International Journal of Applied Earth Observation and Geoinformation 126 (2024) 103624

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

DeepAqua: Semantic segmentation of wetland water surfaces with SAR imagery using deep neural networks without manually annotated data

Francisco J. Peña [a,b,*], Clara Hübinger [a], Amir H. Payberah [b], Fernando Jaramillo [a]

[a] Department of Physical Geography, Stockholm University, Svante Arrhenius väg 8, Stockholm, 106 91, Sweden
[b] Division of Software and Computer Systems, KTH Royal Institute of Technology, Kistagången 16, Kista, 164 40, Sweden

ARTICLE INFO

Dataset link: https://github.com/melqkiades/deep-wetlands

Keywords:
Deep learning
Semantic segmentation
Remote sensing
Wetland mapping
Vegetated water
Automated data labeling
CNN

ABSTRACT

Deep learning and remote sensing techniques have significantly advanced water surface monitoring; however, the need for annotated data remains a challenge. This is particularly problematic in wetland detection, where water extent varies over time and space, demanding multiple annotations for the same area. In this paper, we present DeepAqua, a deep learning model inspired by knowledge distillation (a.k.a. teacher–student model) to generate labeled data automatically and eliminate the need for manual annotations during the training phase. We utilize the Normalized Difference Water Index (NDWI) as a teacher model to train a Convolutional Neural Network (CNN) for segmenting water from Synthetic Aperture Radar (SAR) images. To train the student model, we exploit cases where optical- and radar-based water masks coincide, enabling the detection of both open and vegetated water surfaces. DeepAqua represents a significant advancement in computer vision techniques for water detection by effectively training semantic segmentation models without any manually annotated data. Experimental results show that DeepAqua outperforms other unsupervised methods by improving accuracy by 3%, Intersection Over Union by 11%, and F1-score by 6%. This approach offers a practical solution for monitoring wetland water extent changes without the need of ground truth data, making it highly adaptable and scalable for wetland monitoring.

1. Introduction

Wetlands provide essential ecosystem services such as water purification, flood regulation, and carbon sequestration, and are critical for sustainable development (Jaramillo et al., 2019). They are, however, increasingly threatened by climate change and human activity (Thorslund et al., 2017). The comprehensive monitoring of wetland surface water extent, including both open and vegetated water surfaces, is essential for their conservation and management.

The recognition of water surfaces in wetlands is usually achieved by combining optical and Synthetic Aperture Radar (SAR) imagery. While optical imagery helps recognize open water surfaces, SAR can additionally identify some water surfaces covered by vegetation. Combining optical and SAR imagery with deep learning improves the semantic segmentation of wetlands to map their water extent (Jamali et al., 2022; Jamali and Mahdianpari, 2022). Semantic segmentation refers to the classification of different parts of an image (e.g., water and non-water surfaces). However, annotating the data required for training deep learning models is often time-consuming and costly. Hence, there is a need to develop deep learning models that identify water surfaces without manually annotated data.

In this paper, we present DeepAqua, a deep learning model that eliminates the need for manual annotation during the training phase and that is inspired by the concept of knowledge distillation (a.k.a. teacher–student model). Knowledge distillation is the process of transferring knowledge from a large model to a smaller one. Utilizing the Normalized Difference Water Index (NDWI) (McFeeters, 1996) as the "teacher" model, we train a "student" U-Net (Ronneberger et al., 2015) to recognize water boundaries in SAR images. Using the signal from non-vegetated water as a guide, the student model learns from the teacher model, eventually recognizing both open and vegetated water bodies. The NDWI "teacher" model can generate annotated images of water surfaces without training. We aim to eliminate the costs of collecting training data through fieldwork or manual annotations of water on satellite imagery. In contrast to the traditional knowledge distillation paradigm, where knowledge is transferred from a large neural network to a smaller one, our model transfers the knowledge from a thresholding method like NDWI into a neural network, using NDWI as a replacement for manual annotations.

We test our model in three wetlands in Sweden with open water and water surfaces covered by grassy and floating vegetation. We use

* Corresponding author at: Department of Physical Geography, Stockholm University, Svante Arrhenius väg 8, Stockholm, 106 91, Sweden.
E-mail address: francisco.pena@natgeo.su.se (F.J. Peña).

https://doi.org/10.1016/j.jag.2023.103624
Received 19 May 2023; Received in revised form 8 December 2023; Accepted 12 December 2023
Available online 19 December 2023

C-band SAR to detect these covered water surfaces and map the total surface water extent. The model could be even adjusted to work with SAR sensors with longer wavelengths, such as L-band, to detect water under thicker canopies (e.g., mangroves).

In summary:

- We present DeepAqua, a novel method for training semantic segmentation models inspired by knowledge distillation without manually annotated data.
- We employ NDWI masks in our method as proxies for semantic labels and optical images as an auxiliary modality to supervise a SAR-based U-Net.
- We present DeepAqua as a highly adaptable and scalable model, as it does not require ground truth for training.
- Experimental results on three temperate wetlands in Sweden show that DeepAqua has superior performance of the semantic segmentation of water from SAR images when compared to other unsupervised methods.
- DeepAqua can monitor changes in surface water coverage, encompassing both in open water surfaces and those covered by vegetation.

2. Related work

The challenge of accurately detecting wetlands — a critical task for environmental and societal applications such as flood mapping and water resource management — relates to the existence of variable spectral signatures of water resulting from illumination, turbidity, and vegetation. Multispectral optical imagery from satellites such as Sentinel-2 has been widely used to map wetlands using deep learning methods (Jiang et al., 2019; Cui et al., 2020; Dang et al., 2020; Jamali et al., 2021b; Pham et al., 2022; Onojeghuo and Onojeghuo, 2023). For example, Rezaee et al. (2018) perform fine-tuning of AlexNet (Krizhevsky et al., 2017) pre-trained on the ImageNet dataset (Deng et al., 2009) to classify wetland patches using optical imagery. Mahdianpari et al. (2018) explore multiple Convolutional Neural Network (CNN) architectures to determine the one producing the most accurate wetland classification. However, these approaches cannot detect water hidden under vegetation, which is crucial for monitoring water extent in wetlands.

Radar imagery from satellites such as Sentinel-1, which has a SAR C-band sensor that can penetrate vegetation and clouds (Geudtner et al., 2014), have additionally been used for this purpose. Slagter et al. (2020) combine Sentinel-1 with optical Sentinel-2 and fieldwork data to map wetlands using a random forest classifier. The WetNet model (Hosseiny et al., 2021) is an ensemble of three classifiers that uses multitemporal images to map wetlands: a 2D-CNN trained on radar imagery, a 3D-CNN trained on multispectral and multitemporal imagery, and a Recurrent Neural Network (RNN) trained with multivariate temporal information. Jamali et al. (2022) introduce the 3DUnetGSFormer model, which uses a Generative Adversarial Network (GAN) (Goodfellow et al., 2020) to generate synthetic data with similar characteristics as the ground-truth data and a Swin transformer (Liu et al., 2021) to classify the wetland images. Similar approaches have used optical and radar imagery to map wetlands , such as Jamali et al. (2021a) and Jamali and Mahdianpari (2022).

All these approaches have a common limitation: *they require manually annotated data to train their models*. The manually annotated data usually comes from fieldwork, and is very costly and time-consuming to acquire due to logistics, equipment maintenance, sampling, etc. Moreover, these approaches assume that the surface water extent is constant over time, although the water extent usually varies across the season and is dependent on weather conditions. For instance, one wetland location could have been labeled as “open water” because the image was taken in April when the surface water extent of the wetland had increased due to snowmelt. On the other hand, the same location could be dry in July, leading to an inaccurate prediction of water extent.

3. Background and problem formulation

This paper addresses the problem of *detecting water surfaces under vegetation without requiring fieldwork or manually annotated data*. To create and train a model without fieldwork or manually annotated data, we combine remote sensing (Section 3.1) with deep learning techniques (Sections 3.2, 3.3, and 3.4). Here, we recall some of their basic concepts.

3.1. Detecting surface water using remote sensing

Traditionally, optical sensors based on reflected solar radiation have been used to detect water. NDWI is one of the most popular optical methods to delineate open waters (McFeeters, 1996), as it helps differentiate open water from soil since: (1) water reflects green light, (2) water has low reflectance of Near-Infrared (NIR) light, and (3) terrestrial vegetation and soil have a high reflectance of NIR light. For each pixel in an image, the NDWI is calculated using the normalized difference between the green light intensity and the NIR light intensity — the results of the NDWI index range from -1 to $+1$. Water surfaces have positive values, while soil and terrestrial vegetation have zero or negative values because they typically have a higher reflectance of NIR than green light. Fig. 1 shows how NDWI is used to delineate open water.

Other index-based methods to detect water from optical imagery include the Modified NDWI (MNDWI) (Xu, 2006), High Resolution Water Index (HRWI) (Yao et al., 2015), Two-step Urban Water Index (TSUWI) (Wu et al., 2018) and Automated Water Extraction Index (AWEI) (Feyisa et al., 2014).

Although NDWI provides reliable information on open water, its usage is limited to cloud-free days. Moreover, it becomes inaccurate in areas with low albedo and shadows (Feyisa et al., 2014). An alternative approach to optical imagery is SAR, which is unaffected by sunlight and capable of penetrating clouds and vegetation (Mondini et al., 2021). Unlike optical imagery, SAR can reveal hidden water under vegetation cover. However, interpreting SAR imagery alone poses challenges as it is hard to distinguish water from some unvegetated or sparsely vegetated surfaces (Tsyganskaya et al., 2018; Hardy et al., 2019).

3.2. Semantic segmentation of images

Semantic segmentation is a computer vision technique to identify and classify objects within an image at the pixel level. Instead of just identifying objects as a whole, semantic segmentation identifies the exact boundaries of each object and assigns a specific label or class to each pixel within the object’s boundary. This allows for a more precise and detailed analysis of the image, making it useful for applications such as self-driving cars, medical imaging, and, in our case, water delineation.

CNNs are one of the most efficient deep learning approaches for image processing (LeCun et al., 2015). They work by reading an image through a series of layers that extract features such as edges and shapes and then using those features to recognize patterns in the image. They extract a varying level of abstraction from the data in different layers with the added benefit of not requiring prior feature extraction and having more generalization capability. CNNs use convolutional layers to extract features, pooling layers to downsample the output, and activation functions to introduce non-linearity. CNNs have proven effective in image recognition tasks because they can learn to recognize patterns and features in images without being explicitly programmed. CNNs have high prediction accuracy because they (1) can retain the geometrical properties from two-dimensional images, (2) can be trained with large amounts of data and perform consistently across varied data, and (3) do not require expert input of features; the CNNs can learn the features.

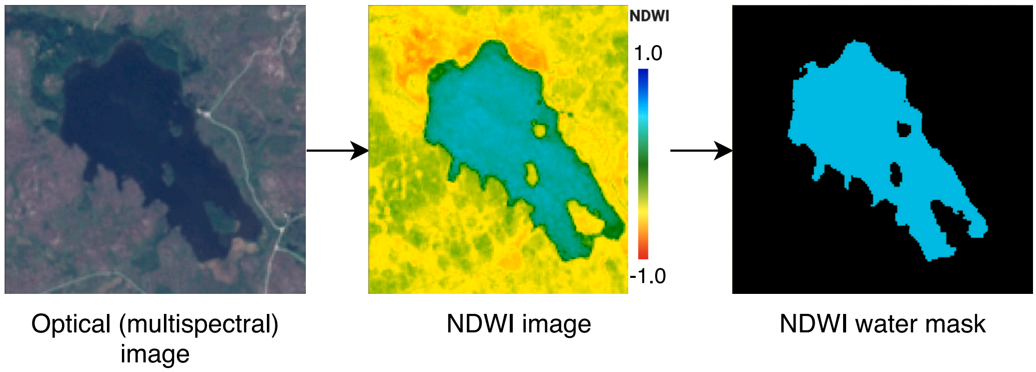

Fig. 1. Water delineation using NDWI. The left image comprises red, green, and blue bands (RGB). The middle image is generated using the NDWI index. The right image shows an NDWI image where pixels with positive values are cyan, and the rest are black, thus completing the water delineation process. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We use a particular CNN architecture called U-Net (Ronneberger et al., 2015) for producing semantic segmentations. U-Net is designed in a way that it has a contracting path followed by an expanding one. The contracting path consists of a series of convolutional layers, reducing the spatial dimensions of the image while increasing the number of channels. This helps to extract high-level features from the input image. The expanding path then consists of a series of up-convolutional layers, which upsample the feature maps to restore the original spatial dimensions of the image, generating a segmentation map with the same dimensions as the input image. Other architectures for semantic segmentation include Resnet (He et al., 2016), MSResNet (Dang and Li, 2021), and MSCENet (Kang et al., 2021).

Here, we use the concept of knowledge distillation to train a U-Net without requiring manually annotated data.

3.3. Knowledge distillation

Knowledge distillation (Hinton et al., 2015; Xu et al., 2020), which is also known as the teacher–student model, is a process of teaching a smaller and simpler model — the student model — to mimic the behavior of a larger and more complex model — the teacher model — and achieve similar performance. During the distillation process, the teacher model generates predictions for training data, which are then used to train the student model. However, instead of training the student model to directly predict the correct output, the student model is trained to learn from the teacher model's predictions.

By doing so, the student model can learn from the teacher model, including the relationships between different input features and the patterns in the data, which is difficult to learn from the training data alone. The result is a smaller and faster model that performs similarly to the larger and more complex teacher model. Additionally, the student model may generalize better than the teacher model under specific conditions (Deng et al., 2022; Beyer et al., 2022), as it has learned to capture the most important aspects of the teacher's behavior while ignoring the noise.

In this paper, we tweak the knowledge distillation process: instead of having a small model learn from a large model, we make a *radar-based* model (a CNN model as a student) learn from an *optical-based* model (a NDWI model as a teacher). We exploit the fact that it is easier to identify water from optical images rather than from SAR images. This process is called cross-modal knowledge distillation (Hu et al., 2020). Finally, we create a model that is able to generate its own training data and learn from it.

3.4. Learning without annotated data

One of the biggest bottlenecks in deep learning models, including CNNs, is requiring large amounts of annotated data to make accurate predictions. Particularly in semantic segmentation of images, manually annotating each image is costly and time-consuming, aggravated by the fact that deep learning models require thousands of images to produce accurate predictions. Techniques like transfer learning (Garcia-Garcia et al., 2018) and data augmentation (Shorten and Khoshgoftaar, 2019) have alleviated the need for large amounts of annotated data. Nevertheless, they still require a minimum amount of data. Self-supervised learning is a machine learning method that allows algorithms to learn without needing human-annotated samples (Shurrab and Duwairi, 2022).

The following section shows how we achieve automatic labeling of training data inspired by the knowledge distillation architecture. We automatically generate ground truth data using NDWI and train a SAR-based CNN to detect water. Our approach has the advantage of requiring "zero" manually annotated data.

4. DeepAqua model

Semantic segmentation projects typically involve a human annotator delineating images to indicate which parts correspond to a particular object or feature, such as water surfaces in SAR images. This process can be time-consuming and expensive, often making data annotation the bottleneck of deep learning projects. Fig. 2(a) shows this traditional architecture with a human annotator delineating the water in radar images and a CNN trained based on these annotated images to recognize water in previously unseen images.

In this study, as seen in Fig. 2(b), instead of relying on human-annotated water masks, our teacher–student architecture uses the NDWI model as the teacher and the CNN as the student, where the teacher extracts knowledge about the location of the water surface from optical imagery and produces segmented images that the student will try to mimic. Unlike most knowledge distillation approaches, our teacher and student models rely on different data types; particularly, the teacher uses optical imagery to produce water segmentations, while the student uses radar imagery. For each batch of images, we calculate the Dice loss (Soomro et al., 2018) between the student predictions and the ground truth masks provided by the teacher and minimize this loss function to improve the performance of the student model. We then backpropagate the loss to update the weights of the CNN. By eliminating the need for manual annotation, we aim to streamline the model training process and reduce overall project costs.

4.1. DeepAqua *framework and workflow*

Fig. 3 shows the DeepAqua's overall framework and workflow. Our method consists of two models: teacher and student models. The teacher model is a thresholding model that generates water masks by applying the NDWI index to optical images, and the student model

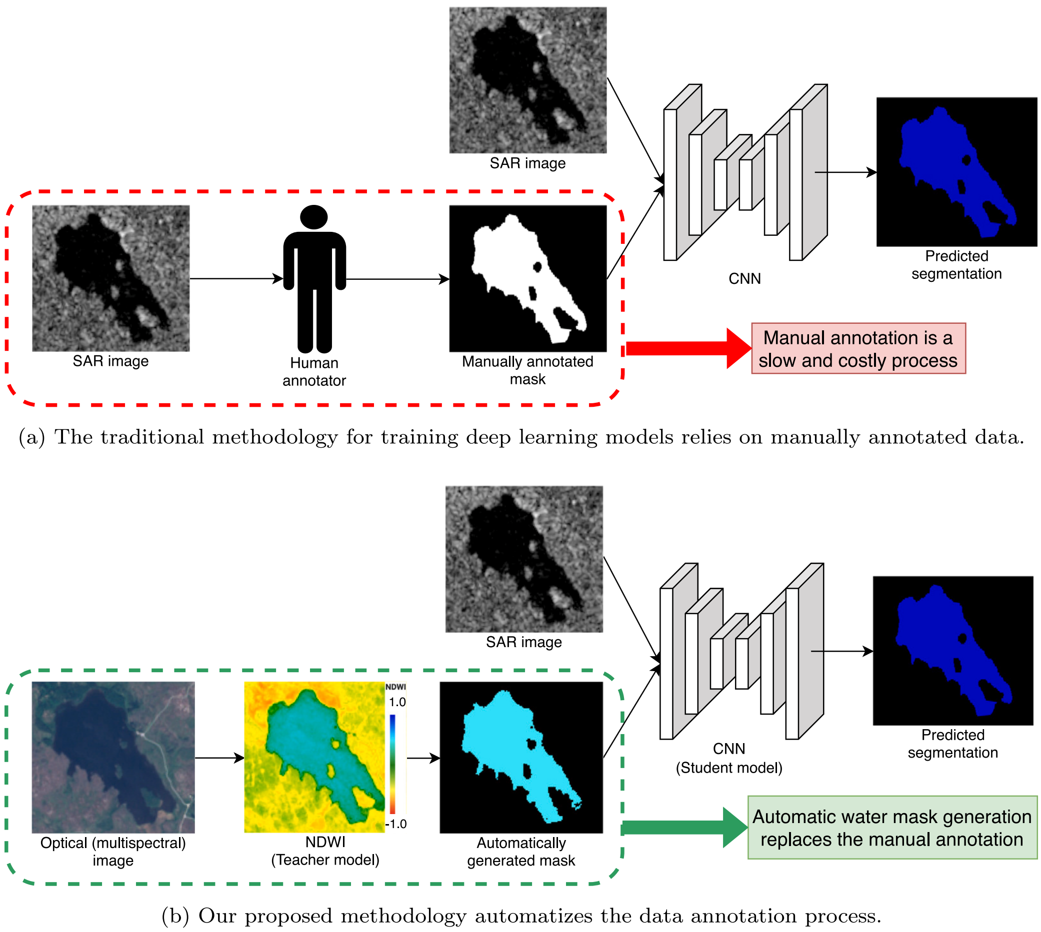

Fig. 2. The training of a CNN model to recognize water boundaries from SAR imagery using NDWI water masks as ground truth.

which is a U-Net (Ronneberger et al., 2015) that takes SAR images as input and produces segmentation masks as output. The teacher and student models are trained jointly by minimizing the Dice loss between their outputs. The workflow of our method is as follows:

- Step 1: We create a training set by selecting images that fulfill the following conditions: First, Sentinel-1 (SAR) and Sentinel-2 (multispectral) image availability on the same date for the region of interest. Second, a maximum of 1% of cloud cover on the Sentinel-2 image, and (3) no missing values on the Sentinel-1 and Sentinel-2 images.
- Step 2: Given a pair of optical and SAR images that are co-registered and cover the same geographic area, we feed the optical image to the teacher model and obtain an NDWI mask as its output.
- Step 3: We feed the SAR image to the student model and obtain a segmentation mask as its output.
- Step 4: We compute the Dice loss between the teacher and student output to measure their similarity.
- Step 5: We update the student weights using backpropagation based on the Dice loss.
- Step 6: We repeat steps 2–5 for all pairs of optical and SAR images in the training set until convergence.

4.2. The teacher model

The teacher model generates water masks from optical images using the NDWI index. Selecting the optimal threshold for NDWI values is challenging, as highlighted by Ji et al. (2009) and Reis et al. (2021). Following the recommendation of McFeeters (1996), we adopted a threshold value of 0.0 to delineate open water. The teacher model outputs a binary water mask matching the input optical image size, where 0 indicates ground and 1 denotes water. This mask then guides the student model.

4.3. The student model

The student model is a U-Net (Ronneberger et al., 2015) model that takes SAR images as input and produces segmentation masks as output. We use U-Net as the student model for two reasons. First, U-Net is a simple and effective model for semantic segmentation that can achieve good results with limited data and computational resources. Second, U-Net is compatible with the teacher model regarding input and output sizes, facilitating the cross-modal learning process.

We train U-Net from scratch on SAR images without requiring annotated data. We use SAR images from Sentinel-1, a satellite mission that provides C-band SAR images with a resolution of 10m ×10m per pixel. For enhanced contrast in these images, we exclude pixel values below the 1st percentile and above the 99th percentile. Subsequently, we normalize the images. The output of the U-Net is a segmentation mask that has the same size as the input SAR image. Within this mask, values span from 0 to 1, with higher values suggesting increased water presence. This segmentation mask serves as the desired output for model optimization.

4.4. The cross-modal learning process

The cross-modal learning process is the core of our method that transfers knowledge from the teacher to the student model. The teacher

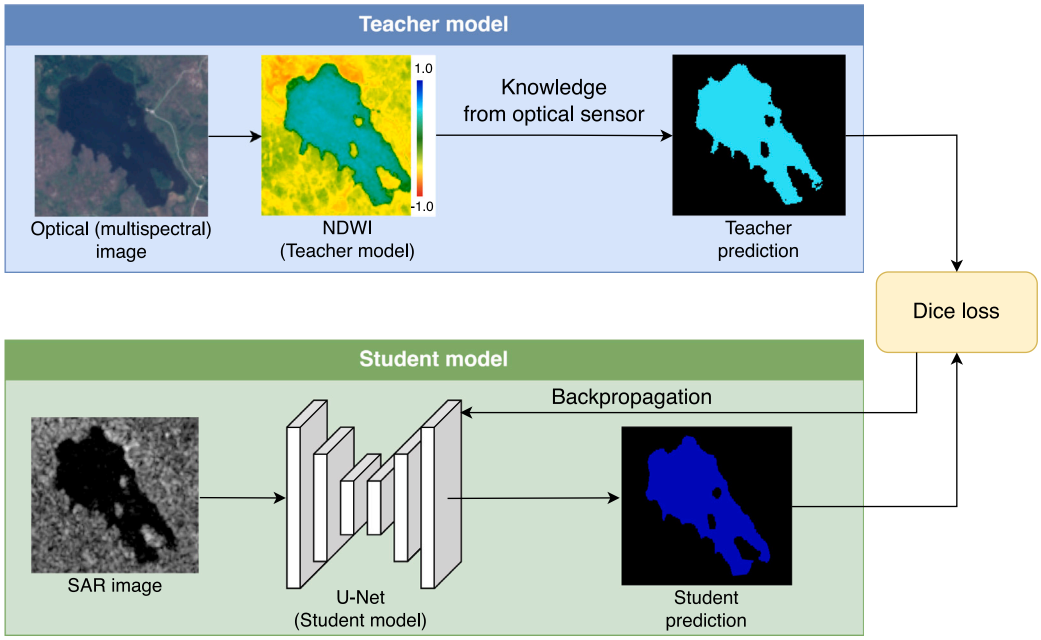

Fig. 3. The process of training a CNN model to recognize water boundaries from SAR imagery by learning from a NDWI model.

model produces a hard NDWI water mask from an optical image, and the student model produces a segmentation mask from a SAR image. The hard NDWI mask and the segmentation mask are aligned in terms of spatial resolution and geographic area, as they are generated from co-registered optical and SAR images that cover the same scene. The cross-modal learning process aims to minimize the Dice loss between the NDWI mask and the segmentation mask, which measures their similarity.

The Dice loss (Dice, 1945) is a loss function frequently employed for semantic segmentation tasks. It is particularly suited for imbalanced data, where one class (e.g., water pixels) might be significantly under-represented compared to another (e.g., land pixels). The Dice loss is defined as:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \times |Y_T * Y_S| + \epsilon}{|Y_T| + |Y_S| + \epsilon} \tag{1}$$

where Y_T and Y_S are the teacher output and the student output, respectively, $|\cdot|$ denotes the sum of all elements in a matrix, $*$ denotes element-wise multiplication, and ϵ is a small constant to avoid division by zero. The Dice loss ranges from 0 to 1, where lower values indicate higher similarity.

By minimizing the Dice loss, the student model learns to mimic the teacher model's output and thus segment SAR images without requiring annotated data. The Dice loss provides a soft and smooth supervision signal for the student model, as it considers true positives in the numerator and true positives, false positives, and false negatives in the denominator. This Dice loss formula helps solve the issue of imbalanced training data and does not require defining weighting parameters between different classes (in our case, the ground and water). Besides, the function works well for binary segmentation tasks (Soomro et al., 2018).

4.5. The backpropagation algorithm

The backpropagation algorithm is the algorithm that updates the student weights based on the Dice loss. The backpropagation algorithm consists of two steps: forward and backward propagation. In the forward propagation, we compute the teacher output, the student output, and the Dice loss for a given pair of optical and SAR images. In the backward propagation, we calculate the gradient of the Dice loss with respect to the student weights and update the weights using an optimizer.

The steps of the backpropagation algorithm are as follows:

- Step 1: Given a pair of optical image X_O, and SAR image X_S, we feed X_O to the teacher model and X_S to the student model. We then obtain Y_T and Y_S as the teacher's and student's outputs, respectively.
- Step 2: Compute $\mathcal{L}_{Dice}$ using Y_T and Y_S as inputs (as in Eq. (1)).
- Step 3: Compute $\partial\mathcal{L}_{Dice}/\partial W_S$ using the chain rule, where W_S are the student weights.
- Step 4: Update W_S using an optimizer (e.g., Adam Kingma and Ba, 2015).
- Step 5: Repeat steps 1–4 for all pairs of optical and SAR images in the training set until convergence.

5. Evaluation

We evaluated the performance of DeepAqua using SAR-Vertical-Horizontal (VH) imagery downloaded from Google Earth Engine (Gorelick et al., 2017).

5.1. Training, validation, and testing datasets

To train DeepAqua, we used Sentinel-1 (SAR) and Sentinel-2 (multispectral) images of the entire county of Örebro in Sweden. Örebro has an area of (∼ 8550 km^2). We used Sentinel-2 multispectral images and then applied NDWI to generate water masks. To this end, we first split the entire Örebro region into tiles of 64 × 64 pixels. Each pixel had a resolution of 10 m. Then, we repeated the same procedure to generate a SAR dataset using Sentinel-1 images from the same region. This resulted in a total of 45 500 multispectral-SAR pairs. Fig. 4 illustrates how we generated the data to train our model. Once we generated all the data, we randomly selected 80% of the tiles to create a training set, and we took the remaining 20% to create a validation set.

Our testing set is composed of imagery from Svartådalen, Hjälstaviken and Hornborgasjön, three wetlands that belong to the Ramsar convention, as shown in Fig. 5. The three wetlands are located in flat areas of Southern and Western Sweden. These are shallow wetlands that are not connected to any important river stream and are rather

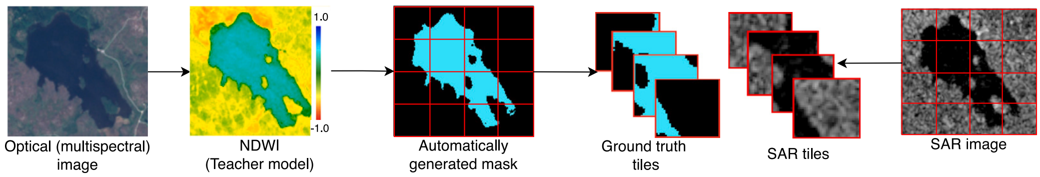

Fig. 4. The process of automatically generating the training set. We create multispectral-SAR (X_O, X_S) pairs by splitting the satellite images into smaller tiles.

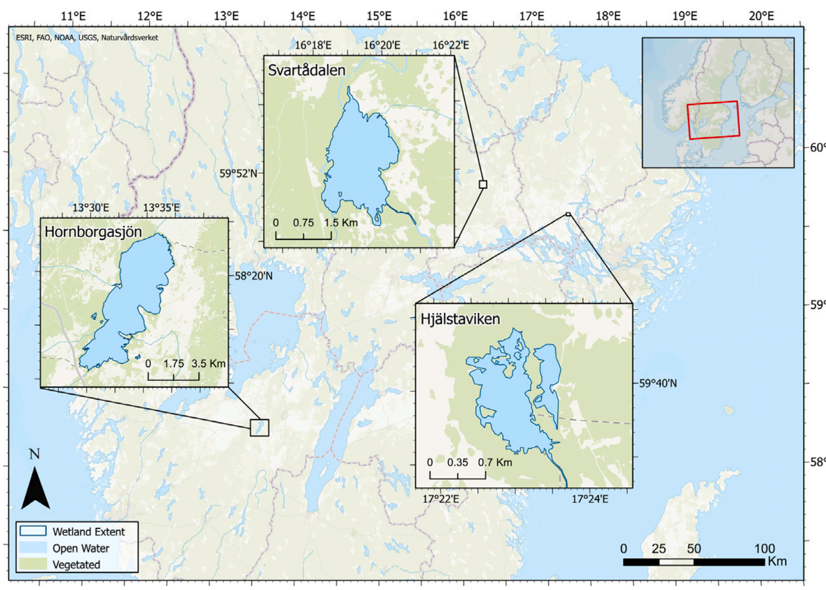

Fig. 5. Map of Sweden containing the location of the three wetlands with manually annotated data that compose the testing set.

fed by small streams with low flow rates. The wetlands are covered by emerging grassy vegetation consisting of mires and open water bodies. The grassy vegetation is often flooded during the rainy season and during spring when the wetlands receive snowmelt from upstream. Some borders of the wetlands also have tree canopies that do not allow the penetration of C-band signals.

Svartådalen is a mixed wetland complex of 1 977 ha comprised of mires, bogs, and fens. Hjälstaviken is a limnic complex of 808 ha. Hornborgasjön is a human-made mire complex of 6 197 ha, one of the largest single nature conservation projects ever carried out in Sweden (Gunnarsson and Löfroth, 2014; Matthews et al., 1993). We manually delineated the water in radar imagery from the study sites. Each site contains 40 images between 2018 and 2022. We excluded the months of January, February, March, and December to avoid images that contained snow and ice.

5.2. Evaluation metrics

To quantitatively evaluate the performance of our semantic segmentation model on radar imagery of wetlands, we employed a set of evaluation metrics as detailed in Everingham et al. (2015). These metrics, derived from TP, TN, FP, and FN counts, measure both pixel-level accuracy and the overall quality of the segmented regions.

1. **Pixel Accuracy (PA)**: This metric computes the proportion of correctly classified pixels in the entire image. It provides an overall sense of how well the model is performing but may not capture errors distributed across different classes effectively: $\frac{TP+TN}{TP+TN+FP+FN}$.
2. **Intersection Over Union (IOU)**: Also known as the Jaccard index, Intersection Over Union (IOU) offers a measure of the overlap between the predicted and true areas, with values ranging from 0 (no overlap) to 1 (perfect overlap): $\frac{TP}{TP+FP+FN}$
3. **Precision**: This quantifies the proportion of positive identifications (i.e., water pixels) that were actually correct. A high precision indicates that the model has fewer false positives: $\frac{TP}{TP+FP}$
4. **Recall**: Also known as sensitivity, recall measures the proportion of actual positives (water pixels in ground truth) that are identified correctly. A high recall means the model has fewer false negatives: $\frac{TP}{TP+FN}$
5. **F1-Score**: The harmonic mean of precision and recall, the F1-score gives a balanced measure of the model's performance, especially when the class distribution is imbalanced: $2 \times \frac{Precision \times Recall}{Precision+Recall}$

By employing these metrics, we aim to comprehensively evaluate our model's capabilities, considering both the fine and broad contexts of water segmentation in radar imagery.

5.3. Baseline methods

We compared the performance of DeepAqua against Otsu's method (Otsu, 1979) and the model by Carvalho Júnior et al. (2011). We selected these methods because they are unsupervised, aligning with our study's central theme of working with non-manually annotated data, making it distinct from the other methods discussed in Section 2.

We decided to concentrate on unsupervised methods because DeepAqua's core advantage is its capacity to train without manual annotations. Introducing supervised methods into our comparison would present complications related to the quality of training data, which might shift attention away from our primary goal: demonstrating the effectiveness of a strong unsupervised solution.

Otsu's method is a technique for automatically determining the optimal threshold value for image segmentation or binarization. This

Table 1
Semantic segmentation performance over various wetland areas in Sweden. The best performance is highlighted in bold.

Model	Svartådalen					Hjälstaviken					Hornborgasjön				
	PA	IOU	Prec	Recall	F1	PA	IOU	Prec	Recall	F1	PA	IOU	Prec	Recall	F1
Dynamic World	0.92	0.59	0.92	0.63	0.75	0.90	0.35	0.57	0.47	0.51	0.88	0.57	0.91	0.60	0.73
Otsu	0.90	0.64	0.69	0.89	0.78	0.81	0.34	0.36	0.85	0.51	0.79	0.49	0.65	0.66	0.66
Otsu + Gaussian filter	0.93	0.73	0.75	**0.96**	0.84	0.83	0.38	0.40	**0.89**	0.55	0.83	0.56	0.73	0.70	0.72
Carvalho Júnior	0.96	0.82	0.90	0.91	0.90	0.89	0.47	0.53	0.82	0.64	0.96	0.85	0.89	0.95	0.92
DeepAqua-NDWI	**0.97**	**0.88**	**0.98**	0.90	**0.93**	**0.96**	0.68	0.81	0.81	**0.81**	**0.98**	**0.94**	**0.98**	0.96	**0.97**
DeepAqua-MNDWI	**0.97**	0.85	0.95	0.89	0.92	0.95	0.68	0.78	0.84	**0.81**	**0.98**	0.93	0.96	**0.97**	0.96
DeepAqua-AWEI	**0.97**	0.84	**0.98**	0.85	0.91	**0.96**	0.68	**0.86**	0.77	**0.81**	**0.98**	**0.94**	**0.98**	0.95	**0.97**
DeepAqua-HRWI	**0.97**	0.86	0.97	0.88	0.92	**0.96**	**0.69**	0.82	0.81	**0.81**	**0.98**	**0.94**	0.97	0.96	**0.97**

method and other thresholding approaches, such as the one described in Carvalho Júnior et al. (2011), seek to find a threshold value that augments the distinction between an image's foreground (water, in our context) and background (soil). Otsu's method computes the variance between these two classes of pixels for every conceivable threshold value. The threshold that mitigates the variance within each class while amplifying the variance between the classes is adopted as the prime choice. We executed our experiments using OpenCV's Python implementation of the Otsu method. We also implemented the model from Carvalho Júnior et al. (2011), which finds the threshold value that maximizes the Dice score (Dice, 1945) from a SAR image and its corresponding NDWI mask.

SAR images are naturally noisy, so filtering techniques are often used to improve the segmentation process (Tan et al., 2023; Zhou et al., 2020; Li et al., 2020). We incorporated a Gaussian filter variation of the Otsu method in response to these recognized challenges. This variation assists in diminishing the noise impact, ensuring more accurate segmentation without drastically diverging from the raw data. Likewise, we used the Gaussian filter in conjunction with the method from Carvalho Júnior et al. (2011). It is noteworthy that while the Gaussian filter aids in reducing noise, our core objective remains to accentuate the potential of our proposed CNN model to train effectively without manual annotations. This commitment to reducing dependency on manually annotated data is in line with our strategy for minimal preprocessing.

We benchmarked DeepAqua using the Dynamic World dataset (Brown et al., 2022), which classifies land cover based on Sentinel-2 optical imagery. We identified *water* and *flooded vegetation* as positive classes, with others as negative. We favored Dynamic World for its five-day update frequency, in contrast to the yearly updates of datasets like Esri (Karra et al., 2021) and the European Space Agency (Zanaga et al., 2022). However, its reliance on optical imagery could limit vegetated water detection compared to radar-based methods.

5.4. Implementation details

We implemented our methods using PyTorch (Paszke et al., 2019) with an Adam (Kingma and Ba, 2015) optimizer with a learning rate of 5×10^{-5}. We minimized the Dice loss function (Dice, 1945) as described in Section 4.4. We trained our method for 20 epochs with a batch size of 32 on a MacBook Pro with an M1 processor and 16 GB of RAM. The total training time was 277 min.

We trained the CNN using the raw data to show the true power of our approach in handling complex and noisy images. We did not apply any filters or preprocessing techniques to clean the original SAR images, except for removing outlier pixel values by discarding values lower than percentile 1 and higher than percentile 99. We also applied min–max scaling to bring the pixel values to the range $[0, 1]$. We could increase prediction performance using techniques such as image denoising, data augmentation, image contrasting, transfer learning, and model ensembling; however, in this paper, we focused only on the potential of the NDWI water masks to train SAR-based CNNs.

We selected and tuned the hyperparameters of our method using a grid search based on the performance of the validation set. For the learning rate, we tried the values $[1 \times 10^{-6}, 5 \times 10^{-6} 1 \times 10^{-5}, 5 \times 10^{-5}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}]$, and for the batch size, we tried the values $[1, 2, 4, 8, 16, 32, 64, 128, 256]$. We stopped at 20 epochs because the model converged at this time. We also experimented with other water indexes such as MNDWI (Xu, 2006), AWEI (Feyisa et al., 2014) and HRWI (Yao et al., 2015). We applied the Otsu and Carvalho Júnior et al. (2011) methods with a Gaussian filter using a 5×5 kernel size.

One of our challenges was that the models trained on 2018 data only showed good performance for 2018 and 2019 but poor performance in 2020, 2021, and 2022. For this reason, we had to train two models based on data from different years. The first model was trained on satellite images taken on July 4th, 2018, and worked well for all 2018–2019 images. The second model was trained on satellite images taken on June 23rd, 2020, working well for all 2020–2022 images. However, after inspecting the images, we realized that the SAR images from 2018–2019 had more speckle and noise than those from 2020–2022, possibly due to an adjustment on the Sentinel-1 sensors. Therefore, we provide a pre-trained version of our model for both 2018 and 2020.

The code, testing dataset, and pre-trained models are available at https://github.com/melqkiades/deep-wetlands.

5.5. Quantitative results

As Table 1 shows, DeepAqua outperforms the Otsu and Carvalho Júnior et al. (2011) models on accuracy, IOU, recall, and F1-score by a significant margin on all three study areas, demonstrating the effectiveness of our approach in leveraging cross-modal learning without requiring annotated data.

Using data from The metrics presented in Table 1, are derived from a weighted aggregation of DeepAqua's performance in the three wetlands, factoring in the varying sizes of each site. The DeepAqua-NDWI model, which is the best overall performer, shows an accuracy of 98%, reflecting its consistent efficacy across different terrains. An IOU of 92% highlights its precision in mapping the overlap between predicted and actual water extents. With a precision of 97%, the model robustly pinpoints true water pixels in its positive predictions, and a recall of 94% indicates its capacity to identify most water pixels within the dataset. This balance between precision and recall results in an F1-score of 96%. While the performance of the various DeepAqua models is similar, we can see that DeepAqua-NDWI has a slight edge. This could be due to the NDWI index it uses, which has a 10-m resolution, compared to the 20-m resolution of MNDWI and AWEI. DeepAqua surpasses the baseline models in almost every single metric across the three study areas.

We can also see that the model from Carvalho Júnior et al. (2011) outperforms the Otsu models with a Gaussian filter on pixel accuracy, IOU, precision, and F1-score. However, the Otsu method exhibits a high recall, indicating its proficiency in recognizing water bodies, albeit with some propensity to misclassify non-water areas. While the Dynamic World model struggles to detect vegetated water, its precision surpasses the Otsu and Carvalho Júnior et al. (2011) models across all three study areas, indicating that its water pixel predictions are often accurate.

DeepAqua demonstrates effective water surface detection and consequent water surface extent estimation, with errors being a small portion

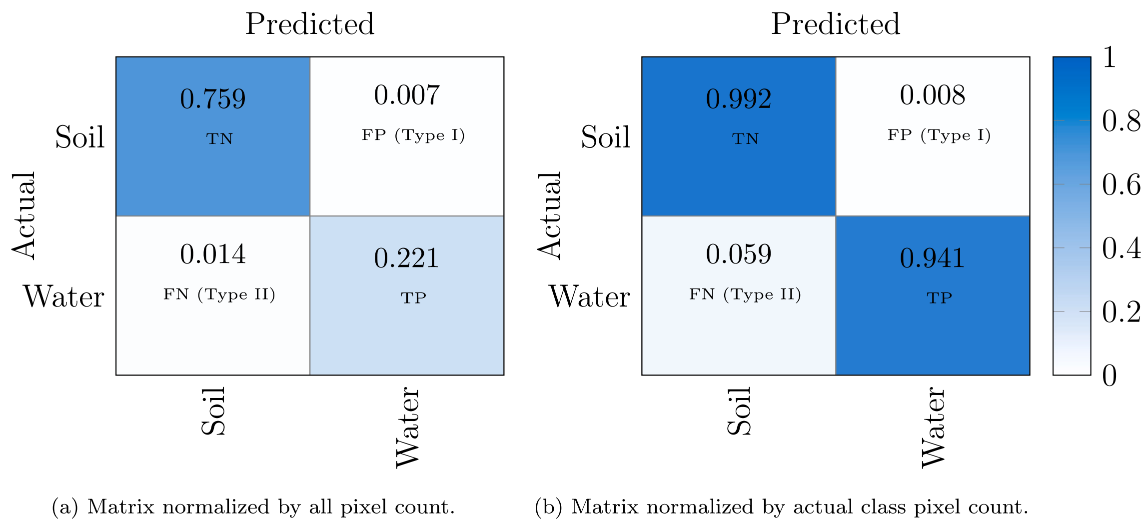

Fig. 6. Comparative confusion matrices for DeepAqua's performance. Fig. 6(a) displays normalization against the total pixel count, while Fig. 6(b) emphasizes performance metrics specific to each class.

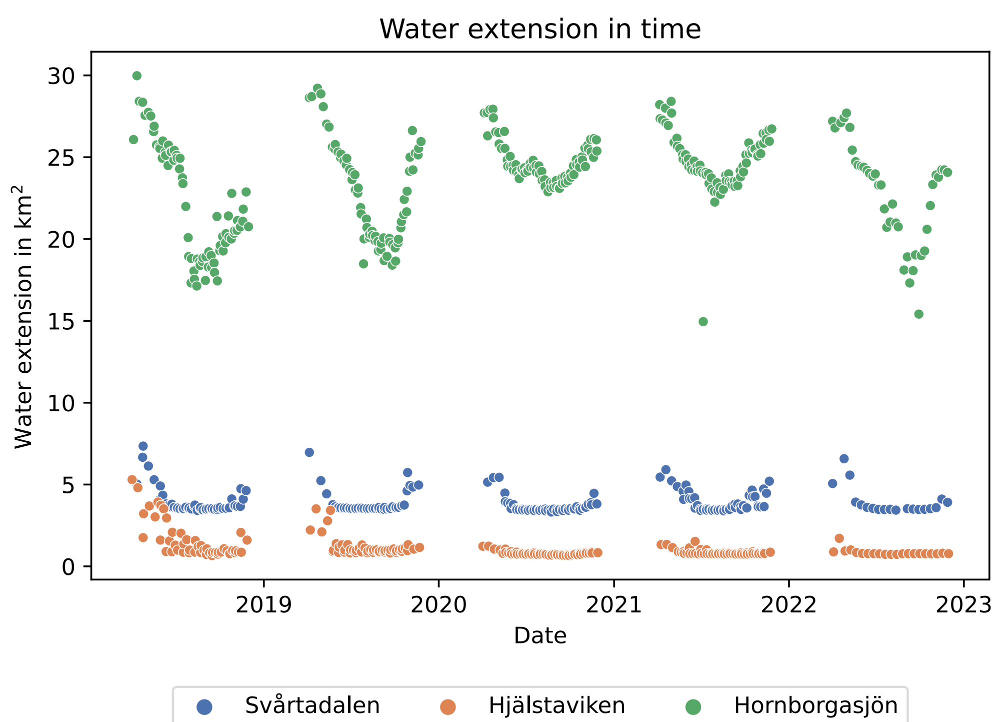

Fig. 7. Dots indicate the water extension of three Swedish wetlands over 2018–2022. Only the months of April through November are considered.

of its results, as evident from the confusion matrices in Fig. 6. The 0.7% FP rate indicates an overestimation of water surfaces. On the other hand, the 1.4% FN rate suggests the incapability of detecting all water surfaces. Yet, these figures highlight the model's ability to distinguish soil from water, especially when considering the challenges of detecting both open and vegetated water surfaces in dynamic wetland environments.

5.6. Qualitative results

We applied DeepAqua to assess the surface water extent in three study areas from 2018 to 2022. Fig. 7 shows total water extent in the wetlands as measured by our model. Typically, the wetlands experience increased water levels during spring due to snowmelt, effect that fades as summer arrives. As autumn approaches, the wetland surface water extent increases once again.

To underscore the accuracy of DeepAqua, Fig. 8 offers predictions for the three study areas across different months. The leftmost column displays the input SAR image, while the rightmost presents the manually annotated ground truth. Intermediate columns feature predictions from all models.

The top row captures the Svårtadalen wetland during summer on July 4th, 2018. It is evident that the Otsu model and the method from Carvalho Júnior et al. (2011) have difficulty with speckle noise; however, filtering mitigates part of it. The DeepAqua model appears

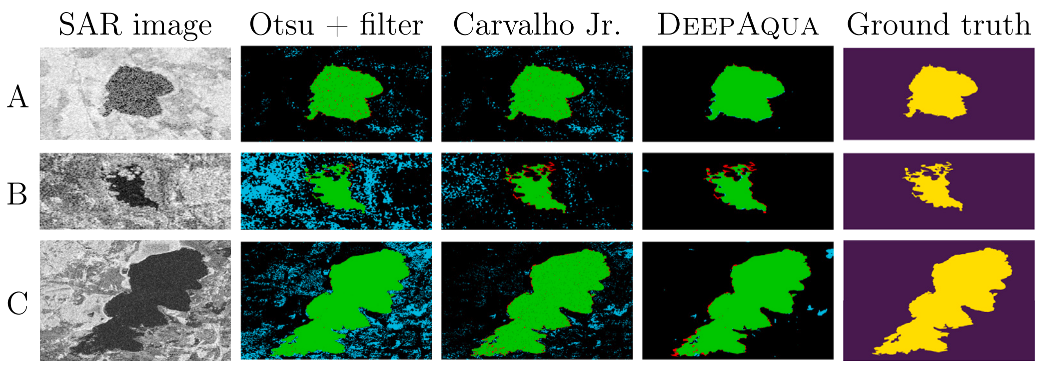

Fig. 8. Illustration of the performance of our model for different areas and times of the year: Svårtadalen, July 4th, 2018 (A); Hjälstaviken, October 4th, 2020; and Hornborgasjön April 19th, 2021 (C). From left to right: original SAR images, segmentation using Otsu's method with a Gaussian filter, segmentation using the model from Carvalho Júnior et al. (2011), segmentation using DeepAqua, and manually annotated data. Green pixels denote TP, cyan pixels denote FP, red pixels denote FN and black pixels denote TN. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

resilient to this noise and aligns closely with the ground truth. The middle row captures the Hjälstaviken wetland during autumn on October 4th, 2020. This image is more noisy than the previous one, causing challenges particularilly for the Otsu method. Instead of reducing the noise, the Gaussian filter amplifies it. The model from Carvalho Júnior et al. (2011) is less affected by the noise than the Otsu. Overall, DeepAqua offers clear predictions of the water surface and no noise. Yet, as indicated by the red pixels, it could not detect water surfaces in some areas at the top of the image. The bottom row shows the Hornborgasjön wetland in spring on April 19th, 2021. Both Otsu and the method from Carvalho Júnior et al. (2011) show a low prediction potential due to noise (cyan areas), which is not present in the prediction of DeepAqua.

Notably, DeepAqua accurately predicts surface water extent and is capable of identifying the "land islands" within the wetlands. While it is not flawless, with some FP and FN errors evident, its noise reduction capability is commendable and achieved without resorting to additional filtering or preprocessing. As the red and cyan pixels in Fig. 8 show, errors often appear around wetland shores or where water levels are low. In these regions, differentiating between water and soil in SAR images can be challenging due to the mixed signals from the water–soil interface. We emphasize that DeepAqua does not use filtering or pre-processing techniques on the SAR images.

To summarize, DeepAqua excels in recognizing water surfaces from SAR images. While the method effectively tracks water extent changes in time, it does not inherently distinguish between open and vegetated waters due to the inherent limitations of doing such from SAR imagery . Although not the primary focus, coupling our approach with water index methods like NDWI can easilly help differentiate these two tyoes of surfaces. It can also be updated with more extended wavelengths than the C-band to even detect waters below thicker canopies such as mangroves.

6. Conclusion

We present DeepAqua, a novel method that uses cross-modal learning to train a CNN for semantic segmentation of water in SAR imagery without requiring annotated data. Our method consists of two models: a teacher model that creates NDWI water masks from optical images and a student model that learns to segment water in SAR images. We used U-Net to implement the student model. The teacher and student models are trained jointly by minimizing the Dice loss between their outputs. Our experiments confirmed that our model can accurately segment images and confidently detect water. The model here is trained and tested in three wetland environments and overall can be applied to wetlands with emerging vegetation. Future studies may adapt the approach for radar sensors with longer wavelengths to expand the applicability in wetlands with thicker vegetation, such as Mangroves. This model can help applications where water detection is crucial, such as flooding detection, river and lake mapping, and water availability assessments in time and space.

CRediT authorship contribution statement

Francisco J. Peña: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Clara Hübinger:** Writing – review & editing. **Amir H. Payberah:** Methodology, Supervision, Writing – original draft, Writing – review & editing. **Fernando Jaramillo:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

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.

Data availability

All the source code, testing dataset, and pre-trained models can be found in our online open-source repository at https://github.com/melqkiades/deep-wetlands. Please cite this article when using the DeepAqua model.

Acknowledgments

This work was supported by Digital Futures. We gratefully acknowledge the Bolin Centre for Climate Research for providing us with a GPU server and the National Academic Infrastructure for Super computing in Sweden for giving us access to their high computing cluster. We thank Ezio Cristofoli, Johanna Hansen, and Ioannis Iakovidis for their ideas and for annotating data.

References

Beyer, L., Zhai, X., Royer, A., Markeeva, L., Anil, R., Kolesnikov, A., 2022. Knowledge distillation: A good teacher is patient and consistent. In: Proc. IEEECVF Conf. Comput. Vis. Pattern Recognit. pp. 10925–10934.

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., et al., 2022. Dynamic world, near real-time global 10 m land use land cover mapping. Sci. Data 9 (251).

Carvalho Júnior, O.A., Guimarães, R.F., Gillespie, A.R., Silva, N.C., Gomes, R.A., 2011. A new approach to change vector analysis using distance and similarity measures. Remote Sens. 3, 2473–2493.

Cui, B., Zhang, Y., Li, X., Wu, J., Lu, Y., 2020. WetlandNet: Semantic segmentation for remote sensing images of coastal wetlands via improved UNet with deconvolution. In: Genet. Evol. Comput. Proc. Thirteen. Int. Conf. Genet. Evol. Comput. Novemb. 1–3 2019 Qingdao China 13, Springer, pp. 281–292.

Dang, B., Li, Y., 2021. Msresnet: Multiscale residual network via self-supervised learning for water-body detection in remote sensing imagery. Remote Sens. 13 (3122).

Dang, K.B., Nguyen, M.H., Nguyen, D.A., Phan, T.T.H., Giang, T.L., Pham, H.H., Nguyen, T.N., Tran, T.T.V., Bui, D.T., 2020. Coastal wetland classification with deep U-Net convolutional networks and Sentinel-2 imagery: A case study at the Tien Yen estuary of Vietnam. Remote Sens. 12 (3270).

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conf. Comput. Vis. Pattern Recognit. Ieee, pp. 248–255.

Deng, X., Zheng, J., Zhang, Z., 2022. Personalized education: Blind knowledge distillation. In: Comput. Vision–ECCV 2022 17th Eur. Conf. Tel Aviv Isr. Oct. 23–27 2022 Proc. Part XXXIV, Springer, pp. 269–285.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

Everingham, M., Eslami, S.M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2015. The pascal visual object classes challenge: A retrospective. Int. J. Comput. Vis. 111, 98–136.

Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated water extraction index: A new technique for surface water mapping using landsat imagery. Remote Sens. Environ. 140, 23–35.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., Garcia-Rodriguez, J., 2018. A survey on deep learning techniques for image and video semantic segmentation. Appl. Soft Comput. 70, 41–65.

Geudtner, D., Torres, R., Snoeij, P., Davidson, M., Rommen, B., 2014. Sentinel-1 system capabilities and applications. In: 2014 IEEE Geosci. Remote Sens. Symp. IEEE, pp. 1457–1460.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Commun. ACM 63, 139–144.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. http://dx.doi.org/10.1016/j.rse.2017.06.031.

Gunnarsson, U., Löfroth, M., 2014. The Swedish wetland survey: compiled excerpts from the national final report. Naturvårdsverket.

Hardy, A., Ettritch, G., Cross, D.E., Bunting, P., Liywalii, F., Sakala, J., Silumesii, A., Singini, D., Smith, M., Willis, T., et al., 2019. Automatic detection of open and vegetated water bodies using Sentinel 1 to map African malaria vector mosquito breeding habitats. Remote Sens. 11 (593).

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. pp. 770–778.

Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. ArXiv Prepr. ArXiv150302531 arXiv:1503.02531.

Hosseiny, B., Mahdianpari, M., Brisco, B., Mohammadimanesh, F., Salehi, B., 2021. WetNet: A spatial–temporal ensemble deep learning model for wetland classification using sentinel-1 and sentinel-2. IEEE Trans. Geosci. Remote Sens. 60, 1–14.

Hu, H., Xie, L., Hong, R., Tian, Q., 2020. Creating something from nothing: Unsupervised knowledge distillation for cross-modal hashing. In: Proc. IEEECVF Conf. Comput. Vis. Pattern Recognit. pp. 3123–3132.

Jamali, A., Mahdianpari, M., 2022. Swin transformer and deep convolutional neural networks for coastal wetland classification using sentinel-1, sentinel-2, and LiDAR data. Remote Sens. 14 (359).

Jamali, A., Mahdianpari, M., Brisco, B., Granger, J., Mohammadimanesh, F., Salehi, B., 2021a. Deep forest classifier for wetland mapping using the combination of sentinel-1 and sentinel-2 data. GIScience Remote Sens. 58, 1072–1089.

Jamali, A., Mahdianpari, M., Brisco, B., Granger, J., Mohammadimanesh, F., Salehi, B., 2021b. Wetland mapping using multi-spectral satellite imagery and deep convolutional neural networks: A case study in newfoundland and labrador, Canada. Can. J. Remote Sens. 47, 243–260.

Jamali, A., Mahdianpari, M., Brisco, B., Mao, D., Salehi, B., Mohammadimanesh, F., 2022. 3DuNetGSFormer: A deep learning pipeline for complex wetland mapping using generative adversarial networks and swin transformer. Ecol. Inf. 72, 101904.

Jaramillo, F., Desormeaux, A., Hedlund, J., Jawitz, J.W., Clerici, N., Piemontese, L., Rodríguez-Rodriguez, J.A., Anaya, J.A., Blanco-Libreros, J.F., Borja, S., Celi, J., Chalov, S., Chun, K.P., Cresso, M., Destouni, G., Dessu, S.B., Baldassarre, G.Di., Downing, A., Espinosa, L., Ghajarnia, N., Girard, P., Gutiérrez, A.G., Hansen, A., Hu, T., Jarsjö, J., Kalantari, Z., Labbaci, A., Licero-Villanueva, L., Livsey, J., Machotka, E., McCurley, K., Palomino-Ángel, S., Pietron, J., Price, R., Ramchunder, S.J., Ricaurte-Villota, C., Ricaurte, L.F., Dahir, L., Rodríguez, E., Salgado, J., Sannel, A.B.K., Santos, A.C., Seifollahi-Aghmiuni, S., Sjöberg, Y., Sun, L., Thorslund, J., Vigouroux, G., Wang-Erlandsson, L., Xu, D., Zamora, D., Ziegler, A.D., Åhlén, I., 2019. Priorities and interactions of sustainable development goals (sdgs) with focus on wetlands. Water 11, http://dx.doi.org/10.3390/w11030619, https://www.mdpi.com/2073-4441/11/3/619.

Ji, L., Zhang, L., Wylie, B., 2009. Analysis of dynamic thresholds for the normalized difference water index. Photogramm. Eng. Remote Sens. 75, 1307–1317.

Jiang, Z., Von Ness, K., Loisel, J., Wang, Z., 2019. Arcticnet: A deep learning solution to classify arctic wetlands. ArXiv Prepr. ArXiv190600133 arXiv:1906.00133.

Kang, J., Guan, H., Peng, D., Chen, Z., 2021. Multi-scale context extractor network for water-body extraction from high-resolution optical remotely sensed images. Int. J. Appl. Earth Obs. Geoinf. 103, 102499.

Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J.C., Mathis, M., Brumby, S.P., 2021. Global land use/land cover with sentinel 2 and deep learning. In: 2021 IEEE international geoscience and remote sensing symposium IGARSS. IEEE, pp. 4704–4707.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (Eds.), 3rd Int. Conf. Learn. Represent. ICLR 2015 San Diego CA USA May 7-9 2015 Conf. Track Proc.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Commun. ACM 60, 84–90.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

Li, N., Lv, X., Xu, S., Li, B., Gu, Y., 2020. An improved water surface images segmentation algorithm based on the Otsu method. J. Circuits Syst. Comput. 29, 2050251.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proc. IEEECVF Int. Conf. Comput. Vis.. pp. 10012–10022.

Mahdianpari, M., Salehi, B., Rezaee, M., Mohammadimanesh, F., Zhang, Y., 2018. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. Remote Sens. 10 (1119).

Matthews, G.V.T., et al., 1993. The Ramsar Convention on Wetlands: Its History and Development. Ramsar Convention Bureau Gland.

McFeeters, S.K., 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432.

Mondini, A.C., Guzzetti, F., Chang, K.T., Monserrat, O., Martha, T.R., Manconi, A., 2021. Landslide failures detection and mapping using synthetic aperture radar: Past, present and future. Earth Sci. Rev. 216, 103574.

Onojeghuo, A.O., Onojeghuo, A.R., 2023. Wetlands mapping with deep ResU-Net CNN and open-access multisensor and multitemporal satellite data in Alberta's parkland and grassland region. Remote Sens. Earth Syst. Sci. 1–16.

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32.

Pham, H.N., Dang, K.B., Nguyen, T.V., Tran, N.C., Ngo, X.Q., Nguyen, D.A., Phan, T.T.H., Nguyen, T.T., Guo, W., Ngo, H.H., 2022. A new deep learning approach based on bilateral semantic segmentation models for sustainable estuarine wetland ecosystem management. Sci. Total Environ. 838, 155826.

Reis, L.G.D.M., Souza, W.D.O., Ribeiro Neto, A., Fragoso, Jr., C.R., Ruiz-Armenteros, A.M., Cabral, J.J.D.S.P., Montenegro, S.M.G.L., 2021. Uncertainties involved in the use of thresholds for the detection of water bodies in multitemporal analysis from landsat-8 and sentinel-2 images. Sensors 21, 7494.

Rezaee, M., Mahdianpari, M., Zhang, Y., Salehi, B., 2018. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 11, 3030–3039.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Int. Conf. Med. Image Comput. Comput.-Assist. Interv. Springer, pp. 234–241.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J. Big Data 6, 1–48.

Shurrab, S., Duwairi, R., 2022. Self-supervised learning methods and applications in medical imaging analysis: A survey. PeerJ Comput. Sci. 8, e1045.

Slagter, B., Tsendbazar, N.E., Vollrath, A., Reiche, J., 2020. Mapping wetland characteristics using temporally dense sentinel-1 and sentinel-2 data: A case study in the St,Lucia wetlands, South Africa. Int. J. Appl. Earth Obs. Geoinf. 86, 102009.

Soomro, T.A., Afifi, A.J., Gao, J., Hellwich, O., Paul, M., Zheng, L., 2018. Strided u-net model: Retinal vessels segmentation using dice loss. In: 2018 Digit. Image Comput. Tech. Appl. DICTA. IEEE, pp. 1–8.

Tan, J., Tang, Y., Liu, B., Zhao, G., Mu, Y., Sun, M., Wang, B., 2023. A self-adaptive thresholding approach for automatic water extraction using sentinel-1 sar imagery based on otsu algorithm and distance block. Remote Sens. 15 (2690).

Thorslund, J., Jarsjo, J., Jaramillo, F., Jawitz, J.W., Manzoni, S., Basu, N.B., Chalov, S.R., Cohen, M.J., Creed, I.F., Goldenberg, R., et al., 2017. Wetlands as large-scale nature-based solutions: Status and challenges for research, engineering and management. Ecol. Eng. 108, 489–497.

Tsyganskaya, V., Martinis, S., Marzahn, P., Ludwig, R., 2018. Detection of temporary flooded vegetation using sentinel-1 time series data. Remote Sens. 10 (1286).

Wu, W., Li, Q., Zhang, Y., Du, X., Wang, H., 2018. Two-step urban water index (tsuwi): A new technique for high-resolution mapping of urban surface water. Remote sens. 10 (1704).

Xu, H., 2006. Modification of normalised difference water index (ndwi) to enhance open water features in remotely sensed imagery. Int. J. Remote Sens. 27, 3025–3033.

Xu, G., Liu, Z., Li, X., Loy, C.C., 2020. Knowledge distillation meets self-supervision. In: Comput. Vision–ECCV 2020 16th Eur. Conf. Glasg. UK August 23–28 2020 Proc. Part IX, Springer, pp. 588–604.

Yao, F., Wang, C., Dong, D., Luo, J., Shen, Z., Yang, K., 2015. High-resolution mapping of urban surface water using zy-3 multi-spectral imagery. Remote Sens. 7, 12336–12355.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.E., Xu, P., Ramoino, F., Arino, O., 2022. Esa worldcover 10 m 2021 v200. http://dx.doi.org/10.5281/zenodo.7254221.

Zhou, S., Kan, P., Silbernagel, J., Jin, J., 2020. Application of image segmentation in surface water extraction of freshwater lakes using radar data. ISPRS Int. J. Geo-Inf. 9 (424).