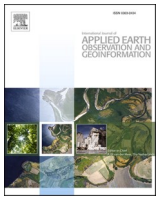
International Journal of Applied Earth Observations and Geoinformation 107 (2022) 102682

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



A learning approach for river debris detection

Àlex Solé Gómez, Leonardo Scandolo, Elmar Eisemann

Delft University of Technology, the Netherlands

ARTICLE INFO

Keywords:
Segmentation
Hyperspectral Imaging
Machine learning

ABSTRACT

Plastic pollution in the sea is an environmental hazard, negatively impacts marine life, and causes economic damage all over the world. It is estimated that each year 8 million tonnes of plastic are deposited in seas, the vast majority coming from rivers. In recent years, publicly available satellite imagery has been used to attempt to track floating plastic debris, using specialized hand-crafted features. In this work, we present an automatic learning approach based on satellite imagery that can detect floating plastic debris in rivers with high precision. This approach is based on well-proven image segmentation architectures, U-Net (RonnebeRger et al., 2015) and DeeplabV3+ (Chen et al., 2018), which we adapt to process high-dimensional multispectral images. To train and test the approach, we also present a dataset of images from different rivers around the world containing floating plastic debris, which is a key step to creating an automated learning solution. We test the predictive accuracy of our network, showing that our approach can correctly identify floating debris in images from regions not seen in the training set. Our results also show that a more extensive labeled dataset is necessary to generalize the approach to some types of rivers. Furthermore, we also demonstrate how our solution can also be used to monitor single areas over time to understand and predict floating debris accumulation.

1. Introduction

Plastic pollution in oceans and rivers is one of the most pressing environmental problems nowadays. The production of plastics reached 359 million tons in 2018 (PlasticsEurope, 2019), and by 2050 there will be a yearly production of 2 billion tons (United Nations Environment Programme, 2016). It is estimated that roughly 8 million tons of plastic are deposited at sea each year, which constitutes 80% of the total ocean waste (United Nations Environment Programme, 2017). The costs associated with the problems caused by this pollution to the fishing industry, wildlife, and tourism are predicted to be at least 8 billion dollars (United Nations Environment Programme, 2017). When plastic reaches the water, depending on its composition, part of it sinks and part of it remains afloat (Biermann et al., 2020). Plastic gradually breaks down in the water and becomes *microplastics*, which are often ingested by marine life, and are much more difficult to extract and recycle (Biermann et al., 2020). If current plastic pollution trends continue, it is estimated that by 2050 there will be more plastic than fish in the oceans and that 99% of seabirds will have ingested plastic (United Nations Environment Programme, 2017). One large contributor to ocean plastics are rivers that carry garbage as they cross cities and ultimately deposit it at sea (Lebreton et al., 2012; Cózar et al., 2015). Cleaning up this garbage is a complicated task, and the first step towards it is identifying garbage patch locations so that it can be cleaned up and recycled.

In the past, fishermen looked at the reflection of the water in the clouds to find areas containing shoals of fish, as they have different reflectance properties than seawater. More recently aerial and satellite imagery allow us to use the same principle but at a larger scale for the purpose of detecting floating plastic and debris. Jakovljević et al. (2019) developed an algorithm to detect garbage from the Drina river using neural networks and high-resolution satellite images (ground sample distance of 46 cm) from WorldView-2 (European space agency, 2009). They study several areas of the river with large amounts of garbage and show a methodology for using Google Earth Engine (Gorelick et al., 2017) for the preparation and preprocessing of the data. Garaba et al. (2018), from The Ocean Cleanup Foundation, collected airborne imagery from the Pacific ocean and studied the spectral fingerprint of different types of debris. Their approach confirms that wavelengths around 1215 nm and around 1732 nm have potential use for plastic detection applications using multispectral images, which coincides with other publications (Jakovljević et al., 2019; Garaba and Dierssen, 2018; Martínez-Vicente et al., 2019; Serranti et al., 2018). Garaba and Dierssen (2018) extracted macro- and microplastics from the ocean to study its composition and reflective properties, noting that typical floating plastic debris has a varied composition and is not an exact match for any single polymer source, although it still exhibits high absorption characteristics in the infrared band. Martínez-Vicente et al. (2019) explore the requirements for a satellite sensor platform that can detect marine plastic debris, and state that the NIR and SWIR bands are the most relevant to marine plastic debris detection. Serranti et al. (2018) analyze marine debris collected by surface-trawling plankton nets with a hyperspectral sensor and use the SWIR wavelengths to characterize different polymer

https://doi.org/10.1016/j.jag.2022.102682
Received 29 July 2021; Received in revised form 29 December 2021; Accepted 9 January 2022
Available online 19 January 2022

types. Hu (2021) conducted a theoretical and practical study on the requirements and potentials of using the visible range and near-infrared range for the detection of marine debris. They theoretically determine the minimum proportion of subpixel coverage necessary for the detection of microplastics based on the sensitivity and the SNR of the sensor, suggesting that microplastic detection is technically impossible using all existing sensing platforms, but that macroplastic detection is possible. They also conclude that separating macro plastics and other types of debris is very difficult given their very similar spectrum, although marine debris is usually a mixture of different materials, so differentiation may not be necessary. Topouzelis et al. (2019) explore the usage of a lower resolution satellite with global coverage, Sentinel-2, for the detection of debris in coastal areas. Near the beach of Tsamakia on the island of Lesbos (Greece), they deployed floating platforms filled with different types of plastics in order to study the results when captured by an aerial sensing platform and the Sentinel-2 satellite mission. This is part of the University of the Aeagan's Plastic Litter Project (University of the Aegean Marine Remote Sensing Group, 2018; University of the Aegean Marine Remote Sensing Group, 2019; University of the Aegean Marine Remote Sensing Group, 2020). They repeated the experiment for three years and presented a dataset detailing the dates and coordinates of their platforms. Fig. 1 shows these platforms as seen from Sentinel-2, and we can observe that in the Near-Infrared Band (NIR) band, we can visualize the platforms better than in the RGB bands. Their work demonstrates the different spectral responses corresponding to different plastic types, as well as the percentage of area coverage within the footprint of a satellite image. Recent work by Biermann et al. (2020) shows the feasibility of using Sentinel-2 for the detection of garbage and other floating materials. Their approach is to select specific Sentinel-2 pixels of areas suspected of containing floating plastic debris found by monitoring news and social media. The satellite imagery pixels are then validated by analyzing their spectral footprint and classified using a naïve Bayes classification algorithm. They present a new feature called *Floating Debris Index* (FDI), which is based on the *Floating Algae Index* (FAI) (Hu, 2009). This new feature makes use of the linear interpolation between the NIR band and its two contiguous bands, the Red Edge 2 (RE2) and the Short Wave Infrared Band 1 (SWIR1). The purpose of the FDI is to improve the NIR band via subtracting the linear interpolation of its contiguous bands in order to minimize sensitivity to atmospheric changes and be able to detect floating objects through thin clouds.

Images from satellite sensors capture reflected light in visible and non-visible frequency ranges on an almost global scale with high periodicity and resolution. In particular, Sentinel-2 is a European Space Agency (ESA) (European Space Agency, 2015a) mission consisting of two satellites that capture nearly all global inland regions, coastal areas, and the Mediterranean Sea every two to five days. Sentinel-2 captures 12 spectral bands, ranging from the visible spectrum to shortwave infrared. Each wavelength is captured at different spatial resolutions, from 10 to 60 meters per pixel. In this paper, we present a learning-based image segmentation approach that is able to identify floating plastics and other debris in rivers present in Sentinel-2 imagery. We first create a dataset of labeled images from various rivers and urban areas that we use for training. We compiled and manually labeled such collection based on previous work (Jakovljević et al., 2019), and news reports of heavily polluted rivers. We use this dataset to test and validate several state-of-the-art neural network architectures typically used for natural image segmentation and classification. Additionally, we propose several architectural modifications in order to accommodate the increase in input channels and improve precision, resulting in high segmentation accuracy that can be used to identify new areas with floating debris accumulation or to monitor specific regions known to accumulate debris.

2. Materials and methods

In this section, we will first introduce our dataset, which comprises manually labeled images of rivers from the Sentinel-2 mission. This dataset was compiled to be used for training and testing different neural network architectures designed for the purpose of multispectral image segmentation. We will then detail the neural network architectures we consider for the segmentation task and details regarding their training. We tested two well-known and high-performing architectures for classification and segmentation: U-Net (Ronneberger et al., 2015; Roy et al., 2019) and Deeplab V3+ (Chen et al., 2018). We additionally propose and evaluate a modified U-Net architecture, which we named *U-Net3DE*, and uses 3D convolutions along the encoding path in order to exploit contextual spectral information.

2.1. Sentinel-2 satellite imagery

Sentinel-2 is a terrestrial observation mission from the European Space Agency (ESA) (European Space Agency, 2015a) that consists of two satellites, named S2A and S2B, phased 180° that orbit around the earth synced with the sun. It has a coverage of almost all inland areas between 59° South to 84° North and all coastal waters around them, as well as the Mediterranean Sea, and revisits areas near the equator every 5 days, and mid-latitude areas every 2–3 days. It is designed for tasks such as forest monitoring, agricultural fields monitoring, and managing natural disasters. Its sensors offer 12 bands ranging from the visible spectrum to near-infrared and short wave infrared at different resolutions depending on the band, and slightly different wavelengths per-satellite (see Table 1).

Previous works (Garaba et al., 2018; Jakovljević et al., 2019; Garaba and Dierssen, 2018; Martínez-Vicente et al., 2019) point at wavelengths around 1215 nm and around 1732 nm as the most useful to identify patches of plastic and debris. SWIR1 is closest to 1732 nm and has a resolution of 20 meters. Bands B9, B8A, B8 are close to 1215 nm, with a resolution of 60, 20, 10 meters respectively. Band B8 is especially useful because it displays high reflectance properties for garbage when compared to the RGB spectrum (Topouzelis et al., 2019). Further, it has a good spatial resolution.

The Sentinel-2 catalog offers two types of images with different corrections, Level 1C imagery and Level 2A imagery. Level-1C imagery applies radiometric and geometric corrections to the images, including

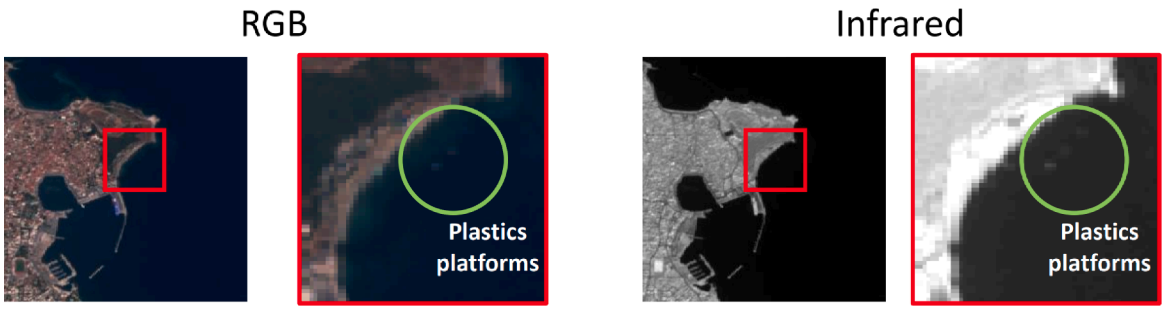


Fig. 1. Sentinel-2 images from the platforms installed in Lesbos (Greece) as part of the Plastic Litter Project (Topouzelis et al., 2019).

Table 1
Sentinel-2 bands with their respective resolution, wavelength for each satellite (S2A and S2B), and brief description.

Band	Wavelength (S2A/S2B)	Resolution	Description
B1	443.9 nm/ 442.3 nm	60 meters	Aerosols
B2	496.6 nm/ 492.1 nm	10 meters	Blue
B3	560.0 nm/ 559.0 nm	10 meters	Green
B4	664.5 nm/ 665.0 nm	10 meters	Red
B5	703.9 nm/ 703.8 nm	20 meters	Red Edge 1
B6	740.2 nm/ 739.1 nm	20 meters	Red Edge 2
B7	782.5 nm/ 779.7 nm	20 meters	Red Edge 3
B8	835.1 nm/ 833.0 nm	10 meters	NIR
B8A	864.8 nm/ 864.0 nm	20 meters	Red Edge 4
B9	945.0 nm/ 943.2 nm	60 meters	Water vapor
B11	1613.7 nm/ 1610.4 nm	20 meters	SWIR 1
B12	2202.4 nm/ 2185.7 nm	20 meters	SWIR 2

orthocorrection. This process unifies the resolution of the band images in order to have the same resolution for all bands. Further, it provides masks for clouds and land/water, and compresses images using JPEG2000. Level 2-A imagery atmospherically corrects the Level-1C imagery with Sen2Cor (Main-Knorn et al., 2017), and also creates a scene classification map. Atmospheric correction consists of removing artifacts caused by the atmosphere but needs to be carefully considered, as it can modify or delete the spectral fingerprint of some materials. We used level 2-A images in our dataset because in our tests the atmospheric correction provided by level 1-C images yielded no improvements over the original data.

2.2. Generation of the datasets

To access and process Sentinel-2 images, we used Google Earth Engine (Gorelick et al., 2017), which is a platform for geospatial analysis and provides tools to create scripts for visualization and classification. To the best of our knowledge, there is no publicly available labeled dataset of satellite imagery of floating plastics or other debris. Previous works (Garaba et al., 2018; Biermann et al., 2020) provide insights into the spectral signature of plastics in seas and oceans and mention analyzed areas, which provides a good starting point to locate debris in multispectral images.

We focused our efforts on classifying debris in rivers and near-coastal areas, since they are covered by Sentinel-2. In particular, we spent considerable effort finding areas with large and visible pollution patches for manual labeling. Although large amounts of floating plastic debris exist in many rivers, it is not always possible to identify it manually through Sentinel-2 imagery. Most debris accumulation happens in the wake of flooding, as flooded rivers next to garbage dumps can carry large amounts of debris. Nevertheless, these areas typically remain clouded for days, which can prevent satellite imagery from being useful. Another problem for manual labeling is the low resolution of Sentinel-2 imagery in some regions, which makes debris detection difficult, as was the case in Accra, Ghana (Chasant, 2020). Based on various articles, news, and publications found via hashtags in social media, we identified areas with garbage content large enough to be detected in Sentinel-2 imagery, and which we could confidently determine as floating garbage and not any other phenomenon.

2.3. Dataset locations

Our dataset primarily consists of satellite imagery from three rivers, the Drina River, the Los Angeles River, and the Yangtzee River. The Drina River is located in the Balkans and borders between Serbia, Bosnia, and Herzegovina. River level rises cause garbage from adjacent landfills to be dragged downstream (Jakovljević et al., 2019; Associated Press, 2020). There is a net installed along the river upstream from the Višegrad Hydroelectric Power Plant to prevent garbage from continuing downstream. The garbage it accumulates is visible on Sentinel-2 imagery (see Fig. 2, left). The Los Angeles River in Los Angeles, USA uses special floating screens (Barboza, 2020) to prevent debris from reaching the sea (see Fig. 2, middle). The accumulation of debris in the net is visible in Sentinel-2 imagery. The Yangtze River in China is considered one of the most polluted rivers and one of the major polluters of the oceans at its mouth in the East China Sea. The largest hydroelectric dam in the world, the Three Gorges Dam, is located near Sandouping, Yichang, Hubei province and acts as a retaining wall for a lot of the floating debris in the river (Three Gorges dam 'could be blocked by rubbish', 2020). Zhang et al. (2015) demonstrated the presence of a large number of microplastics in the Three Gorges Dam basin. According to Hu (2021), it is not possible to detect microplastics with the sensitivity of the Sentinel-2 sensors. However, the presence of microplastics is a strong indicator that there is a nearby area where macroplastics are breaking down. We identified dates when considerable amounts of debris are present (see Fig. 2, right).

Our dataset (see Table 2) contains images from selected dates in the previously described areas, where debris is visible, and several dates when there is no visible debris. The surrounding areas from the Drina River and the Yangtze River consist mostly of vegetation, whereas the Los Angeles River is surrounded mainly by urban and industrial areas. In order to provide further examples of urban environments, we added images of the industrial and the harbor areas from San Francisco, as well as images from the center of Barcelona. We followed a labeling methodology based on three classes: water, debris, and *other*. This last label comprises soil, forest, city, vehicles, etc.

Previous works (Garaba et al., 2018; Biermann et al., 2020) have shown that the spectral reflectance of plastic materials has a minimum around 945 nm (B9 in Sentinel-2) and a maximum around 1613.7 nm (B11 in Sentinel-2), and plastics, wood, and seaweed have a maximum in the B8 band. These findings aided our manual labeling efforts and can be seen in the debris spectrum of our dataset (Fig. 3). Identified debris pixels have a maximum around the B8 and B8A band, a minimum in the B9 band, and another maximum in the B11 band. Furthermore, in the Yangtze and Drina River images, there is a sharp increase at the B5 band (703.9 nm) in pixels classified as *other*, which is expected as vegetation reflects light above $\sim 700nm$ (Biermann et al., 2020; Myneni et al., 1995). In contrast, the Los Angeles River images show high reflectance in almost all bands, caused by highly reflective urban construction materials. Fig. 4 shows an example of the band values for a single image of the Drina River.

We followed the methodology described by Biermann et al. (2020) to manually label our dataset. First, we monitored news publications, hashtags, and keywords to find articles or social networks posts that show regions suitable for our work (Jakovljević et al., 2019; Associated Press, 2020; SCPR, 2019; Trash Accumulates at Three Gorges Dam, 2020; Three Gorges dam 'could be blocked by rubbish', 2020). We then matched the dates and the locations with sentinel imagery, and once a region with debris was established, we revised historical data on the same region to obtain further samples. For the validation of the garbage patches, we used Snap software to check the spectra, as well as well-known index values (FDI, NDVI, and NDWI), aided also by available reference values corresponding to similar publications on the topic (Biermann et al., 2020; Garaba et al., 2018; Jakovljević et al., 2019; Garaba and Dierssen, 2018; Martínez-Vicente et al., 2019).

2.4. Neural networks for image segmentation

In recent years, the use of neural networks for image classification tasks has grown immensely, spearheaded by the work of LeCun et al. (1998). Simonyan and Zisserman (2015) introduced the 16 layer Visual Geometry Group (*VGG16*) architecture for image classification and identification tasks. Long et al. (2015) extended the work from Simonyan et al. and created the Fully Convolutional Network (*FCN*). This network was developed for image segmentation purposes and was able to predict images instead of vectors. Ronneberger et al. (2015)

Fig. 2. Sentinel-2 images of debris at the Drina River (left), the Los Angeles River (center), and the Three Gorges Dam along the Yangtze River in China (right).

Table 2
Labeled dataset used for our experiments. Used labels are debris (D), water (W) and *other* (O).

Region	Coordinates	Date	Label	Usage
Drina River	43.756455, 19.273323	15/01/2019	D/W/O	Training
		09/02/2019	D/W/O	
		24/02/2019	W/O	
		01/03/2019	D/W/O	
		31/03/2019	D/W/O	
		25/04/2019	D/W/O	
		14/06/2019	D/W/O	
		06/07/2019	D/W/O	
		22/10/2019	W/O	
		20/03/2020	D/W/O	
		04/05/2020	D/W/O	
		28/06/2020	W/O	
		03/07/2020	W/O	
		08/07/2020	W/O	
Yangtze River	30.828716, 111.008711	04/08/2018	D/W/O	Training
		14/08/2018	D/W/O	
		19/08/2018	D/W/O	
		03/09/2018	D/W/O	
		25/07/2019	D/W/O	
		28/09/2019	D/W/O	
		09/07/2019	D/W/O	
		15/04/2020	W/O	
		10/05/2020	W/O	
Barcelona	41.390528, 2.149071	17/05/2020	O	Training
San Francisco	37.757826, −122.378724	04/05/20	W/O	Training
Los Angeles	33.764164, −118.205894	09/01/2019	D/W/O	Test
		24/01/2019	D/W/O	
		05/12/2019	D/W/O	
		04/01/2020	D/W/O	
		29/03/2020	W/O	

developed a network for medical image segmentation called *U-Net*, which was inspired by the *FCN* architecture, with the addition of skip layers in all the pooling layers. U-Net is one of the best-known architectures for image segmentation tasks.

Zhou et al. (2019, 2018) modified U-Net using modern backbones such as the Residual Neural Network (*Resnet*) (He et al., 2016), Dense Convolutional Network (*Densenet*) (Huang et al., 2017), *Inception* (Szegedy et al., 2015) or *Xception* (Chollet, 2017).

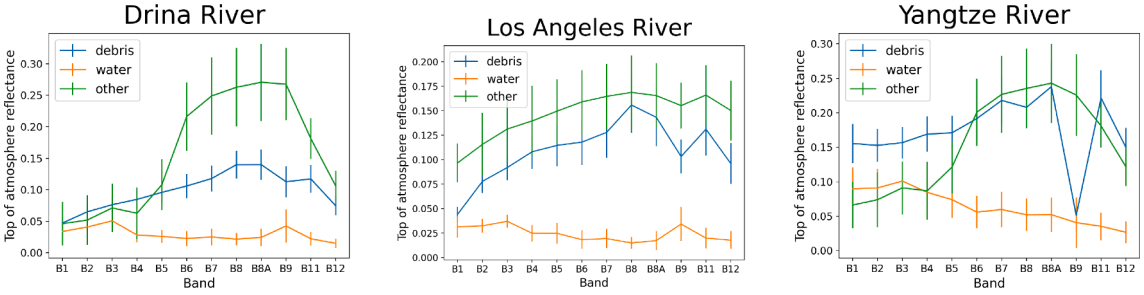
Other architectures for image segmentation have been proposed, such as the Pyramid Scene Parsing Network (*PSPNet*) (Zhao et al., 2017). This approach exploits pyramid pooling modules to improve segmentation results. Following the same idea, Chen et al. proposed a convolutional network architecture for image segmentation containing fully-connected conditional random fields called DeeplabV3+ (Chen et al., 2018). DeeplabV3 + makes use of the pyramid to have multiple representations of the same features, as well as *atrous* convolutions. DeeplabV3 + achieves sharper segmentation results and is currently one of the state-of-the-art approaches for image segmentation.

2.5. Tested Network architectures

We tested three different architectures in the task of floating debris detection in our described dataset (Table 2): U-Net, U-Net3DE, and Deeplab V3+.

U-Net (Ronneberger et al., 2015) is one of the most widely used network architectures for image classification and segmentation tasks. It is based on the Fully Convolutional Neural Network (FCN) (Long et al., 2015) architecture, with the addition of a skip layer between each pooling layer and a transposed convolution upsample path.

Our proposed U-Net3DE architecture is based on U-Net. We modified the original U-Net architecture, similar to the work of Roy et al. (2019, 2020), to use 3D convolutions as their work shows that using 3D convolutional layers improves classification-task results over using separate per-band convolutions, as filters can access and encode information on neighboring spectral bands. The encoder stage of our U-Net3DE uses 3D convolutional layers with a 2D decoder for segmentation. At each skip layer we reshape the output-feature cubes to 2D feature maps by concatenating the output of the 2D convolutional transposed layer. At



(a) Drina River. (b) Los Angeles River. (c) Yangtze River.

Fig. 3. Top of atmosphere spectral values from the different regions of our dataset according to their classification.

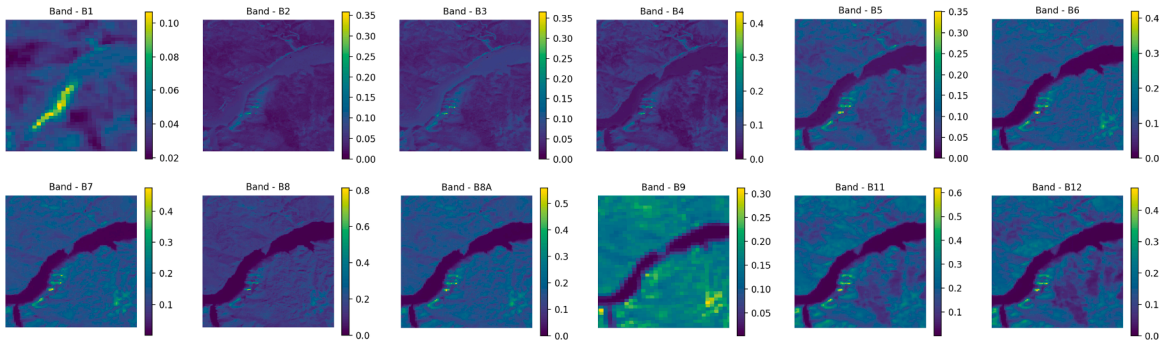


Fig. 4. Top of atmosphere reflectance of the different bands of a Drina River image of our dataset, from 20/03/20.

the end of the encoder we also reshape the output cubes to 2D feature maps and use these as an input to the decoder stage. For the pooling layers, we implemented a 3D pooling that halves the height, length, and number of bands. Since Sentinel-2 has 12 bands, the 3D pooling layers are only implemented in the first two pooling layers. Past the second pooling layer, we only have three bands and the pooling layers only halve the height and the length of the cubes, not the number of bands. Although this modified architecture leads to slightly fewer parameters than a conventional U-Net network for this task (see Table 3), the training is slower given the larger convolutional kernels. Fig. 5 shows an overview of the U-Net3DE architecture and a detailed view of the encoder.

Deeplab V3+ (Chen et al., 2018) is the current state-of-the-art architecture for semantic segmentation. Hoeser and Kuenzer (2020) tested different deep learning architectures for multispectral and hyperspectral images and concluded that DeeplabV3 + reaches the best results for segmentation tasks. This architecture is the latest update from the DeepLab family (Chen et al., 2017) and it adds an encoder-decoder structure like U-Net instead of using naïve decoder architectures like the rest of the DeepLab family. This change in the architecture refines the results, especially for label boundaries. Fig. 6 shows the implementation used for our experiments.

We tested two backbone implementations for DeepLab V3+: Xception (Chollet, 2017) and MobileNet V2 (Sandler et al., 2018). According to the work of Hoeser and Kuenzer (2020), Xception leads to better results as a backbone. Nevertheless, we also tested MobileNet V2 since it contains significantly fewer parameters, which leads to faster training and can prevent overfitting given limited data.

2.6. Network Training

Data augmentation was performed on the input dataset in order to obtain an extended training set and improve the robustness of the learned features. Standard data augmentation transforms were used, such as affine transforms, noise addition, and blurring and sharpening operations. As the Sentinel-2 dataset regularly contains defective (completely null) pixels, such pixels were also randomly added as an augmentation technique. We assign black regions a special class with null weight in the loss function in order to ignore them during training, which is also employed to handle potential black pixels arising from affine transforms due to zero padding outside the image boundaries.

An issue with our dataset stems from the large class imbalance, as most pixels are labeled as water and *other*, but comparatively few are labeled as debris. To solve this problem, we use a weighted loss function, which emphasizes results for underrepresented classes in the dataset. Our loss function is therefore a weighted cross entropy:

$$\mathscr{L} = -\sum_{n=0}^{3} \omega_i y_i log(S_i) \tag{1}$$

$$\omega_i = \begin{cases} 0 & if \quad i = 0 \\ \dfrac{Total\_Samples}{Samples\_class \cdot 3} & if \quad i \neq 0 \end{cases} \tag{2}$$

In Eq. (1), ω_i is the weight for each class, y_i is the ground truth label, and S_i represents the output from a softmax function. Eq. (2) shows the formula used to compute class weights. Class 0 corresponds to black pixels, which do not contribute to the loss value. For the rest of the classes, we compute their weight as the inverse ratio of class pixels to the total amount of pixels, scaled by $1/3$ to have a similar magnitude.

Based on the work of Choi et al. (2019), we use an Adam optimizer for our experiments due to its fast convergence time and low validation error. The Adam parameters used for our experiment were learning rate $0.001, \beta_1 = 0.9, \beta_2 = 0.999$, and $\epsilon = 10^{-7}$. Due to memory limitations, we used a batch size of 4 images.

Table 3
Trainable parameters from each architecture.

Network	Trainable Parameters
U-Net	31,113,188
U-Net3DE	29,105,972
Deeplab V3 + Mobilenet V2 backbone	2,111,780
Deeplab V3 + Xception backbone	41,053,636

3. Results

The network training was performed on an Intel i7-3820 with 16 Gb of RAM and an Nvidia GTX 980 Ti GPU, as well as using the Google Colab (Bisong, 2019) service. All training and evaluation code was written with the Tensorflow API for Python.

We trained and tested the different architectures introduced in Section 2.5, using the dataset presented in Table 2. As such, our results show the classification performance of each architecture for the Los Angeles River images, which is an area that is not present in the training set. Table 4 shows the mean Intersection over Union (IoU) and the IoU results for each class for all the different architectures. The IoU value for a label consists of the ratio between the intersection of the pixels with a specific label in the segmented image and the ground truth image, over the union of the pixels with that label in both versions. We achieve the best results in terms of garbage detection IoU when using U-Net3DE and Deeplab V3 + using the Xception backbone (DV3X). All architectures achieve similar results in terms of water and *other* pixel detection. This suggests that U-Net and Deeplab V3 + using the Mobilenet backbone (DV3M) are a better choice for water detection tasks in remote sensing images, since they are faster to train due to the lower number of trainable parameters, as noted in Table 3. U-Net3DE was the slowest network to train, but interestingly, it contains less trainable parameters than DeeplabV3 + or U-Net, since parameters are shared between layers when performing 3D convolutions. Its 3D convolutions are also the cause of the slower training performance because they have a large

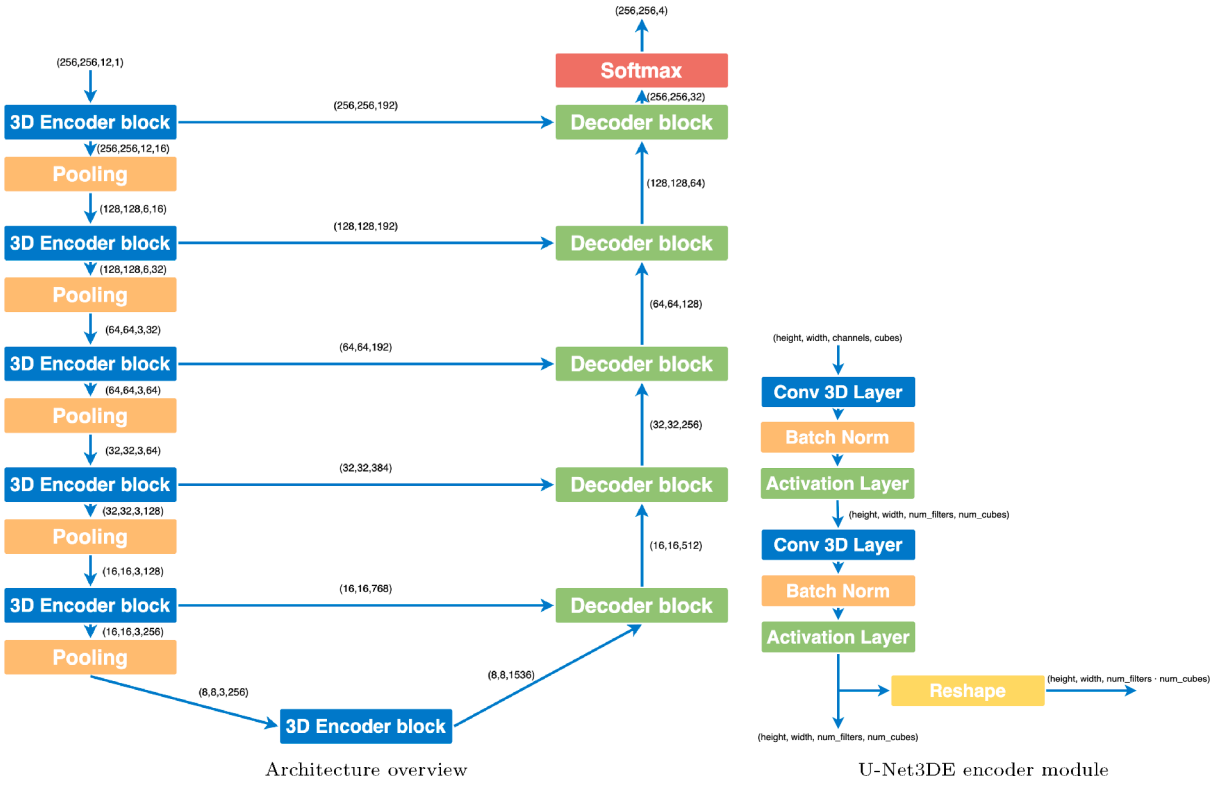


Fig. 5. U-Net3DE architecture (left) and a detailed view of the encoder module (right).

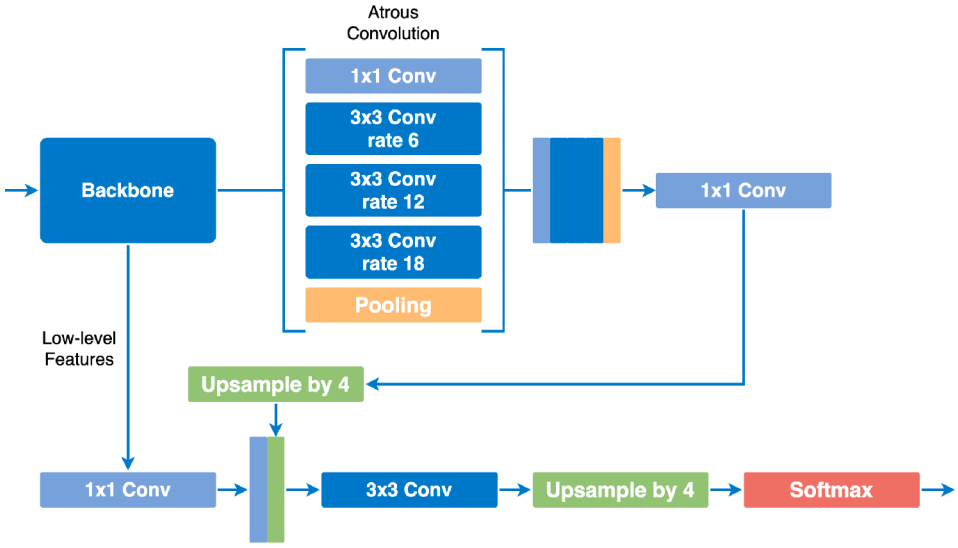


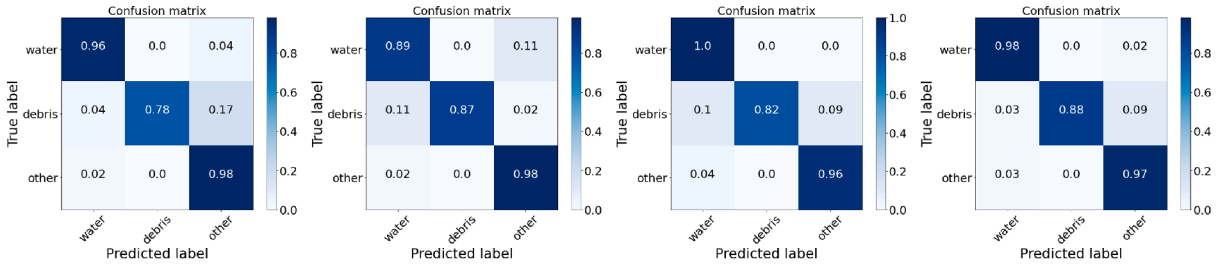
Fig. 6. DeeplabV3 + Architecture.

Table 4
Intersection-over-union test results for the different architectures tested, using the dataset as described in Table 2.

	IoU water	IoU debris	IoU *other*	Mean IoU
U-Net	0.84	0.29	0.96	0.69
U-Net3DE	0.87	**0.61**	**0.97**	**0.82**
Deeplab V3 + Mobilenet V2 backbone (DV3M)	0.81	0.26	0.96	0.68
Deeplab V3 + Xception backbone (DV3X)	**0.89**	**0.61**	**0.97**	**0.82**

memory footprint and are not as well optimized in our framework as 2D convolutions.

Fig. 7 shows the confusion matrix for the different architectures, providing further insight into their per-class performance. In all cases, we achieve roughly 80% to 90% accuracy for debris pixels and very high prediction performance for the remaining two labels. These results are affected by the class imbalance in our dataset. The debris class has a low representation in our dataset, so mislabeling even a few pixels can greatly alter the confusion matrix, which partly explains the lower correct classification percentage for such pixels. For the case of DV3M, water pixels are sometimes also more commonly mislabeled as *other* pixels. This is a common issue in satellite image classification, especially



(a) DeeplabV3 Xception (b) DeeplabV3 Mobilenet V2 (c) U-Net (d) U-Net3DE

Fig. 7. Confusion matrices for all the tested architectures for the Los Angeles River test set.

in the presence of shadows and buildings (Xu, 2006).

Fig. 8 shows the RGB channels from the input image, the ground truth, and the predicted labels for the different architectures tested in a subset of the test set corresponding to the Drina River. DV3X shows the best results in terms of similarity with the ground truth. In all scenes with no debris pixels, classification works correctly and edges between labels are sharp. DV3M shows that using the Mobilenet backbone results in less sharp edges when compared to the Xception backbone, e.g., the two bridges that cross the river in Fig. 8 and spurious debris labels assigned to pixels in the lower-left corner of the image. This explains the reduced performance of DV3M for water pixels since it tends to mislabel water pixels in the river banks as *other* pixels. U-Net3DE shows an improvement over U-Net, mainly due to more accurate predictions at edges between debris and *other* labeled pixels, as well as avoiding misclassification of pixels in the lower-left part of Fig. 8, which are highlighted with red circles. We observed that dark areas due to shadows corresponding to the *other* label are sometimes classified as water in both approaches; this is highlighted with blue circles in Fig. 8. In practice, confusion between water and *other* pixels is not of great importance in our task, and can additionally be corrected by using freely available geographic databases, such as open OpenStreetMap (Haklay, 2010).

Additionally, we trained and tested a more conventional fully connected multi-layer perceptron for the same task. Such an architecture provides no implicit spatial ordering of the pixels of the image. As such, multilayer perceptrons are typically outperformed by convolution-based networks in image tasks, for which convolutional networks can exploit the image structure and regularity. We observed similar results for our use case, as seen in Fig. 9. The multilayer perceptron architecture can mostly distinguish water from *other* labeled pixels, but it misclassifies a large part of the image as debris. The used architecture contained two fully connected layers, the first one with 512 neurons and the second one with 1024 neurons for a total of 535,043 trainable parameters. The number of layer quantity and their sizes were determined using hyperparameter tuning.

In Fig. 9, we also showcase the results of the method proposed by Biermann et al. (2020), which relies on a naïve Bayes classifier that uses the FDI and NDVI values of the image pixels for classification. We trained the classifier using our dataset, balancing classes to avoid overfitting. This approach resulted in an IoU score of 0.3, 0.0023, and 0.378 for the water, debris, and *other* pixels respectively, indicating that the handcrafted indices are not a good fit for this particular classification

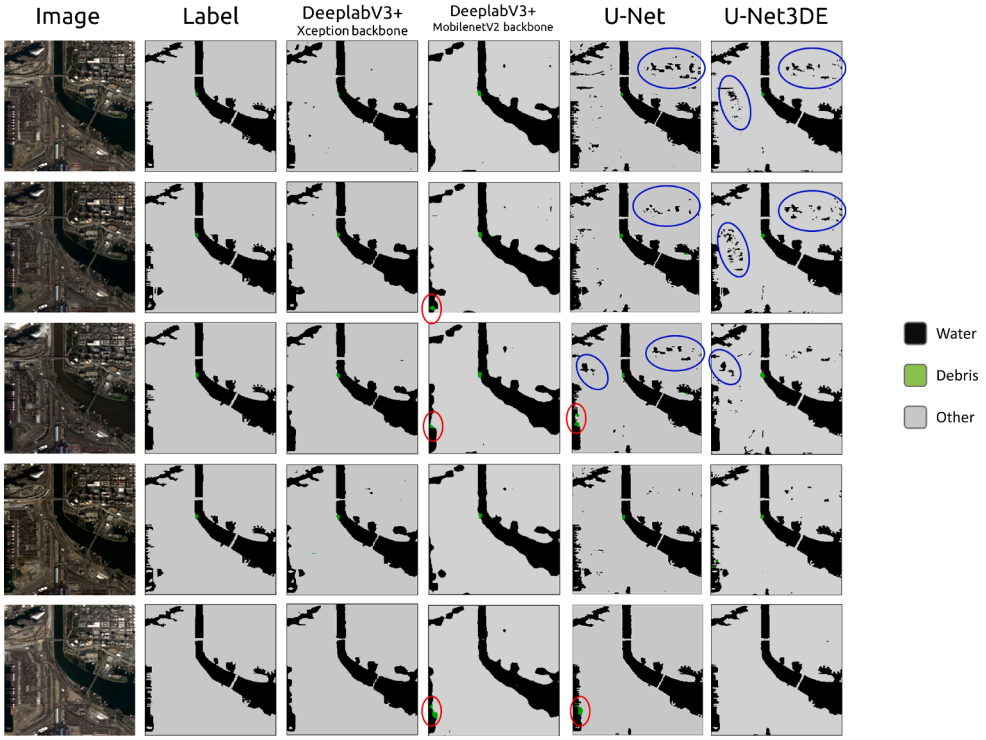


Fig. 8. Test images for the different architectures tested. Red circles show areas with a considerable amount of false positive debris predictions, and blue circles show areas with a considerable amount of false positives water predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

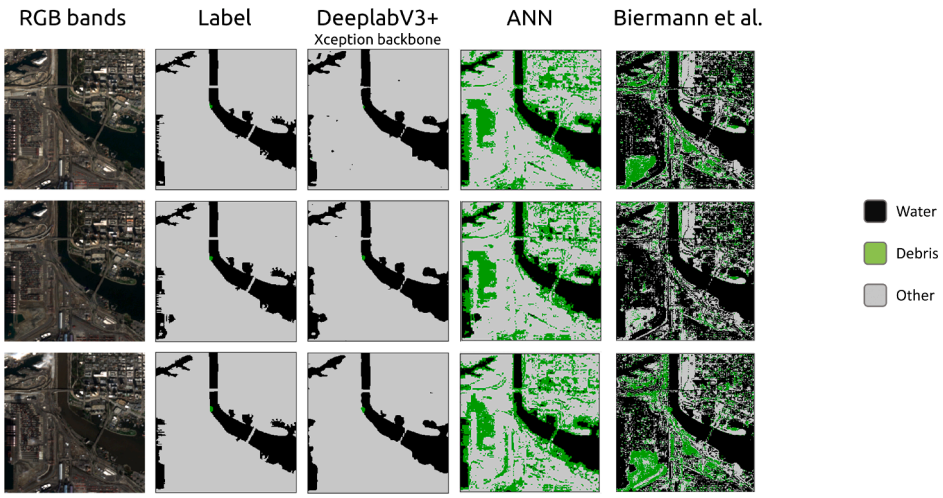


Fig. 9. Segmentation results when using DV3X, a fully connected multilayer perceptron (ANN), and the method proposed by Biermann et al. (2020) based on the floating debris index (FDI). Both the ANN and the Biermann et al. approaches achieve significantly less accurate results than DV3X.

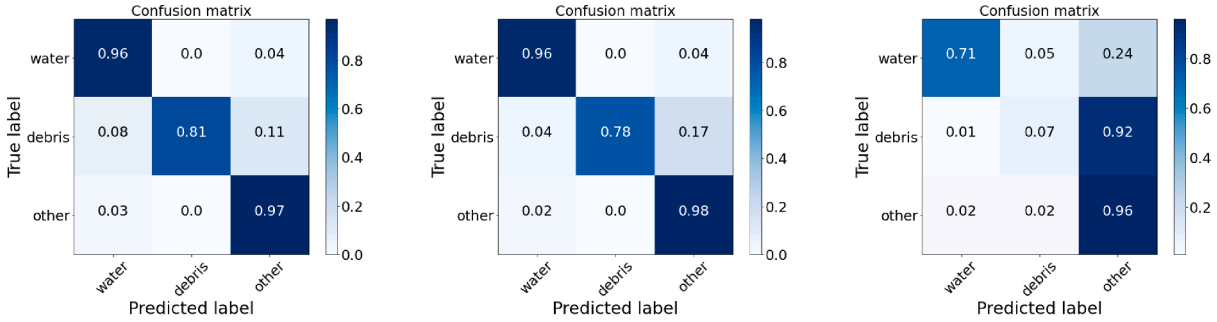
problem.

3.1. Cross-validation

We also performed cross-validation of our dataset to test classification performance for the different regions. The DV3X architecture was used for these experiments since they achieved the best prediction rates in our previous tests (see Table 4). Fig. 11 shows the visual results of this experiment, and Fig. 10 shows the respective confusion matrices. We obtain similarly good results for the Los Angeles River and Drina River if used as test sets: roughly 0.5 debris IoU and 0.8 of debris accuracy. The first three columns from Fig. 11 show that the detection is almost perfect for the Drina River test set, except for the case of the fourth image, where prediction accuracy is low. We believe the latter result is caused by weather conditions, the mountains are covered in snow, which is highly reflective, a phenomenon that is not present in the training dataset for that region. Conversely, for the case of the Yangtze River our approach largely failed at correctly predicting debris pixels, as we obtain a 0.02 IoU and 0.07 accuracy. The prediction accuracy for the debris label was low due to misclassification as *other* pixels. In this case, the nature of the river is different from the Drina and Los Angeles River, being much deeper and carrying more sediment. This also affects the nature and reflectivity of the accumulated debris, as shown in Fig. 3. This can be seen even in the RGB images in Fig. 11, where the third and fourth images correspond to times where a large amount of sediment is present in the water south of the dam resulting in a brown appearance to the naked eye. Correspondingly, the network incorrectly assigns the *other* label to most of the water pixels. We believe that such errors stem from the unique nature of the Yangtze River, i.e., very large discharge and sediment presence. These can be overcome by expanding our training set, or, as future work, using more sophisticated data augmentation procedures that can recreate such characteristics. However, given the high accuracy results for the Drina River and the Los Angeles River image sets, we believe that our approach generalizes well to typical urban and rural rivers.

3.2. Area monitoring

As a final evaluation of real-world applications, we tested the performance of our system for one of our test areas over the complete period during which Sentinel-2 has been in service. This would be the typical use case for monitoring an area or to obtain statistics and predict debris accumulation around the year. For this task, we selected the Drina River region for evaluation and used all the labeled images from our dataset for training, plus four additional training images to cover the various seasons. These extra training images correspond to slightly cloudy days that were not present in the original set, and examples of autumn landscapes, where vegetation is mostly absent. For our test set, we use all other Sentinel-2 images from the same region, except those with heavy cloud cover. As seen in Fig. 12, the segmentation result shows a recurrent annual pattern, where debris accumulation is higher at the beginning of the year and quickly decreases as the garbage net in the region is cleared. The figure also allows us to extract the encountered maximum amount of debris. This maximum can be verified with the corresponding RGB image that shows a heavy accumulation in the area, even though it is partially hidden under clouds. Fig. 12 also shows examples of moderate debris accumulation and no debris accumulation from different years in the area.



(a) Drina River as test. (b) Los Angeles River as test. (c) Yangtze River as test.

Fig. 10. Cross validation confusion matrices using the DV3X architecture.

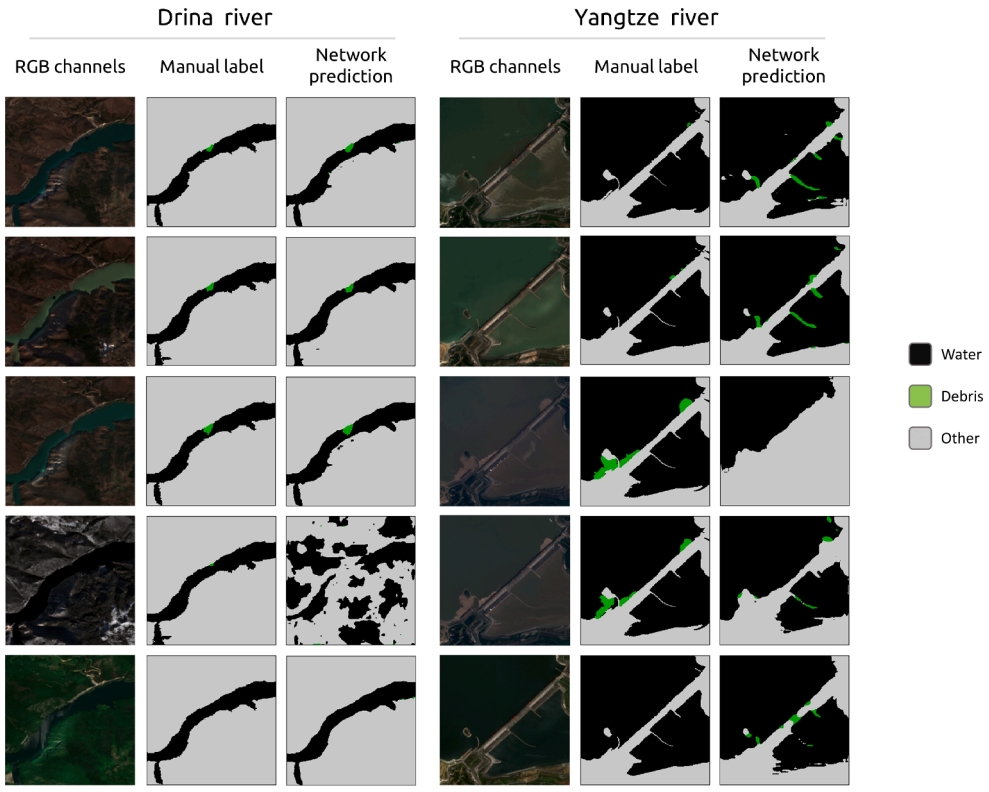


Fig. 11. RGB channels, ground truth from manual labeling, and prediction of part of our cross-validation tests when using the Drina River (left) and Yangtze River (right) as test dataset.

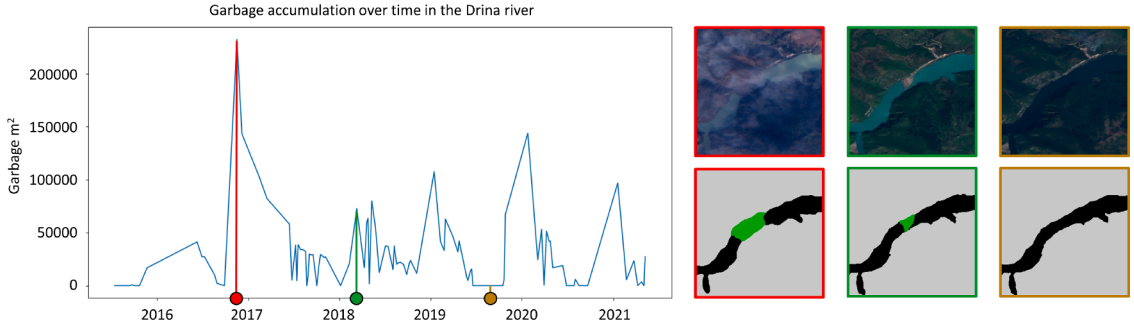


Fig. 12. Left: Debris area detected over time across all Sentinel-2 images of the Drina River with no heavy cloud cover. Right: RGB bands and segmentation result corresponding to the times indicated in the graph.

4. Conclusion

We have presented a learning-based method to segment satellite imagery to identify accumulated surface debris in rivers. Our dataset was compiled from the Sentinel-2 dataset and, based on previous work, we manually labeled pixels with a high likelihood of corresponding to man-made debris that has accumulated in river shores. We tested several state-of-the-art CNN architectures, identifying DV3X and our proposed U-Net3DE architecture as the best performing, with roughly 0.8 mean IoU across all labels. Moreover, cross-validation tests show that, for a large part of our dataset, such network architectures can successfully identify debris despite being trained with satellite images from different regions, meaning that our approach can generalize well from only a few exemplars to different environments. As such, our solution can be used to monitor or detect debris accumulation using publicly available satellite imagery. Nevertheless, cross-validation results for the Yangtze River, however, expose that a larger training database should be created, such that exemplars from more varied bodies of water and weather conditions are available. Given that convolutional approaches can lead to spatial overfitting, i.e., learning to classify based on the spatial disposition of training pixel labels rather than content, future work can explore how to mix convolutional and purely spectral approaches to improve classification performance. Furthermore, future work can also focus on using the presented method with aerial imaging or data from other remote sensing platforms, such as the upcoming Sentinel satellite missions (European Space Agency, 2015b), whose increased spatial and spectral resolutions should lead to better classification results.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially funded by the Netherlands Prize for ICT Research.

References

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV), pp. 801–818.

PlasticsEurope, Plastics - the Facts, https://www.plasticseurope.org/en/resources/publications/1804-plastics-facts-2019, accessed: 2020-09-06, 2019.

United Nations Environment Programme, Marine Plastic Debris and Microplastics, URL https://www.un-ilibrary.org/content/publication/0b228f55-en, 2016.

United Nations Environment Programme, UN declares war on ocean plastic, https://www.unenvironment.org/news-and-stories/press-release/un-declares-war-ocean-plastic, accessed: 2020-09-06, 2017.

Biermann, L., Clewley, D., Martinez-Vicente, V., Topouzelis, K., 2020. Finding plastic patches in coastal waters using optical satellite data. Scientific reports 10 (1), 1–10.

Lebreton, L.-M., Greer, S., Borrero, J.C., 2012. Numerical modelling of floating debris in the world's oceans. Marine pollution bulletin 64 (3), 653–661.

Cózar, A., Sanz-Martín, M., Martí, E., González-Gordillo, J.I., Ubeda, B., Gálvez, J.Á., Irigoien, X., Duarte, C.M., 2015. Plastic accumulation in the Mediterranean Sea. PloS one 10 (4), e0121762.

Jakovljević, G., Govedarica, M., Taboada, F.Á., 2019. Remote sensing data in mapping plastics at surface water bodies. Conference: FIG Working Week.

European space agency, WorldView-2 satellite mission, https://earth.esa.int/eogateway/missions/worldview-2, accessed: 2020-03-03, 2009.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment 202, 18–27.

Garaba, S.P., Aitken, J., Slat, B., Dierssen, H.M., Lebreton, L., Zielinski, O., Reisser, J., 2018. Sensing ocean plastics with an airborne hyperspectral shortwave infrared imager. Environmental science & technology 52 (20), 11699–11707.

Garaba, S.P., Dierssen, H.M., 2018. An airborne remote sensing case study of synthetic hydrocarbon detection using short wave infrared absorption features identified from marine-harvested macro-and microplastics. Remote Sens. Environ. 205, 224–235.

Martínez-Vicente, V., Clark, J.R., Corradi, P., Aliani, S., Arias, M., Bochow, M., Bonnery, G., Cole, M., Cózar, A., Donnelly, R., et al., 2019. Measuring marine plastic debris from space: initial assessment of observation requirements. Remote Sensing 11 (20), 2443.

Serranti, S., Palmieri, R., Bonifazi, G., Cózar, A., 2018. Characterization of microplastic litter from oceans by an innovative approach based on hyperspectral imaging. Waste Manage. 76, 117–125.

Hu, C., 2021. Remote detection of marine debris using satellite observations in the visible and near infrared spectral range: Challenges and potentials. Remote Sens. Environ., vol. 259, 112414, ISSN 0034–4257, doihttps://doi.org/10.1016/j.rse.2021.112414, https://www.sciencedirect.com/science/article/pii/S0034425721001322.

Topouzelis, K., Papakonstantinou, A., Garaba, S.P., 2019. Detection of floating plastics from satellite and unmanned aerial systems (Plastic Litter Project 2018). Int. J. Appl. Earth Obs. Geoinf. 79, 175–183.

University of the Aegean Marine Remote Sensing Group, Plastic Litter Project, https://mrsg.aegean.gr/?content=&nav=55, accessed: 2020-09-06, 2018.

University of the Aegean Marine Remote Sensing Group, Plastic Litter Project, https://mrsg.aegean.gr/?content=&nav=65, accessed: 2020-09-06, 2019.

University of the Aegean Marine Remote Sensing Group, Plastic Litter Project, https://mrsg.aegean.gr/?content=&nav=88, accessed: 2020-09-06, 2020.

Hu, C., 2009. A novel ocean color index to detect floating algae in the global oceans. Remote Sens. Environ. 113 (10), 2118–2129.

European Space Agency, Sentinel-2 mission, https://sentinel.esa.int/web/sentinel/missions/sentinel-2, accessed 2020-09-06, 2015a.

Roy, S.K., Krishna, G., Dubey, S.R., Chaudhuri, B.B., 2019. HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. 17 (2), 277–281.

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for sentinel-2, in: Image and Signal Processing for Remote Sensing XXIII, vol. 10427, International Society for Optics and Photonics, 1042704, 2017.

Chasant, M., 2020. Plastic pollution in Ghana: urban trash heroes, https://www.muntaka.com/plastic-pollution-in-ghana/, accessed: 2020-08-10, 2020.

Associated Press, Garbage clogs once crystal clear Bosnia rivers amid neglect, Inquirer https://technology.inquirer.net/86266/garbage-clogs-once-crystal-clear-bosnia-rivers-amid-neglect, accessed: 2020-03-03.

Barboza, T., 2020. Tons of Los Angeles river trash will be captured before it hits the sea, Los Angeles Times https://latimesblogs.latimes.com/lanow/2011/11/massive-la-river-trash-capturing-project-completed.html, accessed: 2020-03-03.

Three Gorges dam 'could be blocked by rubbish', The Telegraph https://www.telegraph.co.uk/news/worldnews/asia/china/7922348/Three-Gorges-dam-could-be-blocked-by-rubbish.html, accessed: 2020-03-03.

Zhang, K., Gong, W., Lv, J., Xiong, X., Wu, C., 2015. Accumulation of floating microplastics behind the Three Gorges Dam. Environ. Pollut. 204, 117–123.

Myneni, R.B., Hall, F.G., Sellers, P.J., Marshak, A.L., 1995. The interpretation of spectral vegetation indexes. IEEE Trans. Geosci. Remote Sens. 33 (2), 481–486.

S.C.P.R. (SCPR), https://www.scpr.org/programs/take-two/2019/02/04/19395/, accessed: 2020-09-06, 2019.

Trash Accumulates at Three Gorges Dam, The Wall Street Journal https://www.wsj.com/articles/SB10001424052748704271804575404460453360890, accessed: 2020-03-03.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

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

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging 39 (6), 1856–1867.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. in: Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, 3–11, 2018.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890.

Roy, S.K., Chatterjee, S., Bhattacharyya, S., Chaudhuri, B.B., Platoš, J., 2020. Lightweight spectral–spatial squeeze-and-excitation residual bag-of-features learning for hyperspectral classification. IEEE Trans. Geosci. Remote Sens. 58 (8), 5277–5290.

Hoeser, T., Kuenzer, C., 2020. Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. Remote Sensing 12 (10), 1667.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence 40 (4), 834–848.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. Mobilenetv 2: Inverted residuals and linear bottlenecks, in. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520.

Choi, D., Shallue, C.J., Nado, Z., Lee, J., Maddison, C.J., Dahl, G.E., 2019. On empirical comparisons of optimizers for deep learning, arXiv preprint arXiv:1910.05446.

Bisong, E., 2019. Google colaboratory, in: Building Machine Learning and Deep Learning Models on Google Cloud Platform, Springer, 59–64, 2019.

Xu, H., 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International journal of remote sensing 27 (14), 3025–3033.

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environment and planning B: Planning and design 37 (4), 682–703.

European Space Agency, Sentinel-2 mission, https://https://sentinels.copernicus.eu/web/sentinel/missions/sentinel-3, accessed 2020-09-06, 2015b.