, “An introduction to convolutional neural networks” arXiv preprint arXiv:1511.08458, 2015.

[9] N. Audebert, B. Le Saux and S. Lefèvre, “Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images” Remote Sensing, vol. 9, no. 4, pp. 368, 2017.

[10] B. Yu. L. Yang and F. Chen, “Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 9, pp. 3252-3261, 2018.

[11] V. Badrinarayanan, A. Kendall and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation” IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pp. 2481-2495, 2017.

[12] L. Weng, Y. Xu, M. Xia, Y. Zhang, J. Liu and Y. Xu, “Water areas segmentation from remote sensing images using a separable residual Segnet network” ISPRS International Journal of Geo-Information, vol. 9, no. 4, pp. 256, 2020.

[13] Z. Du, W. Yang, W. Huang and C. Ou, “Training SegNet for cropland classification of high resolution remote sensing images” AGILE Conference, 2018.

[14] T.S. Akiyama, J.M. Junior, W.N. Gonçalves, P.O. Bressan, A. Eltner, F. Binder and T. Singer, “. Deep Learning Applied to Water Segmentation” The International Archives of

Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 43, pp. 1189-1193, 2020.

[15] A. Eltner, P. Bressan, W. Goncalves, T. Akiyama, J. Marcato Junior. “Using deep learning for automatic water level measurement” Water Resources Research, vol. 55(3), e2020WR027608, 2021.

[16] A. Gargia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez and J. Garcia-Rodriguez, “A review on deep learning techniques applied to semantic segmentation” arXiv preprint arXiv:1704.06857, 2017.

[17] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong and Q. He, “A comprehensive survey on transfer learning” Proceeding of the IEEE, vol. 109, no. 1, pp. 43-76, 2020.

5. EVALUATING THE INFLUENCE OF THE NUMBER OF IMAGES ON MODEL PERFORMANCE

This section aims to analyze how the number of images on the training dataset influences the developed models. The manuscript is under development and will be submitted in the future.

ABSTRACT

Convolutional neural networks require a large dataset to train the models efficiently. This limitation is considered important because there are several challenges in building a suitable database, such as obtaining many images, high-quality label annotations, and computational resources. These requirements take a lot of time-consuming and are often considered costly. Due to this problem, this chapter investigates how variation in the target dataset impacts the effectiveness of semantic water segmentation. The number of images in the training dataset was varied randomly into partitions of 80% (2,480), 60% (1,860), 40% (1,239), 20% (619), 5% (154), and 1% (30), and for better validation, these variations were performed three times. Two types of developed models were tested in our study, and the metrics pixel accuracy and IoU were used to validate the accuracy of the segmentation. The results show that for variations between 80% and 20%, the segmentation does not deteriorate and presents good accuracy (above 80% for both metrics). Thus, it is possible to define that for our study, it would be sufficient to use up to 20% of the images for training. When the variation is between 5% and 1%, both the delimitation of the water bodies and the accuracy (around 70% for both metrics) present a worsening. These achieved results are not suitable for the task at hand, which requires well-defined study water bodies boundaries.

5.1. INTRODUCTION

As seen in the previous chapters, deep learning is an area where artificial neural networks are developed and applied to solve different types of problems emerging with the advancement of technology. Since the algorithms developed for learning contain several possible architectures, complex computations and overloaded tasks influence the training process. For instance, Zhang et al. (2017) show the relationship between training dataset sizes, input data size, and depth of a network to fully memorize the training dataset parameters. Due to these arguments, the best architecture must be analyzed and developed to make the training process faster and more consistent.

Çayir and Navruz (2021) mention that dataset size and quality are factors that most affect accuracy in deep learning studies. In addition, very large datasets take much longer to

train and validate the data, and it is often complicated to assemble a robust dataset (Linjordet and Balog, 2019). Thus, an alternative would be to evaluate the influence of dataset training variation on model accuracy performance in detail.

A few studies aim to evaluate the impact of the size of the training dataset on model performance. Linjordet and Balog (2019) investigated how large the dataset should be to obtain good training performance from models on different neural architectures on the response selection task. A result considered surprising is that many models did not perform as well as desired even if the size of the training dataset was increased. Barbedo (2018) investigated the impact of dataset size on plant disease classification. Çayir and Navruz (2021) show the effects of dataset size in speech recognition. Soekhoe, Van der Putten and Plaat (2016) show the relation between transfer learning and dataset size. Based on the justifications mentioned, this chapter analyzes the impact of the dataset size (Dataset 2 - Chapter 4) on the performance of the Segnet model 1 (Section 4.2.3.1) and model 3 (Section 4.2.3.3).

5.2. METHODOLOGY

This work aims to complement some analyses in relation to the models presented in Chapter 4. Using Smartphones, fixed cameras and Unmanned Aerial Vehicles (UAVs), 5,169 images containing diverse water bodies from diverse locations around the world were collected in order to build the dataset. This dataset was randomly divided into training (60%), validation (10%), and test datasets (30%). In this way, the training dataset consisted of 3,101 images. We varied the number of training images on the dataset to verify if this procedure would influence the performance of the Segnet developed models. So, three trials were performed varying the original dataset randomly by 80%, 60%, 40%, 20%, 5%, and 1%. Table 1 presents the number of images for each variation.

Table 1. Number of images for each variation

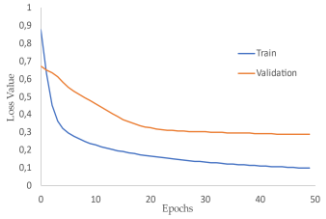
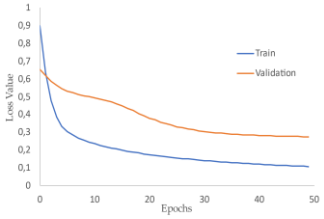
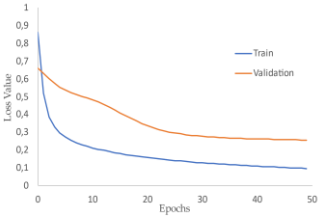
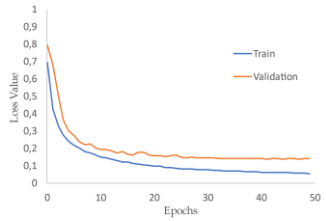
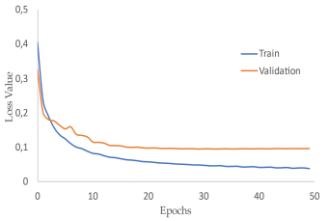
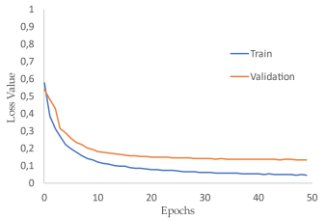
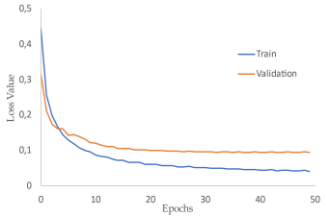
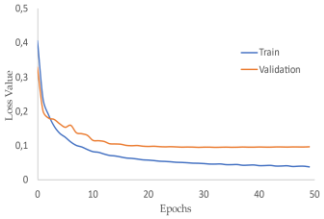
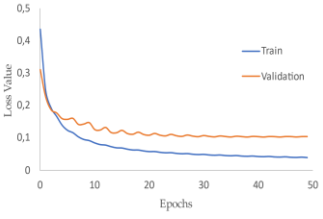
Dataset 2	Images
100%	3,101
80%	2,480
60%	1,860
40%	1,239
20%	619
5%	154
1%	30

Model 1 consists of a SegNet trained from scratch, which the dataset was used for this procedure, and the encoder weights obtained from VGG16 are used. The number of epochs was 50, and the optimizer was stochastic gradient descent (SGD) based on a learning rate of 0.001.

Model 3 consists of the transfer learning and fine-tuning technique. From a pre-trained SegNet model, all parameters were retrained on the dataset with a low learning rate. While validating the dataset, the parameters are adjusted, thus achieving improvements by adapting the features of a pre-trained model to the new data instead of training from scratch.

5.3. RESULTS

First, the Loss Function of each model was analyzed for each variation in the training dataset. Figures 1 and 2 show the values according to each model. The graphics showing the variation of images from 80% to 20% indicate that the losses values before 10 epochs were below 0.5, while most values for the variation from 5% to 1% present higher values. This is due to the smaller number of training images, making it more difficult for the model to learn. The loss value stabilizes faster on the dataset with more images (80%) compared to the datasets with fewer images.

1st Attempt	2nd Attempt	3rd Attempt
		
	(a)	
		
	(b)	
		
	(c)	

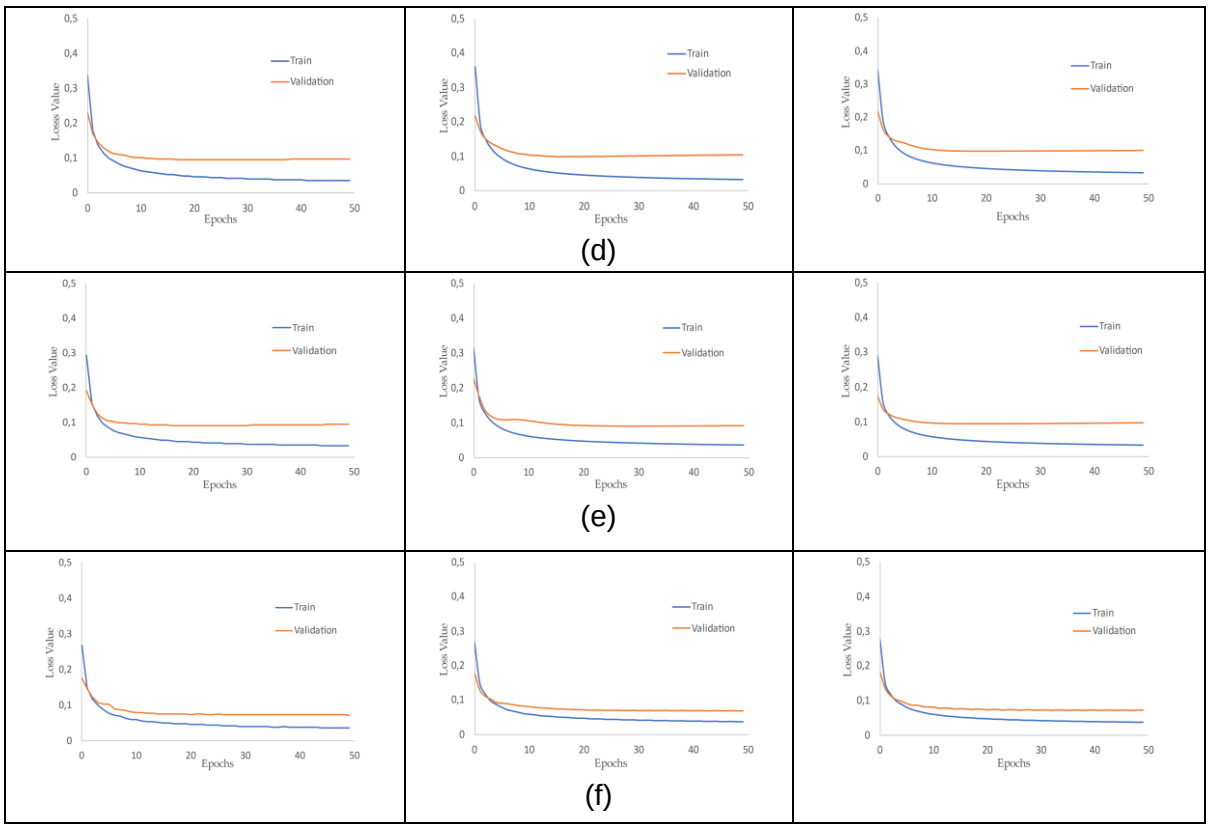
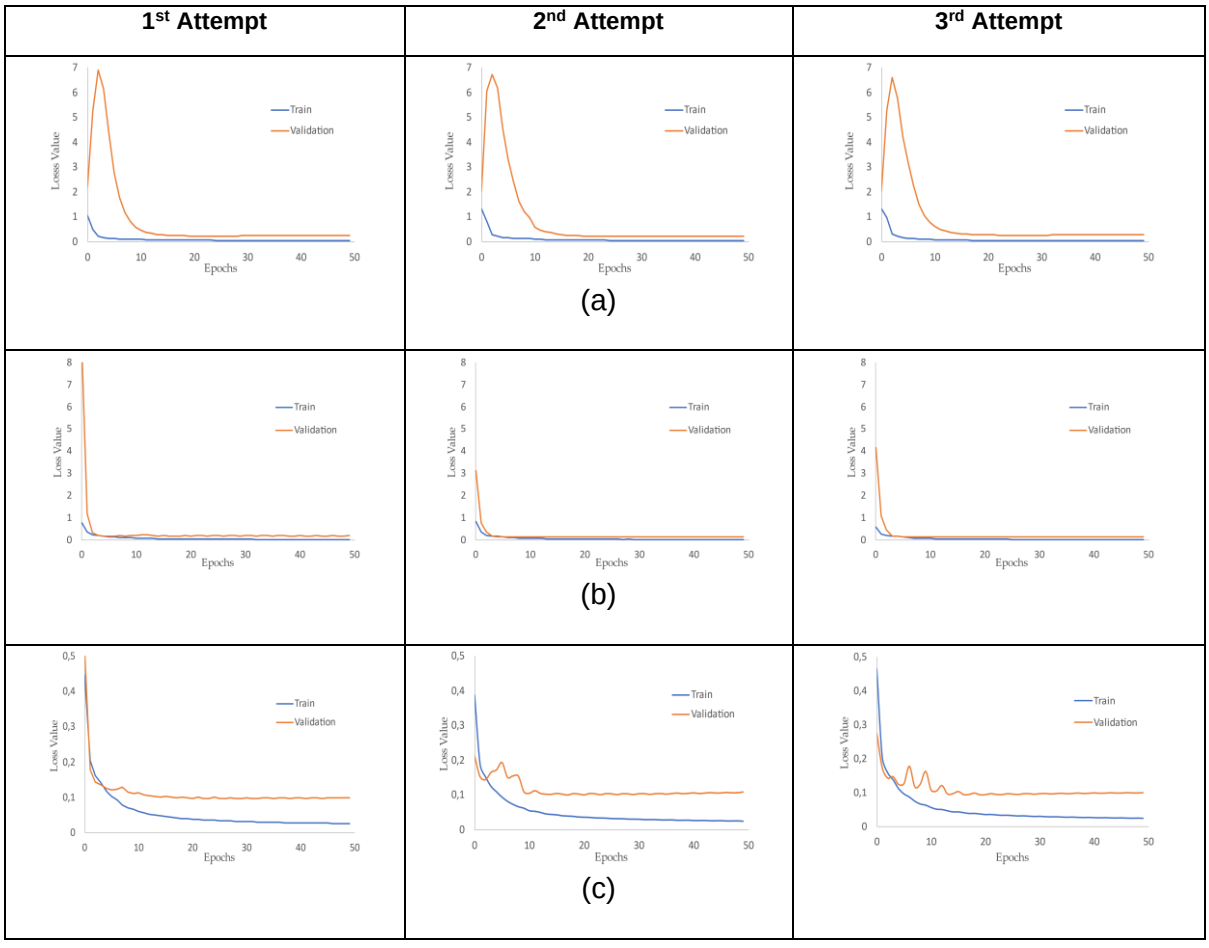


Figure 1. Loss Value of Segnet Model 1. (a)1%, (b)5%, (c)20%, (d)40%, (e)60%, (f)80%.



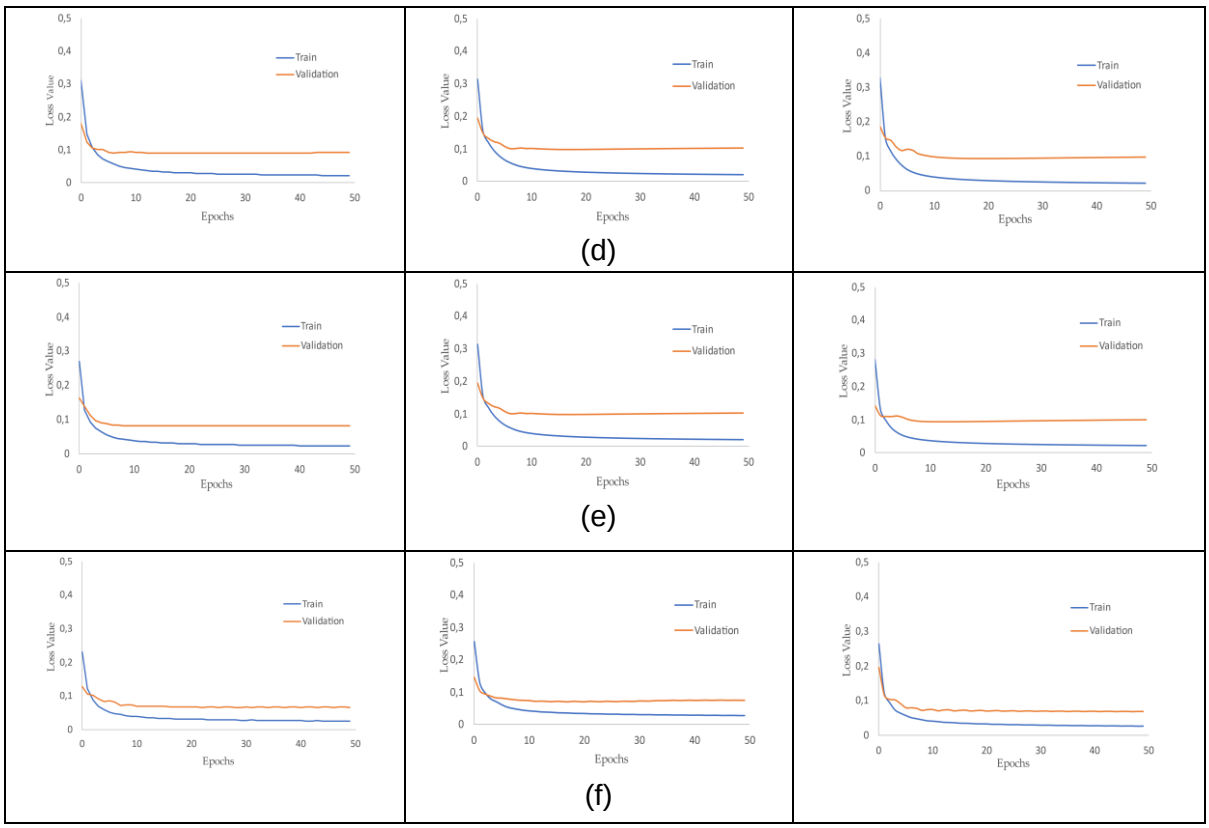


Figure 2. Loss Value of Segnet Model 3. (a)1%, (b)5%, (c)20%, (d)40%, (e)60%, (f)80%.

Tables 2 to 5 assess the performance of the segmentation classification using pixel accuracy and IoU metrics. Both presented good results, indicating that the SegNet Models 1 and 3 are efficient in segmenting images from different sensors. Comparing these two metrics regarding the river segmentation, Pixel Accuracy presents better results than IoU. It is also possible to check that using 1%, the accuracy for rivers decreased compared to the other variations in the number of images.

Table 2. Results using IoU on Model 1.

	IoU							
	1st Attempt		**2nd Attempt**		**3rd Attempt**		**Average**	
	Background	**River**	**Background**	**River**	**Background**	**River**	**Background**	**River**
1%	0.89 ± 0.15	0.68 ± 0.31	0.89 ± 0.15	0.65 ± 0.32	0.90 ± 0.14	0.72 ± 0.29	0.89	0.68
5%	0.97 ± 0.07	0.91 ± 0.15	0.95 ± 0.08	0.82 ± 0.22	0.94 ± 0.09	0.82 ± 0.22	0.95	0.85
20%	0.94 ± 0.10	0.82 ± 0.22	0.96 ± 0.08	0.88 ± 0.16	0.96 ± 0.08	0.88 ± 0.16	0.95	0.86
40%	0.96 ± 0.08	0.88 ± 0.16	0.96 ± 0.07	0.88 ± 0.17	0.96 ± 0.07	0.88 ± 0.16	0.96	0.88
60%	0.97 ± 0.07	0.89 ± 0.16	0.97 ± 0.07	0.89 ± 0.16	0.97 ± 0.07	0.89 ± 0.16	0.97	0.89
80%	0.97 ± 0.06	0.91 ± 0.13	0.97 ± 0.05	0.91 ± 0.13	0.97 ± 0.06	0.91 ± 0.13	0.97	0.91

Table 3. Results using pixel accuracy on Model 1.

Pixel Accuracy								
	1st Attempt		2nd Attempt		3rd Attempt		Average	
	Background	River	Background	River	Background	River	Background	River
1%	0.95 ± 0.10	0.76 ± 0.27	0.97 ± 0.07	0.70 ± 0.31	0.95 ± 0.09	0.80 ± 0.25	0.96	0.75
5%	0.97 ± 0.07	0.91 ± 0.15	0.97 ± 0.06	0.90 ± 0.16	0.98 ± 0.04	0.87 ± 0.18	0.97	0.89
20%	0.99 ± 0.05	0.92 ± 0.12	0.98 ± 0.06	0.93 ± 0.10	0.98 ± 0.06	0.92 ± 0.13	0.98	0.92
40%	0.99 ± 0.04	0.91 ± 0.14	0.99 ± 0.04	0.91 ± 0.14	0.98 ± 0.05	0.93 ± 0.12	0.99	0.92
60%	0.99 ± 0.04	0.92 ± 0.14	0.99 ± 0.02	0.92 ± 0.14	0.98 ± 0.05	0.93 ± 0.12	0.99	0.92
80%	0.99 ± 0.04	0.94 ± 0.10	0.99 ± 0.04	0.94 ± 0.10	0.99 ± 0.04	0.94 ± 0.10	0.99	0.94

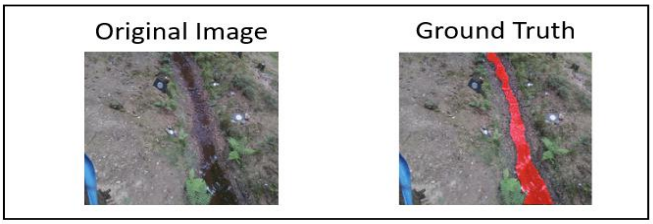
Table 4. Results using IoU on Model 3.

IoU								
	1st Attempt		2nd Attempt		3rd Attempt		Average	
	Background	River	Background	River	Background	River	Background	River
1%	0.81 ± 0.14	0.70 ± 0.31	0.92 ± 0.13	0.74 ± 0.28	0.91 ± 0.15	0.75 ± 0.28	0.88	0.73
5%	0.93 ± 0.12	0.82 ± 0.22	0.95 ± 0.09	0.83 ± 0.22	0.95 ± 0.10	0.82 ± 0.22	0.93	0.82
20%	0.96 ± 0.07	0.89 ± 0.16	0.97 ± 0.07	0.89 ± 0.15	0.96 ± 0.08	0.89 ± 0.15	0.96	0.89
40%	0.97 ± 0.07	0.89 ± 0.15	0.97 ± 0.07	0.88 ± 0.17	0.96 ± 0.07	0.88 ± 0.16	0.97	0.88
60%	0.97 ± 0.06	0.90 ± 0.15	0.97 ± 0.06	0.89 ± 0.16	0.97 ± 0.08	0.90 ± 0.15	0.97	0.90
80%	0.98 ± 0.06	0.92 ± 0.12	0.97 ± 0.06	0.91 ± 0.13	0.97 ± 0.05	0.91 ± 0.13	0.97	0.91

Table 5. Results using pixel accuracy on Model 3.

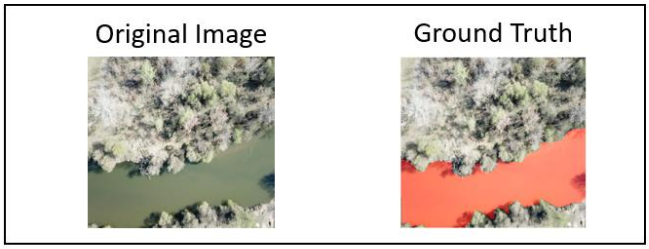
Pixel Accuracy								
	1st Attempt		2nd Attempt		3rd Attempt		Average	
	Background	River	Background	River	Background	River	Background	River
1%	0.97 ± 0.08	0.74 ± 0.30	0.98 ± 0.05	0.79 ± 0.26	0.94 ± 0.13	0.86 ± 0.23	0.96	0.80
5%	0.96 ± 0.10	0.91 ± 0.14	0.97 ± 0.07	0.92 ± 0.15	0.98 ± 0.05	0.86 ± 0.19	0.97	0.90
20%	0.98 ± 0.06	0.93 ± 0.11	0.98 ± 0.06	0.93 ± 0.11	0.98 ± 0.06	0.93 ± 0.11	0.98	0.93
40%	0.99 ± 0.04	0.93 ± 0.11	0.99 ± 0.04	0.91 ± 0.15	0.98 ± 0.05	0.93 ± 0.11	0.99	0.92
60%	0.98 ± 0.05	0.94 ± 0.11	0.99 ± 0.03	0.92 ± 0.13	0.98 ± 0.06	0.94 ± 0.10	0.98	0.93
80%	0.99 ± 0.04	0.95 ± 0.08	0.99 ± 0.05	0.94 ± 0.10	0.99 ± 0.03	0.94 ± 0.10	0.99	0.94

Figures 3 to 8 show the segmentation results for different images. The automated segmentation has worse results considering the variation in 1%. From 5% upwards already show that the performance of the method is valid in delimiting the edges of the water body. However, looking at Figure 8, we notice that the models sometimes have a little difficulty differentiating between water, shadows, and trees. This is because water has similar characteristics to these elements in RGB images, so it is a challenge for future works to eliminate these adversities.



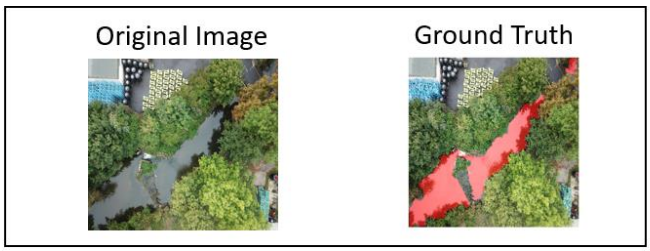
	1%	5%	20%	40%	60%	80%
Model 2						
Model 3						

Figure 3. Segmentation results on Model 2 and 3 varying the number of images.



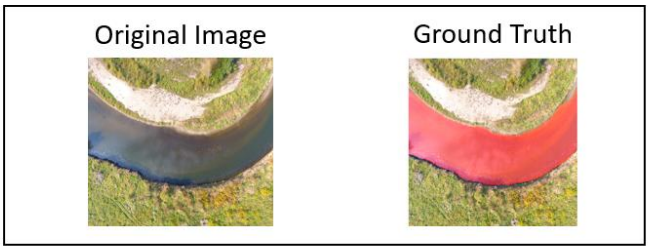
	1%	5%	20%	40%	60%	80%
Model 2						
Model 3						

Figure 4. Segmentation results on Model 2 and 3 varying the number of images.



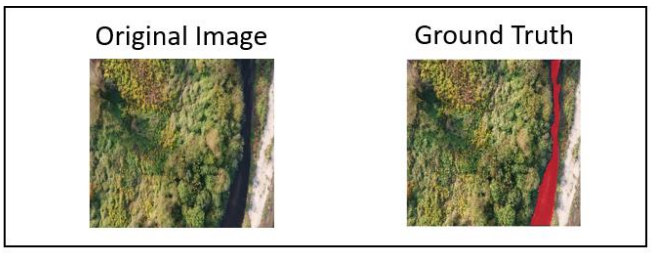
	1%	5%	20%	40%	60%	80%
Model 2						
Model 3						

Figure 5. Segmentation results on Model 2 and 3 varying the number of images.



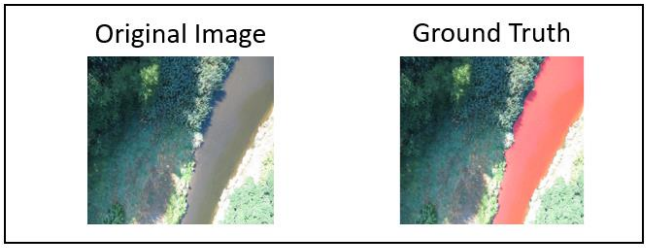
	1%	5%	20%	40%	60%	80%
Model 2						
Model 3						

Figure 6. Segmentation results on Model 2 and 3 varying the number of images.



	1%	5%	20%	40%	60%	80%
Model 2						
Model 3						

Figure 7. Segmentation Results on Model 2 and 3 varying the number of images.



	1%	5%	20%	40%	60%	80%
Model 2						
Model 3						

Figure 8. Segmentation results on Model 2 and 3 varying the number of images.

5.4. DISCUSSION

It was analyzed that both models 1 and 3 presented good accuracy in delimiting the boundaries of water bodies. The variation in the training data set influenced the values of some results, mainly when 1% of the dataset was considered. Regarding how the model was

developed, regardless whether training occurred from the beginning (model 1) or transfer learning and fine-tuning techniques applied (model 3), the performance obtained was considered equal for the task at hand, which consisted in evaluating the segmentation of the water. If model 3 was tested before transfer learning and fine-tuning techniques, the results generated would not be the same. This is confirmed in Chapter 4, which shows the worst segmentation performed for the generalization attempt precisely because it only uses images from the same river and the same camera perspective.

5.5. CONCLUSIONS

In this study, we observed the influence of the number of training images on the performance of different semantic segmentation approaches. Both models presented good results in the task of segmenting different water bodies, where it is possible to observe in most figures the delineation was successful. The worst results and segmentations occur between 5% to 1% range. As many studies mention and it was also verified in this research, the more robust the dataset, the better the results obtained.

The diversity of images containing water bodies, with different means of acquisitions, camera positions, and weather conditions helped in the robustness and creation of a dataset for training, validation, and testing. Furthermore, it was possible to notice that the larger the number of images used for training, the faster is the decrease of the Loss value.

Finally, this study sought to analyze the efficiency of the models according to the number of images. Therefore, it was possible to determine approximately the amount of images needed for the task at hand, consisting of segment water. In future studies, we hope to improve the models so that fewer images are needed, thereby providing accurate models for the task of segmenting bodies of water and helping other studies related to water resources.

5.6. REFERENCES

BARBEDO, J. G. A. Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. **Computers and electronics in agriculture**, v. 153, p. 46-53, 2018.

ÇAYIR, A. N.; NAVRUZ, T. S. Effect of Dataset Size on Deep Learning in Voice Recognition. *In*: 3RD INTERNATIONAL CONGRESS ON HUMAN-COMPUTER INTERACTION, OPTIMIZATION AND ROBOTIC APPLICATIONS (HORA). IEEE, 2021 p.1-5.

LINJORDET, T.; BALOG, K. Impact of training dataset size on neural answer selection models. *In*: EUROPEAN CONFERENCE ON INFORMATION RETRIEVAL. Springer, Cham, 2019. p. 828-835.

SOEKHOE, D.; PUTTEN, P. V. D.; PLAAT, A. On the impact of data set size in transfer learning using deep neural networks. *In*: INTERNATIONAL SYMPOSIUM ON INTELLIGENT DATA ANALYSIS. Springer, Cham, 2016. p. 50-60.

ZHANG, C. et al. Understanding deep learning (still) requires rethinking generalization. **Communications of the ACM**, v. 64, n. 3, p. 107-115, 2021.

6. SEMANTIC SEGMENTATION OF WATER CONSIDERING VIDEO STRUCTURES

This section presents a manuscript called “Semantic segmentation of water considering video structures”, which will be further submitted.

ABSTRACT

Novel automated monitoring techniques for mapping water resources based on deep learning have been growing in recent years, providing a greater facility for researchers and authorities due to optimized data acquisition schemes and measurements with increased accuracy and robustness. However, most of these techniques require large labeled datasets, which is time-consuming. This paper introduces a deep learning method for automated segmentation of rivers and water bodies using only the initial labeled frame. Space-Time Correspondence Network (STCN) is used to model water features in video sequences, which require only one labeled image per video. We obtained videos from different remote sensing platforms, such as UAVs (DJI Phantom 4, DEERC DE25), smartphones, surveillance cameras, from different countries (England, Germany and Brazil). In addition, different weather conditions were considered in order to analyze the performance of this method. A total of 25 videos were applied. Also, we analyzed time-lapsed images (acquired at each 15 minutes) in a public dataset. The results indicated that the model was efficient to segment water bodies, even considering videos taken from different perspectives and weather conditions. It emerges as a new source for deep learning and photogrammetry techniques on providing information of a water body in an innovative way, enabling different methods to monitor water bodies based on camera systems.

6.1. INTRODUCTION

The monitoring of water resources is important and indispensable, providing different types of information, such as chemical components, precipitation rate, superficial runoff or discharge of catchments. These information influence decision-making processes to prevent catastrophes, for instance caused by flooding. Due to sudden high intensity precipitation, urban flooding has become a major concern for the population residing in cities. There are numerous factors that favor the occurrence of floods, such as the growth of cities and increased population density in high-risk areas; the drainage system expansion being slower than the development of cities; and the increase in impermeable areas [1].

Due to the growth and urbanization process in cities, public safety policies must be implemented to ensure the population's well-being and safety. The occurrence of floods is

associated with poor urban planning, thus causing harmful social impacts on the environment and the quality of life [2]. Urban flooding is considered a big issue, since it causes material, economic and environmental losses and, in the worst case, causes the death of humans and other living beings [3].

Flood risks cannot be avoided completely, hence monitoring approaches and measures must be applied and advanced in order to deal with these events and to minimize their impact [4]. Water stage is an important hydrological observation, which helps to assess urban floods. There are different methods to measure it, e.g., based on floating and pressure gauges. However, besides the possibility of losing the instruments during extreme events, these measurements can be expensive amongst other due to maintenance requirements [5].

Another form of monitoring are remote observation techniques, which have the advantage that they do not require sensors in the water body or river. The development and application of image and video-based methods have been growing over the last decade. They can be less expensive and provide an innovative source of data. For instance, studies illustrated the suitability of short videos to track particles on the water surface from ground-based sensors and UAVs (Unmanned Aerial Vehicles) to measure water flow velocities [6], while others determined the water stages of rivers using either stationary Raspberry Pi cameras [7] or utilizing UAVs to improve existing hydraulic models as well as to aid strategies in case of flooding or inundation [8].

To apply cameras for water stage measurements automatic image processing techniques are required. Deep Learning (DL) as a sub-branch of artificial intelligence gained large attention recently in that regard. It allows to learn representations of data at multiple levels of analysis via computational models composed of multiple processing layers [9,10]. The interaction of remote sensing and DL is increasing due to the advancement of computational resources and the increased performance of data evaluation [11]. While DL seeks to obtain information with high degree of automation and speed, remote sensing is amongst other focusing on obtaining information about the Earth's surface. Thus, the application of DL based object detection methods for remote earth observation is a large field of research [12,13], analyzing data in a more robust and efficient way.

There are several methods of interpretation in remote sensing using DL, such as classification, segmentation, and object detection. Regardless of which method is used, the best results from DL models for such tasks were those that required a large dataset for training [14]. Developing a robust dataset for deep neural networks is considered a major problem. Tasks such as acquiring numerous images to build a proper dataset, number of manual annotations, all these requirements demand a lot of time and cost to train or retrain the parameters of a

collected dataset. [15]. Thus, new methodologies are emerging that do not require a large amount of data, such as few-shot, which uses a few labeled images and thus learns to classify images for classes never seen during training [16]. One-shot Learning [17] and Zero-shot Learning [18] are other cases that learn with few training samples, where the first uses only one labeled sample to train and the latter zero annotation. All these methods seek to learn transferable knowledge so that it can be generalized to new classes and thus recognize patterns in images with few annotation examples. This is a challenge for remote sensing imagery, as targets often vary in size and spatial resolution, making it even more difficult for tasks with few annotations.

There are a few studies that have applied the few-shot method for remote sensing. One study used the few-shot learning-based method to detect objects in remote sensing images using the YOLOv3 model [15]. Another used the few-shot learning method for hyperspectral image classification with only a few labeled samples [19]. This method has also been used for remote sensing scene classification tasks [16]. Although few-shot techniques are already in use in remote sensing, there are almost no studies on the issue of segmentation tasks. Furthermore, there are few studies using one-shot and zero-shot on remote sense images. One of these studies based on a one-shot method consisted of using remotely sensed images to train the model in image classification and then testing it in geospatial object detection [20].

Regarding the intuit of recognizing patterns in videos, Video Object Segmentation aims to identify and segment target instances in a video sequence [21]. Video-based approaches are also emerging as a new alternative in object segmentation tasks. It is common to see their application in segmenting humans, cars, houses, and objects seen from everyday routine [22, 23]. These alternatives have also been applied to the semantic segmentation of UAV videos [24]. However, it is hard to find applications focusing on the segmentation of water bodies on video structures.

Based on these reasons, we applied a DL approach based on Space-Time Correspondence Network (STCN), considered a state-of-the-art method in object segmentation from videos, for semantic segmentation of water on remote sensing videos obtained with different sensors (aerial and terrestrial). The videos were taken from different perspectives and weather conditions in order to analyze the performance of the proposed method. The video segmentation extracts waterline contours automatically, which only one label frame is used to segment videos (sequence of frames) in order to store the most relevant temporal information. This generated water segmentation can subsequently be used to calculate the water stage. This work can be beneficial for image-based water extraction at different locations around the world to complement traditional measurement systems and to eventually develop a real-time flood

warning system in regard to smart cities to assist the population in decision making. Our contribution is also in becoming the dataset available for further investigation.

6.2. MATERIALS AND METHODS

In this study, we investigate the performance of the STCN model [21] on water segmentation using videos. In addition, its results are compared with other types of deep learning models where the framework is based on a large number of images.

6.2.1. Study Area

Videos from various parts of the world were used in order to generate a robust dataset that contains different bodies of water, such as rivers, lakes, streams. Thus, several areas of study were observed in the article. However, two study areas were the focus of application and analysis to investigate the performance of our methodology. The first is the Wesenitz River, situated in a medium-scale watershed (227 km²) in the state of Saxony, Germany. Torrential precipitation events have increased over the past 25 years in this area, highlighting the need to intensify monitoring of this region [25].

Figure 1. Wesenitz River, Germany. (a) Position of the camera used to obtain the data; (b) Example of an image obtained.

The second study area is the urban region of the Prosa basin (31.97 km²) located in Campo Grande, Brazil. Multi-temporal analysis showed that the population growth of this area led to an increase of the urban area by 96.63% and a decrease in vegetated areas, causing environmental degradation due to erosion and siltation and causing emerging floods and inundations [26]. The Prosa basin is one of the most vulnerable to flood risk among the various constituents in the municipality [27].

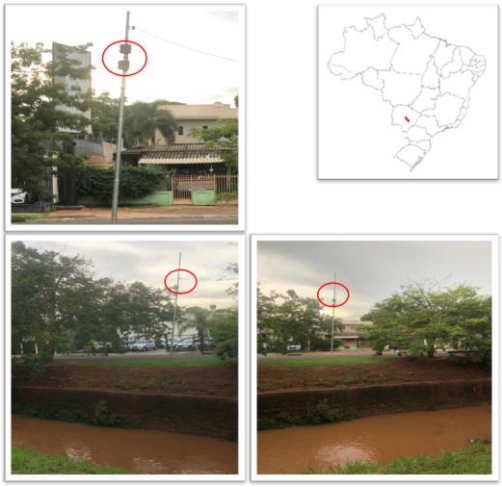


Figure 2. Prosa Stream, Campo Grande – Brazil.

6.2.2. Space-Time Correspondence Networks (STCN)

STCN was proposed to segment videos (sequence of frames) in order to take advantage of temporal information. Thus, the smooth visual change that occurs from one frame to another does not drastically affect the segmentation, as occurs when, for example, a segmentation method is trained in one season of the year and tested in another. In STCN, given the first labeled frame, the other frames are processed sequentially and feed a memory with relevant features. In this way, the feature memory is constantly updated and the abundant use of spatio-temporal information makes it possible to handle changes in appearance smoothly. The frames processed and stored in memory are called memory frames while the current frame to be segmented is called a query frame.

Initially, a frame is encoded in key and value features. In general, key features of the query frame are compared with those of the memory frames to determine when and where to extract relevant memory values. In this way, key features are used to search for relevant information and value features describe the visual characteristics of a region to produce the final segmentation. In STCN, key features are extracted through the RGB frame using ResNet50 while value features are extracted by ResNet18 using both the RGB frame and the mask.

Reading and encoding from memory occurs as follows. Given T memory frames and a query frame to be segmented, memory key k^M, memory value v^M and query key k^Q are extracted by ResNet50 or ResNet18. Then, the similarity $S_{ij} = c(k_i^M , k_j^Q)$ between each pair of keys k_i^M and k_j^Q is computed using the dot product and then softmax-normalized according to Equation 1 to obtain a similarity matrix W. Given W, the query frame features v^Q are computed as the weighted sum of the memory features, $v^Q = v^M.W$. Thus, v^Q have relevant information of the current frame extracted from the features stored in memory. Finally, v^Q are passed to the decoder in order to obtain the segmentation. The decoder gradually processes and upsamples

v^Q at a scale of 2 with a refinement module. The output of the last module is used to obtain the segmentation (mask) through a convolutional layer followed by the softmax function.

An important step of STCN is memory management. For each memory frame, memory stores the items: memory key and memory value. As the visual change in 30fps videos is small, the two items are only stored in memory every *S* frames. STCN was initially proposed with *S=5*, however, we evaluated the influence of *S* on memory management.

6.2.3. Dataset Structure

Dataset - STCN is basically composed of video files that vary in length time from seconds to hours. These videos were captured by different types of sensors (aerial and terrestrial), showing the different situations that a surface covered by water can be found. These situations oscillate in the simplest and calmest, such as streams and rivers on normal days, as well as intense precipitation events in streams and channels, which can cause floods. This dataset was built with the focus on the analysis of the Prosa stream because flood prevention measures are being studied and developed to help the population who lives in Campo Grande to help on decision making.

Figure 3. Examples of images from Dataset - STCN.

6.2.4. Experimental Setup

The Dataset consists of the STCN model methodology for video object segmentation (VOS). All videos were converted into frames per second (fps), and then only the first frame of each video sequence was labeled. From these first frames, considered as references, the model was applied to the following sequence of frames in order to delimit the boundaries of the water bodies. The hyperparameter memory value was varied from 5 to 15 to verify if these changes would influence the model performance.

Pixel accuracy and intersection over union (IoU) [28] were applied to check the performance of the methodologies. They compare the manually annotated measurements (Ground Truth) with those generated by the models (predicted). The pixel accuracy indicates the percentage of correctly classified pixels, contrasting true positives and true negatives [29]. The closer to value 1 is the pixels classified, the more correctly classified it is. The IoU metric calculates the ratio between the number of intersecting pixels of ground truth and predicted mask and the number of unified pixels of both masks.

6.3. RESULTS

The STCN dataset consists of 25 videos. To validate the STCN segmentation, the videos were divided according to the weather conditions and types of sensors used. In addition, the memory value hyperparameter S was varied to check if it would influence the model performance.

6.3.1. Fixed camera and good weather conditions

The videos used in this section were obtained from cell phones and fixed cameras monitoring water bodies such as rivers, springs, streams and lakes. This data was obtained in Germany, Brazil, and the United Kingdom, and to further validate the performance of this method, night time events were also analyzed to analyze the influence of luminosity on the segmentation. Table 1 and Table 2 show the segmentation results of the videos converted into frames, while Figure 4 shows the segmentation generated by the model on normal conditions.

Table 1. Results using pixel accuracy.

Pixel Accuracy	Memory Value			
	5/5		15/15	
	Background	Water	Background	Water
Video 1	0.989 ± 0.001	0.978 ± 0.002	0.991 ± 0.001	0.978 ± 0.002
Video 2	0.929 ± 0.007	0.989 ± 0.002	0.929 ± 0.007	0.989 ± 0.003
Video 3	0.978 ± 0.003	0.997 ± 0.002	0.978 ± 0.003	0.997 ± 0.002
Video 4	0.994 ± 0.001	0.962 ± 0.013	0.994 ± 0.001	0.967 ± 0.013
Video 5	0.994 ± 0.001	0.904 ± 0.007	0.995 ± 0.001	0.899 ± 0.007
Video 6	0.990 ± 0.001	0.965 ± 0.004	0.991 ± 0.001	0.968 ± 0.003
Video 7	0.977 ± 0.037	0.897 ± 0.019	0.978 ± 0.036	0.894 ± 0.020
Video 8	0.984 ± 0.001	0.992 ± 0.001	0.985 ± 0.001	0.992 ± 0.001
Video 9	0.974 ± 0.006	0.976 ± 0.002	0.978 ± 0.004	0.977 ± 0.002

Video 10	0.987 ± 0.003	0.985 ± 0.003	0.987 ± 0.003	0.986 ± 0.003
Video 11	0.977 ± 0.008	0.982 ± 0.006	0.977 ± 0.008	0.982 ± 0.006
Video 12	0.989 ± 0.006	0.998 ± 0.001	0.989 ± 0.006	0.998 ± 0.001
Video 13	0.954 ± 0.010	0.985 ± 0.004	0.955 ± 0.010	0.984 ± 0.004

Table 2. Results using IoU.

IoU	**Memory Value**			
	5/5		**15/15**	
	Background	**Water**	**Background**	**Water**
Video 1	0.960 ± 0.003	0.970 ± 0.002	0.962 ± 0.003	0.971 ± 0.002
Video 2	0.811 ± 0.020	0.984 ± 0.002	0.815 ± 0.022	0.984 ± 0.002
Video 3	0.970 ± 0.003	0.988 ± 0.001	0.970 ± 0.003	0.988 ± 0.001
Video 4	0.951 ± 0.015	0.957 ± 0.013	0.957 ± 0.015	0.962 ± 0.013
Video 5	0.979 ± 0.001	0.873 ± 0.005	0.979 ± 0.001	0.870 ± 0.006
Video 6	0.968 ± 0.003	0.951 ± 0.004	0.970 ± 0.002	0.954 ± 0.004
Video 7	0.902 ± 0.034	0.874 ± 0.029	0.900 ± 0.032	0.871 ± 0.028
Video 8	0.980 ± 0.002	0.967 ± 0.003	0.980 ± 0.002	0.967 ± 0.003
Video 9	0.965 ± 0.007	0.921 ± 0.014	0.968 ± 0.004	0.928 ± 0.009
Video 10	0.978 ± 0.004	0.965 ± 0.007	0.978 ± 0.004	0.966 ± 0.006
Video 11	0.953 ± 0.011	0.967 ± 0.008	0.953 ± 0.012	0.967 ± 0.008
Video 12	0.988 ± 0.006	0.979 ± 0.010	0.988 ± 0.006	0.979 ± 0.010
Video 13	0.936 ± 0.013	0.950 ± 0.010	0.936 ± 0.006	0.951 ± 0.010

	Original Image	5/5	15/15
Video 1			
Video 2			
Video 3			
Video 4			
Video 5			

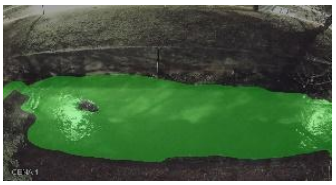
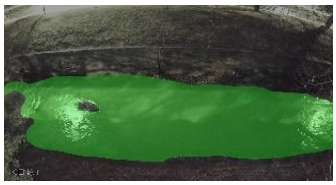
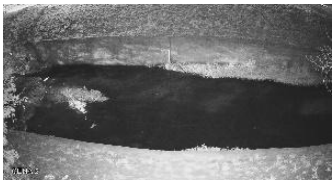
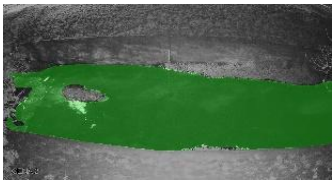
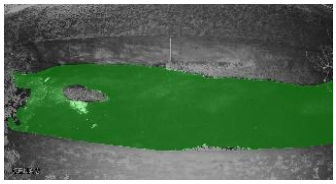
Video 6			
Video 7			
Video 8			
Video 9			
Video 10			
Video 11			
Video 12			
Video 13			

Figure 4. Water segmentation on normal conditions.

Tables 1 and 2 showed that the accuracy for both IoU and pixel accuracy were close to 90% for water and the standard deviation low, showing that the method was sufficient to delineate the water body. Analyzing Figure 4, it is possible to see that in all videos the segmentation was successful, indicating that for fixed cameras with conditions where not many climatic variations occur, the model can be applied successfully.

6.3.2. Fixed camera and turbulent water

The scenarios in this section are based on long-duration videos with the purpose of analyzing the change in water variation when precipitation occurs. Prosa stream in Campo Grande was the object of study, and thus high, and low-intensity precipitation was evaluated. Table 3 and Table 4 show the Pixel Accuracy and IoU validation metrics, respectively. Figure 5 shows the segmentation generated by the model.

Table 3. Results using pixel accuracy.

Pixel Accuracy	**Memory Value**			
	5/5		**15/15**	
	Background	**Water**	**Background**	**Water**
Video 14	0.974 ± 0.001	0.964 ± 0.008	0.975 ± 0.002	0.972 ± 0.003
Video 15	0.983 ± 0.003	0.996 ± 0.001	0.983 ± 0.003	0.996 ± 0.001
Video 16	0.980 ± 0.005	0.940 ± 0.017	0.979 ± 0.005	0.954 ± 0.014
Video 17	0.990 ± 0.005	0.894 ± 0.030	0.990 ± 0.005	0.912 ± 0.026
Video 18	0.984 ± 0.008	0.956 ± 0.020	0.985 ± 0.008	0.964 ± 0.017

Table 4. Results using IoU.

IoU	**Memory Value**			
	5/5		**15/15**	
	Background	**Water**	**Background**	**Water**
Video 14	0.947 ± 0.006	0.934 ± 0.008	0.954 ± 0.002	0.944 ± 0.003
Video 15	0.979 ± 0.003	0.979 ± 0.003	0.980 ± 0.003	0.979 ± 0.003
Video 16	0.938 ± 0.010	0.914 ± 0.012	0.947 ± 0.008	0.928 ± 0.010
Video 17	0.903 ± 0.025	0.884 ± 0.029	0.916 ± 0.022	0.902 ± 0.024
Video 18	0.937 ± 0.024	0.943 ± 0.022	0.945 ± 0.021	0.951 ± 0.019

	Original Image	5/5	15/15
Video 14			
Video 15			

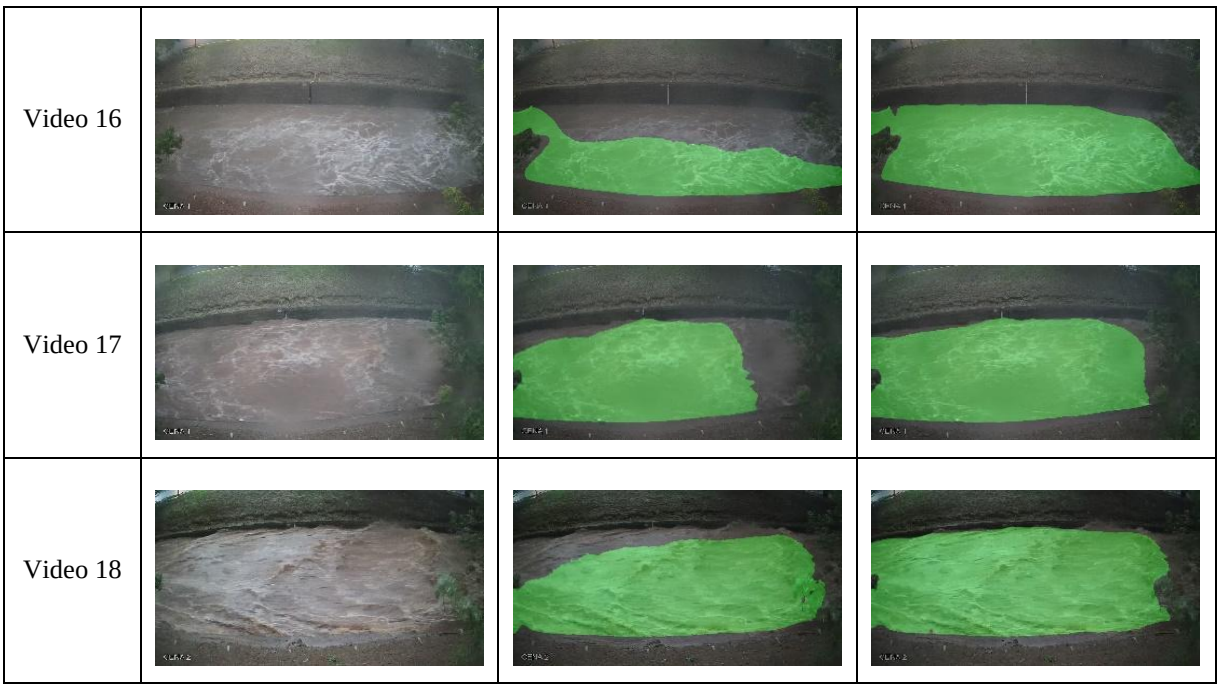


Figure 5. Water segmentation on fixed cameras and turbulent water.

If we analyze Tables 3 and 4, we notice that the hyperparameter of the memory value influenced the results of the metrics, where most of the values for water were better at S equals to 15/15 instead of 5/5. However, looking at Figure 5, it is possible to notice the difference in the contour boundary, especially in Videos 14, 16, 17 and 18, suggesting that the variation in the hyperparameter values can influence the model performance. Even with various adversities, such as variation in water level, precipitation, turbulent water, and other conditions, the model was able to delineate the water contours effectively.

6.3.3. UAV

UAVs were used due to the rotation and translation movements that such devices undergo and because of this expected the object of interest vary on each scene, thus hindering the performance of the segmentation performed by the model. It was also analyzed whether the object of interest being partially hidden (videos 19, 20 and 21) or appearing fully (videos 19a, 20a and 20a) would influence the results. All videos used are from rivers in Germany. The results are shown in Tables 5 and 6, and the results of the water body delineation in Figure 6.

Table 5. Results using pixel accuracy.

Pixel Accuracy	**Memory Value**			
	5/5		**15/15**	
	Background	**Water**	**Background**	**Water**
Video 19	0.917 ± 0.101	0.992 ± 0.005	0.968 ± 0.032	0.984 ± 0.014
Video 19a	0.989 ± 0.003	0.987 ± 0.009	0.990 ± 0.003	0.987 ± 0.010
Video 20	0.887 ± 0.097	0.999 ± 0.002	0.844 ± 0.133	0.998 ± 0.002
Video 20a	0.976 ± 0.012	0.851 ± 0.115	0.976 ± 0.012	0.851 ± 0.115
Video 21	0.719 ± 0.177	0.921 ± 0.099	0.721 ± 0.176	0.924 ± 0.094
Video 21a	0.944 ± 0.035	0.995 ± 0.003	0.958 ± 0.009	0.990 ± 0.007
Video 22	0.971 ± 0.041	0.942 ± 0.064	0.992 ± 0.003	0.943 ± 0.062

Table 6. Results using pixel accuracy.

IoU	Memory Value			
	5/5		15/15	
	Background	**Water**	**Background**	**Water**
Video 19	0.911 ± 0.100	0.896 ± 0.107	0.957 ± 0.036	0.941 ± 0.043
Video 19a	0.978 ± 0.009	0.968 ± 0.024	0.979 ± 0.009	0.970 ± 0.019
Video 20	0.886 ± 0.097	0.893 ± 0.065	0.844 ± 0.133	0.863 ± 0.081
Video 20a	0.878 ± 0.100	0.829 ± 0.111	0.878 ± 0.100	0.829 ± 0.111
Video 21	0.647 ± 0.163	0.787 ± 0.101	0.651 ± 0.163	0.790 ± 0.100
Video 21a	0.936 ± 0.034	0.966 ± 0.011	0.942 ± 0.013	0.966 ± 0.012
Video 22	0.949 ± 0.040	0.894 ± 0.075	0.970 ± 0.017	0.925 ± 0.063

	Original Image	5/5	15/15
Video 19			
Video 19a			
Video 20			
Video 20a			
Video 21			
Video 21a			

Video 22			

Figure 6. Water segmentation on UAV videos.

It is possible to analyze from the results that the accuracy decreases when using images obtained by UAVs. This is mainly due to the sensor not being totally stable at the time of data collection, being susceptible to rotational and translational movements, climatic factors, among other adversities. Furthermore, it is possible to note in Figure 6 that the initial frame used to train the model influenced the segmentation of rivers. When using a frame where the river only partially appears (Videos without "a"), it is possible to notice that many times the model was not able to differentiate the water from the background. The segmentation improvement occurs when using the first frame without the river being partially occluded (Video with "a"). This is perceived in Video 20a, showing that it can differentiate the bridge from the river, and in Video 21a, differentiating the water from the background. The unique experiment in which this did not occur was in Video 19a, showing that the model does not differentiate water from the background when the hyperparameter memory value was varied from 5 to 5, but the segmentation performed better on delineating the water contour when considered from 15 to 15.

6.3.4. Comparing STCN and a semantic segmentation model based on images

In order to seek applications for the generated segmentations, this section compares the results between the segmentation applying the STCN model and the SegNet model [29] adjusted with images from the Wesenitz River used in the current analysis. The latter study is based on a semantic segmentation model using images, and from the segmentation of water contour lines obtained by the SegNet model, it was possible to obtain automated water stages with highly reliable results. Thus, by comparing the results, it is possible to analyze whether the segmentation of the STCN model could be applied in future works related to obtaining information from water bodies. The comparison between the models should be considered because the STCN model is based on videos and SegNet is based on images.

Using a fixed position camera, images of the Wesenitz River were obtained over a one-year period in order to analyze the segmentation performance under different weather conditions. A video sequence was created from the images so that they could be analyzed in the STCN model. It is a challenging scenario for the STCN model because only one labeled image

was considered, and significant modifications occurred during the year. The results are presented in Table 7.

Table 7. Results based on different models.

	SegNet		STCN	
	Background	**Water**	**Background**	**Water**
IoU	0.977 ± 0.008	0.971 ± 0.014	0.975 ± 0.011	0.969 ± 0.012
Pixel Accuracy	0.986 ± 0.008	0.989 ± 0.003	0.992 ± 0.006	0.978 ± 0.013

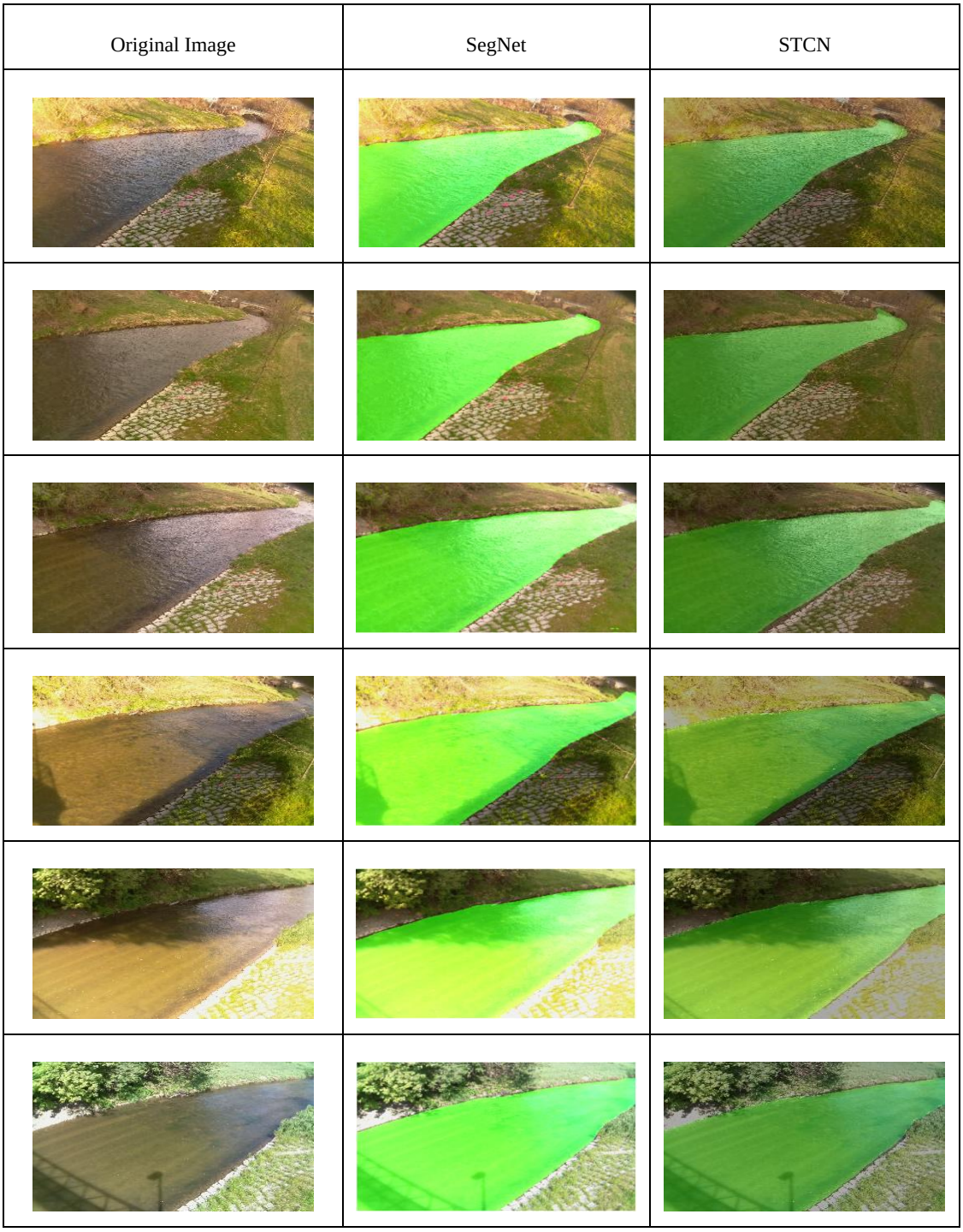


Figure 7. Comparing segmentation between SegNet and STCN.

The results and figures presented showed that the application of the STCN model for this study area also proved to be as effective as SegNet. Thus, the results generated in this work emerge as new alternatives for measuring water stages in an automated way in future works.

6.4. DISCUSSION

Comparing the results in different scenarios, fixed cameras with good weather conditions presented the highest accuracies (around 95% for IoU and pixel accuracy). Due to the stability of the sensor and little adversity, it is normal to expect such accurate results for this evaluation. Analyzing the agitated water, the model performance dropped a bit, but the results still remained good. The accuracies for this method varied around 90%, thus indicating that the model was able to segment the water effectively automatically. The biggest variation in results was when used UAVs and images where the first frame started with the water body partially hidden, where the worst result was 64%. Videos whose water appeared completely in the first frame presented a considerable improvement, thus showing that it is important to consider the issue of the object interest being partially hidden in the first captured frame.

The performance of the STCN model proved to be as accurate as the SegNet framework in segmenting the Wesenitz River even with time-lapsed images (not videos). Analyzing the validation metrics, both IoU and pixel accuracy indicated an accuracy higher than 96% thus, the obtained segmentation emerges as an alternative for automated water level measurement in future works.

The results also showed that storing the features every $S=15$ frames is sufficient to describe temporal information, unlike the original STCN proposal that used $S=5$. This result is relevant because the memory, with $S=15$, can store features for a longer period of time until it fills up and older features are discarded. In this way, a current frame can use features from an older frame, which are relevant to its segmentation and results in greater accuracy and IoU.

6.5. CONCLUSIONS

In this study, we investigated a semantic segmentation of water considering video structures. Our study target was to delineate water bodies in an automated way from images obtained by different types of sensors using the STCN model, considered state of the art in video object segmentation. STCN showed high performance for water feature segmentation tasks. However, when the video was obtained by aerial sensors, it showed greater difficulty in providing good predictions due to the UAVs move during the data acquisition. This fact is of great interest for future studies, which will have the challenge of better delimiting the performance of water bodies in UAV images.

The segmentation generated by STCN showed as reliable results as SegNet model. The SegNet performance depends on a large and robust dataset, requiring more time to obtain data, build an appropriate dataset, and annotate the labels of the classes. STCN, on the other hand, shows that the features extracted by this methodology evolve over time because of the hyperparameter memory value using only the first frame. The first frame is used to extract the characteristics of the river. From the second frame onwards, the model segments the frame and updates the characteristics of this river. Therefore, the attributes referring to the river are adapted. This reason shows that it is possible to use the STCN model for new challenges, such as continuous observations of water stages or even temporal information.

We hope to integrate the segmentation obtained with photogrammetry techniques for automated water stage estimation in future studies. In addition, it would be interesting to combine such information with sensors installed in streams, such as in this study in order to supervise the population in numerous types of sudden events, especially floods. As the advancement in technology occurs gradually, it is necessary that cities also follow in parallel this flow. In this way, we hope that in the future, we will be able to create innovative methodologies to become the cities smarter.

6.6. REFERENCES

1. Li, C.; Cheng, X.; Li, N.; Du, X.; Yu, Q.; Kan, G. A framework for flood risk analysis and benefit assessment of flood control measures in urban areas. *International journal of environmental research and public health* **2016**, 13, 787.

2. Mirza, M.M.Q. Climate change and extreme weather events: can developing countries adapt? *Climate policy* **2003**, 3, 233-248.

3. Yin, J.; Ye, M.; Yin, Z.; Xu, S. A review of advances in urban flood risk analysis over China. *Stochastic Environmental Research and Risk Assessment* **2015**, 29, 1063-1070.

4. Akiyama, T.S.; Junior, J.M.; Gonçalves, W.N.; de Araújo Carvalho, M.; Eltner, A. Evaluating different deep learning models for automatic water segmentation. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021; pp. 4716-4719.

5. Morgenschweis, G. *Hydrometrie: Theorie und Praxis der Durchflussmessung in offenen Gerinnen*; Springer-Verlag: 2010. p. 582.

6. Eltner, A.; Sardemann, H.; Grundmann, J. Flow velocity and discharge measurement in rivers using terrestrial and unmanned-aerial-vehicle imagery. *Hydrology and Earth System Sciences* **2020**, 24, 1429-1445.

7. Eltner, A.; Elias, M.; Sardemann, H.; Spieler, D. Automatic image-based water stage measurement for long-term observations in ungauged catchments. *Water Resources Research* **2018**, 54, 10,362-310,371.

8. Ridolfi, E.; Manciola, P. Water level measurements from drones: A pilot case study at a dam site. *Water* **2018**, 10, 297.

9. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature **2015**, 521 (7553), 436-444.

10. Goodfellow, I.; Bengio, Y.; Courville, A. Deep learning; MIT press: 2016.

11. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine **2017**, 5, 8-36.

12. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, 159, 296-307.

13. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing* **2019**, 152, 166-177.

14. Sun, X.; Wang, B.; Wang, Z.; Li, H.; Li, H.; Fu, K. Research progress on few-shot learning for remote sensing image interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, 14, 2387-2402.

15. Deng, J.; Li, X.; Fang, Y. Few-shot object detection on remote sensing images. *arXiv preprint arXiv:2006.07826* **2020**.

16. Alajaji, D.; Alhichri, H.S.; Ammour, N.; Alajlan, N. Few-shot learning for remote sensing scene classification. In Proceedings of the 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), 2020; pp. 81-84.

17. Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. *Advances in neural information processing systems* **2016**, 29.

18. Fu, Y.; Xiang, T.; Jiang, Y.-G.; Xue, X.; Sigal, L.; Gong, S. Recent advances in zero-shot recognition: Toward data-efficient understanding of visual content. I*EEE Signal Processing Magazine* **2018**, 35, 112-125.

19. Liu, B.; Yu, X.; Yu, A.; Zhang, P.; Wan, G.; Wang, R. Deep few-shot learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing* **2018**, 57, 2290-2304.

20. Zhang, T.; Sun, X.; Zhang, Y.; Yan, M.; Wang, Y.; Wang, Z.; Fu, K. A training-free, one-shot detection framework for geospatial objects in remote sensing images. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, 2019; pp. 1414-1417.

21. Cheng, H.K.; Tai, Y.-W.; Tang, C.-K. Rethinking space-time networks with improved memory coverage for efficient video object segmentation. *Advances in Neural Information Processing Systems* **2021**, 34.

22. Caelles, S.; Maninis, K.-K.; Pont-Tuset, J.; Leal-Taixé, L.; Cremers, D.; Van Gool, L. One-shot video object segmentation. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; pp. 221-230.

23. Fayyaz, M.; Saffar, M.H.; Sabokrou, M.; Fathy, M.; Klette, R.; Huang, F. STFCN: spatio-temporal FCN for semantic video segmentation. *arXiv preprint arXiv:1608.05971* **2016**.

24. Wang, Y.; Lyu, Y.; Cao, Y.; Yang, M.Y. Deep learning for semantic segmentation of UAV videos. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, 2019; pp. 2459-2462.

25. Bernhofer, C.; Schaller, A.; Pluntke, T. Starkregenereignisse von 1961 bis 2015. **2017**.

26. da Cruz-Silva, S.C.B.; Leonel, W.; da Silva, M.H.S.; Mercante, M.A. Dinâmicas de evolução do uso e ocupação da Região Urbana do Prosa, Campo Grande, MS: uma análise multitemporal. In Anais 5º Simpósio de Geotecnologias no Pantanal, 2014; p.661 -670.

27. CONSORCIO RES PLANEJAMENTO EM DRENAGEM URBANA. Plano diretor de drenagem urbana de Campo Grande – Sumário Executivo. 2009.

28. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015; pp. 3431-3440.

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

7. GENERAL CONCLUSIONS

This research investigated and proposed deep learning-based methods for water monitoring. Initially, the study consisted of delimiting water bodies in RGB images using the SegNet semantic segmentation method. Thus, the first dataset was created, consisting of many images of the same river obtained from fixed cameras. The SegNet model was trained, validated and tested for this dataset, proving to be efficient for the task of segmenting the studied river.

In order to seek applications for the generated segmentations, we proposed an image-based approach combining river segmentation and photogrammetry techniques to estimate water stages. The results showed high correlations between traditional and image-based measurement systems (between 87% - 93%), showing that water resource monitoring is possible only with image-based systems, emerging as a new way to obtain data, and possibly being a new tool to support traditional water resource monitoring methodologies.

Other methodologies were developed as the work progressed, such as fine-tuning, transfer learning and the influence of the number of images on the performance of the models, in order to not only segment the same river, but also images containing different types of water bodies for generalization tasks. A second dataset (Dataset 2) was built containing images of different water bodies, such as rivers, lakes, ponds, in order to improve the performance of the semantic segmentation models. All these studies were carried out searching for the best solution to identify water resources and facilitate other researches that pursue the same goal.

Finally, we used a deep learning method based on video structures, which consists in using only one labeled frame to test the model. It differs from the structures developed in previous studies, which were based on a dataset of images. The model used was the STCN, considered the state-of-the-art in video object segmentation. A dataset consisted of videos containing diverse water bodiers to test the model was built. This model showed high performance in segmenting water in videos, regardless whether the camera system is fixed or moving.

The major contribution that this study offers is the optimization of information concerning a body of water using techniques different from what traditional measurement systems are used to. In this way, we show that the development of science tends to offer new tracking options linked to water resources. The advancement of technologies linked to the Internet of Things (IoT), for instance, shows that more alternatives can be developed. A future solution would be to integrate everything that has been developed in this research into real-time surveillance systems, thus helping the population in numerous decisions, especially in hypothetical cases of flooding. This study also emerges as an alternative that different information from a body of water can be obtained from artificial intelligence, such as chemical

components, runoff and flow velocities prediction, among others. Therefore, we hope that in the future more studies related to our area can be developed in order to always contribute to the safety and welfare of society.

7.1. REFERENCES

ACHARYA, T. D.; SUBEDI, A.; LEE, D. H. Evaluation of machine learning algorithms for surface water extraction in a Landsat 8 scene of Nepal. **Sensors**, v. 19, n. 12, p. 2769, 2019.

ASSEM, H. et al. Urban water flow and water level prediction based on deep learning. Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2017, Springer. p.317-329.

ÇAYIR, A. N.; NAVRUZ, T. S. Effect of Dataset Size on Deep Learning in Voice Recognition. *In*: 3RD INTERNATIONAL CONGRESS ON HUMAN-COMPUTER INTERACTION, OPTIMIZATION AND ROBOTIC APPLICATIONS (HORA). IEEE, 2021 p.1-5.

CHEN, J.; HILL, A. A.; URBANO, L. D. A GIS-based model for urban flood inundation. **Journal of Hydrology**, v. 373, n. 1-2, p. 184-192, 2009.

CHOLLET, K. Keras. Disponível em: https://github.com/fchollet/keras. Acesso em 15 ago. 2019.

DETERT, M.; JOHNSON, E. D.; WEITBRECHT, V. Proof-of-concept for low-cost and non-contact synoptic airborne river flow measurements. **International Journal of Remote Sensing**, v. 38, n. 8-10, p. 2780-2807, 2017.

ELTNER, A.; SARDEMANN, H.; GRUNDMANN, J. Flow velocity and discharge measurement in rivers using terrestrial and UAV imagery. **Hydrol. Earth Syst. Sci. Discuss**, v. 2019, p.1-29, 2019.

GAWEHN, E.; HISS, J. A.; SCHNEIDER, G. Deep learning in drug discovery. **Molecular informatics**, v. 35, n. 1, p. 3-14, 2016.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015.

LI, C. et al. A framework for flood risk analysis and benefit assessment of flood control measures in urban areas. **International journal of environmental research and public health**, v. 13, n. 8, p. 787, 2016.

LI, Q. Literature survey: domain adaptation algorithms for natural language processing. **Department of Computer Science The Graduate Center, The City University of New York**, p. 8-10, 2012.

LIN, Tsung-Yi et al. Focal loss for dense object detection. *In:* PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION. 2017. p. 2980-2988.

LINJORDET, T.; BALOG, K. Impact of training dataset size on neural answer selection models. *In*: EUROPEAN CONFERENCE ON INFORMATION RETRIEVAL. Springer, Cham, 2019. p. 828-835.

MIRZA, M. M. Q. Climate change and extreme weather events: can developing countries adapt? **Climate policy**, v. 3, n. 3, p. 233-248, 2003.

MORGENSCHWEIS, G. Hydrometrie: Theorie und Praxis der Durchflussmessung in offenen Gerinnen; Springer-Verlag: 2010. p. 582.

OSCO, L. P. et al. A review on deep learning in UAV remote sensing. **International Journal of Applied Earth Observation and Geoinformation**, v. 102, p. 102456, 2021.

PAN, J. et al. Deep learning-based unmanned surveillance systems for observing water levels. **IEEE Access**, v. 6, p. 73561-73571, 2018.

REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. **arXiv preprint arXiv:1804.02767**, 2018.

REN, S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks. **Advances in neural information processing systems**, v.28. p91-99, 2015.

RIDOLFI, E.; MANCIOLA, P. Water level measurements from drones: A pilot case study at a dam site. **Water**, v. 10, n. 3, p. 297, 2018.

YIN, J. et al. A review of advances in urban flood risk analysis over China. **Stochastic environmental research and risk assessment**, v. 29, n. 3, p. 1063-1070, 2015.

ZHANG, C. et al. Understanding deep learning (still) requires rethinking generalization. **Communications of the ACM**, v. 64, n. 3, p. 107-115, 2021.

ZHU, X. X. et al. Deep learning in remote sensing: A comprehensive review and list of resources. **IEEE Geoscience and Remote Sensing Magazine**, v. 5, n. 4, p. 8-36, 2017.