DEEP LEARNING APPLIED TO WATER SEGMENTATION

T. S. Akiyama [1, *], J. Marcato Junior [1], W. N. Gonçalves [1,2], P. O. Bressan [2], A. Eltner [3], F. Binder [3], T. Singer [4]

[1] Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil - akiyama.thales@gmail.com, jose.marcato@ufms.br
[2] Faculty of Computer Science, Federal University of Mato Grosso do Sul, Campo Grande, Brazil – patrik.bressan@ifms.edu.br, wesley.goncalves@ufms.br
[3] Institute of Photogrammetry and Remote Sensing, Technische Universität Dresden, 01062 Dresden, Germany – felixmbinder@gmx.de, Anette.Eltner@tu-dresden.de
[4] Institute of Hydrology, Technische Universität Dresden, Dresden, Germany – thomas.singer@erfurt.de

Commission III, ICWG II/III

KEY WORDS: Convolutional Neural Network, Semantic Segmentation, Hydrology, Remote Sensing, Urban Floods

ABSTRACT:

The use of deep learning (DL) with convolutional neural networks (CNN) to monitor surface water can be a valuable supplement to costly and labour-intense standard gauging stations. This paper presents the application of a recent CNN semantic segmentation method (SegNet) to automatically segment river water in imagery acquired by RGB sensors. This approach can be used as a new supporting tool because there are only a few studies 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 a period of 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.

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 control the urban flood events.

Due to the increased capacity to evaluate data using computational resources, Zhu et al. (2017) reports 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.

---

* Corresponding author

To the knowledge of the authors, there are only 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 area from high-resolution imagery using DL approaches; Feng and Sester (2018) described a framework to collect, process and analyse pluvial flood relevant information from social media platform applying DL approaches on user generated texts and photos.

Such a monitoring tool could also be applied to potentially support real-time flood warning capabilities as the hardware could 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 labour-intense standard gauging stations.

The main aim of this paper is to automatically segment river water in RGB imagery using 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 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.

2. METHODOLOGY

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 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 labelled 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 labelled images from the dataset.

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

The decoder network is composed by convolutional and a set of upsampling layers, and the memorized max-pooling indices from the encoder feature map(s) are used to upsample 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 a batch normalization is used. The detail preservation can be valuable to delineate what is water area and background with good accuracy. At 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 to the most likely class at each pixel.
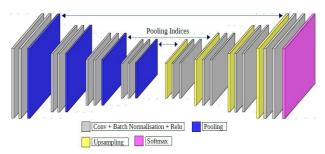
Figure 3. SegNet architecture composed of encoder and decoder. The encoder extracts a low-resolution feature map and the decoder upsamples it to obtain a pixel-wise classification. Source: adapted from Badrinarayanan et al. (2017).

2.3 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 on 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, which defines how the weights are adjusted in the CNN, and to estimate 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.

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 4.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 the 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.

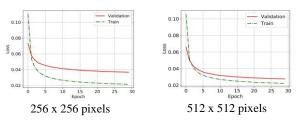| 256 x 256 pixels | 512 x 512 pixels |

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 the 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.

Resolution	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)

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

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.

Resolution	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)

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

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.
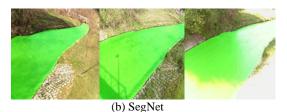

(a) RGB

(b) SegNet

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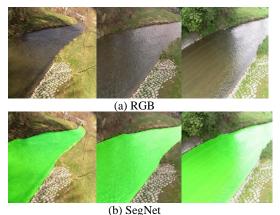

(a) RGB

(b) SegNet

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

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 becoming a new source of information.

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.