



A Deep Multitask Semisupervised Learning Approach for Chlorophyll-a Retrieval from Remote Sensing Images

Melike Ilteralp ^{1,*} , Sema Ariman ² and Erchan Aptoula ³

¹ Department of Computer Engineering, Gebze Technical University, Kocaeli 41400, Turkey

² Department of Meteorological Engineering, Samsun University, Samsun 55070, Turkey; sema.ariman@samsun.edu.tr

³ Institute of Information Technologies, Gebze Technical University, Kocaeli 41400, Turkey; eaptoula@gtu.edu.tr

* Correspondence: melikeilteralp@gtu.edu.tr

Abstract: This article addresses the scarcity of labeled data in multitemporal remote sensing image analysis, and especially in the context of Chlorophyll-a (Chl-a) estimation for inland water quality assessment. We propose a multitask CNN architecture that can exploit unlabeled satellite imagery and that can be generalized to other multitemporal remote sensing image analysis contexts where the target parameter exhibits seasonal fluctuations. Specifically, Chl-a estimation is set as the main task, and an unlabeled sample's month classification is set as an auxiliary network task. The proposed approach is validated with multitemporal/spectral Sentinel-2 images of Lake Balik in Turkey using in situ measurements acquired during 2017–2019. We show that harnessing unlabeled data through multitask learning improves water quality estimation performance.

Keywords: time series analysis; water quality; convolutional neural network; regression; semisupervised learning



Citation: Ilteralp, M.; Ariman, S.; Aptoula, E. A Deep Multi-task Semi-supervised Learning Approach for Chlorophyll-A Retrieval from Remote Sensing Images. *Remote Sens.* **2022**, *14*, 18. <https://doi.org/10.3390/rs14010018>

Academic Editor: Charlotte Pelletier

Received: 3 November 2021

Accepted: 17 December 2021

Published: 22 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Besides containing invaluable aquatic habitats and rich biodiversity, lake water also constitutes a significant resource with a wide range of civil and industrial purposes, ranging from urban water supply and agricultural irrigation to fishery and recreation. Consequently, monitoring its quality is of paramount importance, especially given the growing threat from uncontrolled farming practices (e.g., excessive fertilizer and pesticide use) and industrial pollution [1].

Chlorophyll-a (Chl-a) concentration is one of the most commonly used water quality parameters. Its high levels indicate a state of eutrophication, most often due to an abundance of nutrients [2]. It is known to not only cause the death of aquatic life but also constitutes a threat to public health [3]. The traditional measurement of Chl-a concentration is a prolonged and labor-intensive process. It involves arduous and repeated sample acquisitions from the field and their subsequent laboratory analysis. The end results, in addition, only reflect the water quality of a handful of measurement sites at most, at often large temporal intervals, or even fewer, in cases where the lake is geographically inaccessible [4]. Therefore, developing a computer model that employs as input remote sensing images acquired over the area of interest, and that outputs pixel-level Chl-a estimations, constitutes a highly attractive and efficient alternative solution, capable of the systematic monitoring of an entire water body, rendering both field visits and laboratory analysis redundant.

The research question underlying this study is how to develop such a computer model, capable of exploiting the scarce labels as well as abundant unlabeled data.

1.1. Related Work

The assessment of water quality from remote sensing images in terms of Chl-a concentration constitutes a long-standing challenge, with an abundance of published approaches spanning open oceans [5], and coastal [6] and inland waters [7]. Chl-a retrieval algorithms

can be aggregated into two broad categories: semianalytical and semiempirical [8]; for a comprehensive survey, the reader is referred to [9].

Semianalytical approaches rely on radiative transfer theory and modeling the propagation of light in water, through the radiative transfer equation that connects inherent optical properties of lake water with its radiance levels [10]. Solving this equation via either look-up tables or inversion methods can lead to the estimation of Chl-a concentration. Even though they can be developed with no available field samples, and are broadly applicable thanks to their physical foundation [11], they still require ample prior information about the lake's optically active water constituents. However, their main disadvantage is their high sensitivity to atmospheric effects, as the models rely heavily on precise radiance readings, thus limiting their operational capacity [12].

Semi-empirical approaches, on the other hand, are more robust against atmospheric effects and do not require prior information about a lake's physical characteristics [9]. They rely on feature engineering; in other words, each pixel of an often-optical remote sensing image is described through a numerical feature vector that is subsequently provided to a statistical/machine learning algorithm with the end goal of developing a regression model for Chl-a estimation. Notable examples include Gaussian processes [13], multilayer perceptron [14], and support vector regression [15]. Comprehensive comparative studies of such features have been provided in [16]. As features, various linear and nonlinear band combinations can be encountered in the state-of-the-art [17], involving mostly blue–green and red–near-infrared bands [18], as well as various expertly derived band ratios and indices [19]. Naturally, they require extensive in situ measurements of Chl-a in order to produce models with high estimation accuracy.

The advent of satellites such as Sentinel-2, with shorter revisit times and higher spatial and radiometric resolutions, has not only paved the way for new advances, but has also rendered necessary the use of more elaborate methods, to deal with the greater level of image detail [20]. More precisely, the results of an extensive comparative study across various band ratios, using Sentinel-2 multispectral images with multiple atmospheric correction processors, has been presented in [16]. Jadidi et al. [21] explored spectra-derived features through color space and coordinate transforms, from images belonging to three sources while working on data collected from Central European lakes. The combined use of spectral–spatial features obtained through connected morphological operators has also been investigated [22]. One of the most extensive studies is due to Neil et al. [23], who explored 185 inland and coastal aquatic systems at a global level, through which the performances of 48 distinct Chl-a estimation algorithms were tested.

However, the Chl-a estimation performance of semiempirical approaches depends heavily on the effectiveness of the underlying machine learning method. Consequently, the paradigm-shift that has occurred in the field of machine learning through the advent of deep learning (DL) [24] is of paramount importance in this context. More specifically, DL has led to groundbreaking performances in various highly challenging computer vision tasks, especially via convolutional neural networks (CNNs) [25]. This new family of algorithms enabling the design and efficient training of deep neural networks has rendered, de facto, the process of feature engineering redundant, as it is now delegated to the network itself. Depending on the quality of the provided data, the deep networks often compute features outperforming handcrafted alternatives. For a survey of DL applications in remote sensing, the reader is referred to [26].

It is thus not surprising that DL methods have been already applied to Chl-a estimation. In particular, Peterson et al. [27] have developed an artificial neural network (ANN) with six hidden layers, operating on spectral signatures of Landsat-8 and Sentinel-2 images at 30 m resolution, and applied it to the water quality estimation of lakes in the United States. It constitutes one of the first studies that assesses the suitability of DL for this context, and their results show increased accuracy and robustness with respect to alternative methods. In addition, Pu et al. [28] reported the first application of CNNs to this problem, where small patches of Landsat-8 images, centered on the pixel under study, were used as network input, with the end goal of classifying the water quality of Chinese lakes. Since,

however, Chl-a estimation is inherently a regression problem, such networks estimating the direct Chl-a level of its input have appeared as well [29,30]. Moreover, in an effort to exploit temporal correlations across multitemporal data, long short-term memory networks have also been studied, both individually [31] and together with CNNs [32]. Lastly, a highly comprehensive study was presented in [8], on lakes across four continents, using a relatively large dataset of 2943 samples. They proposed a single cross-mission solution for both Sentinel-2 and Sentinel-3 sensors, based upon a five-layer mixture density network, and obtained remarkable results.

1.2. Aim and Contribution

As with many remote sensing applications (e.g., semantic segmentation), overfitting due to labeled data shortage is a critical problem in this context as well. In fact, even more so, as, for each Chl-a sample of the ground truth, a rigorous in situ sample collection and laboratory analysis workflow is required (as opposed to an expert labeling pixels on a screen), and as such, labeled datasets are often relatively small. This problem is further exacerbated by the notorious labeled data need of contemporary deep networks. Consequently, it is imperative to exploit, to the maximum extent possible, whatever precious little amount of data is available.

To this end, this article's main contribution is a new network structure that learns to estimate Chl-a concentration levels while exploiting both labeled image locations, as well as unlabeled portions of the same visual content. We propose to achieve this via a multitask double-branch CNN, whose main task is Chl-a level regression via training with labeled input, and whose auxiliary task is the classification of unlabeled samples to its month of acquisition. The proposed approach was tested with multispectral Sentinel-2 images of Lake Balik in North Turkey, using samples collected over 3 years. The ablation studies conducted show that the inclusion of unlabeled data improves the correlation coefficient of the resulting model by multiple percentile points.

In the sequel of this article, following an overview of the collected dataset and study area (Section 2.1), we elaborate on the proposed method of Chl-a concentration estimation (Section 2.2), present the results of our regression experiments (Section 3), and discuss our findings, while Section 4 is devoted to concluding remarks.

2. Materials and Methods

2.1. Study Area and Data

The study area for our water quality experiments is Lake Balik, located in the northern part of Turkey, at the Kizilirmak Delta (Figure 1a). It has an area of 13.9 km² and an average depth of 1.5 m. Maximum depth reaches 2.5 m in summer and 3.5 m in winter. Lake Balik has been a designated Ramsar site since 1998 and is home to 74% of the bird species encountered in Turkey. It is surrounded by agricultural areas, and unfortunately is under significant threat from uncontrolled fertilizer and pesticide use. Surface vegetation is present only at the shallows and extends to, at most, 3 m from the shore [33].

In situ measurements were conducted on 32 distinct dates from April 2017 until June 2019 (7 during 2017, 14 during 2018, and 11 during 2019), once, or at most twice, a month, depending on weather conditions. At each field visit, samples were collected from 10 measurement sites, approximately 500 m apart and 250–300 m from the shore (shown in Figure 1a), with no surface vegetation and minimal water movement. Sample collection was arranged to coincide with the satellites' visits over the lake, and lasted at most 3 h. Thus, the resulting dataset contains $10 \times 32 = 320$ samples.

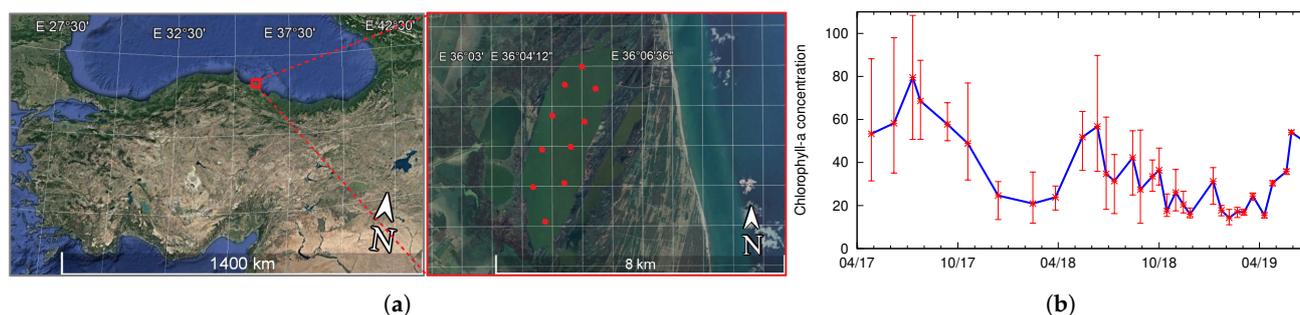


Figure 1. (a) The study area in North Turkey at the Kizilirmak Delta ($36^{\circ}04'23''$ N, $41^{\circ}34'11''$ E); red dots denote measurement positions. (b) The Chl-a concentration distribution of our dataset across time. Each point in the plot denotes the average of 10 measurement sites, while the vertical bar represents the value range across locations; mean = $35.82 \mu\text{g/L}$, standard deviation = $19.43 \mu\text{g/L}$, minimum = $11.04 \mu\text{g/L}$, and maximum = $108.35 \mu\text{g/L}$.

Sample analysis was conducted according to EPA-445 [34]. In particular, 1-liter water samples were collected from a depth of 0.2–0.3 m below the lake surface and filtered on $0.7 \mu\text{m}$ Whatman GF/F filter pads, and then stored in the dark at -20°C pending laboratory analysis. Filtering was performed under subdued light, soon after sampling, since algal populations, and thus Chl-a concentration, are susceptible to change in relatively short periods of time. Chl-a was extracted using 6 mL of 90% acetone for 24 h [34]. The Chl-a concentration was corrected for pheophytin and measured fluorometrically using a TD-7210 Fluorometer (Turner Designs, CA, USA) fitted with a daylight white lamp and properly calibrated chlorophyll optical kit.

Laboratory analysis of the samples shows that the lake is in hypereutrophic state [35], with an annual average Chl-a concentration of $35.82 \mu\text{g/L}$; the distribution of concentrations across time and measurements sites are shown in Figure 1b.

Atmospheric correction was conducted on the Level 1C Sentinel-2 images with the Polymer method v.4.6 [36], using default parameters. Polymer is an atmospheric correction algorithm for processing oceanic waters with and without sun-glint presence [37], and has been reported to show good performance [8]. Finally, all bands were resampled to 10 m/pixel spatial resolution via bilinear interpolation. After removing noisy pixels close to the shore and the in situ sampling locations, approximately 74 thousand lake surface pixels remain (Figure 2c) for which Chl-a estimation values are unknown, forming an unlabeled dataset of approximately 2.3 million samples across all acquisition dates (Table 1).

Table 1. The number of samples per month for labeled and unlabeled datasets.

	Month												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
Unlabeled	148 K	222 K	222 K	222 K	222 K	296 K	222 K	148 K	148 K	296 K	148 K	74 K	2365 K
Labeled	20	30	30	30	30	40	30	20	20	40	20	10	320

2.2. Proposed Method

Chl-a levels of inland waters are known to be positively correlated with water temperature and fertilizer inflow often resulting from rainfall, among other meteorological parameters [38]. In other words, one can expect a certain level of correlation between Chl-a levels and the months of a year, as regular seasonal fluctuations occur (i.e., higher concentrations during sunnier seasons and after rains). Based on this observation, we hypothesize that a network trained to guess the month of a given remote sensing spectral or spectral/spatial sample (with no associated Chl-a information/label) will be able to indirectly learn features helpful for Chl-a level estimation.

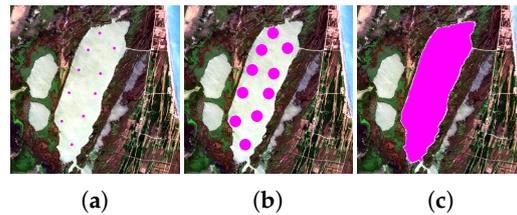


Figure 2. The pink regions denote the unlabeled samples used in our ablation experiments; centered on each of the in situ sampling positions with (a) a disk of diameter 11 pixels, (b) a disk of diameter 45 pixels, and (c) the entire lake surface, yielding 28,160, and 503,040, and 2,364,928 unlabeled samples in 32 images, respectively.

To avoid any confusion, we assume that the date of acquisition is always available for the remote sensing images under study, and in the rest of this article “labeled dataset” will refer to the limited image samples of the water body for which the Chl-a concentration levels are known, and “unlabeled dataset” will refer to the rest of the image samples of the water body.

The aforementioned concepts can be implemented in various ways. For instance, one could first train a convolutional network with unlabeled image locations, to learn the month of acquisition until a reasonable level of performance, and then switch the loss function to regression with a smaller learning rate and continue training with the labeled samples as a form of “forward” transfer learning. However, despite its simplicity, this approach risks directing the network to search space regions too far from the optima of the main task of Chl-a estimation, to which the network is only afterwards introduced.

Instead, a superior alternative is to train the network’s feature extractor for both tasks simultaneously as a multitasking network (Figure 3), where the main task is Chl-a regression and the auxiliary task is month classification. The two tasks form the two branches following the feature extraction layers, and direct the network to learn features useful for both tasks, thus reinforcing each other. Moreover, being able to benefit from the often far greater amount of unlabeled data in this way can have a strong regularization effect, thus avoiding overfitting.

2.2.1. Input Shape

Since the availability of spatial information has been shown to enable networks to exploit correlations among neighboring pixels [28,29], it was chosen to provide the input as a 3D tensor of the shape $b \times k \times k$. It is formed by stacking b multispectral image patches of size $k \times k$ pixels centered on the pixel under study.

2.2.2. Network Architecture

We adapted the 2D CNN architecture presented in [29] as the feature extractor shared across tasks. It consists of four convolutional layers without pooling, as the input patches are already small, and two fully connected (FC) layers. Before the convolutional layers, 1-pixel wide padding is applied to be able to use them with 3×3 kernel sizes. Each convolutional layer is followed by a *tanh* activation function. Then, the feature map generated by the 4th convolutional layer is fed into each of the two task-specific branches, each consisting of two FC layers. These task-specific branches are identical except for the number of neurons in the second FC layer. In this layer, the regression branch has 1 neuron to estimate the continuous Chl-a values, while the classification branch has 12 neurons to predict the sample acquisition month. The number of shared layers (four) and the learning rate of each branch were determined through ablation studies (presented in the next section). The overall network architecture is shown in Figure 3.

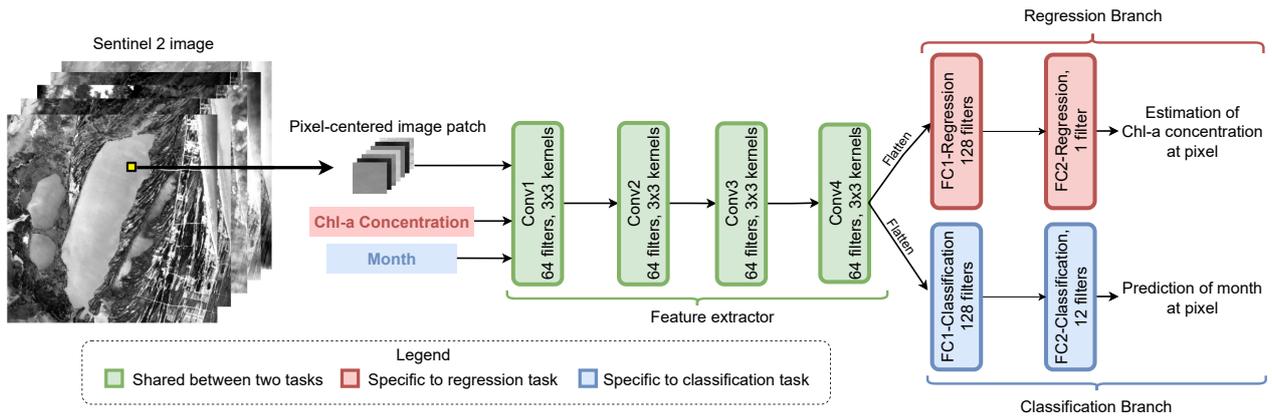


Figure 3. The outline of the proposed multitask learning CNN architecture where Conv and FC denote convolutional layer and fully connected layer, respectively.

2.2.3. Loss Functions

More formally, consider a Chl-a estimation dataset $\mathcal{D}_L = \{(x_i, y_i, m_i)\}_{i=1, \dots, |\mathcal{D}_L|}$, called the labeled dataset, where each patch x_i has Chl-a concentration level y_i and is labeled with the acquisition month m_i for $x_i \in \mathbb{R}^{b \times k \times k}$, $y_i \in \mathbb{R}$, and $m_i \in \{1, 2, \dots, 12\}$. We also have an unlabeled dataset $\mathcal{D}_U = \{(x_i, m_i)\}_{i=1, \dots, |\mathcal{D}_U|}$ where each patch x_i is labeled only with the acquisition month m_i . The labeled dataset \mathcal{D}_L is further split into a training set \mathcal{T} and a test set \mathcal{H} with $|\mathcal{T}|$ and $|\mathcal{H}|$ samples, respectively, where $|\mathcal{T}| + |\mathcal{H}| = |\mathcal{D}_L|$.

The goal is to learn a model Φ^θ with parameters θ to correctly estimate the Chl-a concentration of a given sample x as $y = \Phi^\theta(x)$. Contrary to existing approaches [28–32,39] that rely on a single network architecture for the task of Chl-a retrieval, our approach consists of a feature extractor Φ_{fe}^θ shared between two task-specific branches, namely $\Phi_{reg}^{\theta, FC}$ and $\Phi_{cls}^{\theta, FC}$, such that the model produces an estimation for the main task, $\Phi_{reg}^\theta = (\Phi_{fe}^\theta, \Phi_{reg}^{\theta, FC})$, and a prediction for the auxiliary task, $\Phi_{cls}^\theta = (\Phi_{fe}^\theta, \Phi_{cls}^{\theta, FC})$.

MSE loss: Following the standard approach in Chl-a estimation studies [29,30,32], the MSE (mean squared error) loss is used for the main regression task. The loss is evaluated only for the training set \mathcal{T} of the labeled dataset \mathcal{D}_L ,

$$\mathcal{L}_{Reg}(x_{(i)}, y_{(i)}; \Phi_{reg}^\theta) = \frac{1}{|\mathcal{B}_{reg}|} \sum_i (y_{(i)} - \Phi_{reg}^\theta(x_{(i)}))^2 \quad (1)$$

where (i) represents the i th batch of the training set \mathcal{T} , $|\mathcal{B}_{reg}|$ is the batch size of the regression task, and $\Phi_{reg}^\theta(x_{(i)})$ and $y_{(i)}$ denote, respectively, the estimated and target Chl-a concentrations.

Cross-entropy loss: As far as the auxiliary task is concerned, a “standard” classification loss is used to exploit the unlabeled image samples by predicting their acquisition months. It is evaluated for both the training set \mathcal{T} of the labeled dataset \mathcal{D}_L and the unlabeled dataset \mathcal{D}_U , since the samples of both datasets are labeled with the acquisition month:

$$\mathcal{L}_{Cls}(x_{(i)}, m_{(i)}; \Phi_{cls}^\theta) = -\frac{1}{|\mathcal{B}_{cls}|} \sum_i m_{(i)} \log(\Phi_{cls}^\theta(x_{(i)})) \quad (2)$$

where $\mathcal{D} = \mathcal{T} \cup \mathcal{D}_U$, and $(x_{(i)}, m_{(i)}) \in \mathcal{D}$, (i) represents the i th batch of \mathcal{D} , $|\mathcal{B}_{cls}|$ is the batch size of the classification task, and $\Phi_{cls}^\theta(x_{(i)})$ and $m_{(i)}$ denote, respectively, the predicted and target months.

2.2.4. Multitask Loss

The design of the multitask loss function is of critical importance. A naive multitask loss \mathcal{L}_{MT} could be written as a simple addition of the respective task losses, $\mathcal{L}_{MT} = \mathcal{L}_{Reg} + \mathcal{L}_{Cls}$. However, the range of values produced by \mathcal{L}_{Reg} and \mathcal{L}_{Cls} can vary strongly,

especially considering that they belong to different loss families, thus possibly leading the network to ignore one of the tasks. A better alternative is instead to express the total network loss \mathcal{L}_{MT} as a weighted combination of the task-specific losses $\mathcal{L}_{MT} = c_{Reg}\mathcal{L}_{Reg} + c_{Cls}\mathcal{L}_{Cls}$. It is observed in [40] that by adjusting appropriate values for c_{Reg} and c_{Cls} , the optimization of \mathcal{L}_{MT} ensures that no task is left behind when the other one is being learned. This can be achieved by manually computing the coefficient c_τ of each task \mathcal{L}_τ , where $\tau \in T$ and $T = \{Reg, Cls\}$, but this is a tedious and error-prone process that also does not take into account changes in loss during the training phase. Instead, these loss coefficients c_τ can be learned by the model by adding them to the trainable network parameters θ as $\omega_\tau = (\theta, c_\tau)$. To define the multitask objective function with the aforementioned loss coefficients, we used an implementation (<https://github.com/Mikoto10032/AutomaticWeightedLoss>, accessed on 16 December 2021) of Liebel and Körner's approach [41]:

$$\mathcal{L}_{MT}(x, y_\tau; \omega_\tau) = \sum_{\tau \in T} \frac{1}{2c_\tau^2} \mathcal{L}_\tau(x, y_\tau; \omega_\tau) + \ln(1 + c_\tau^2) \quad (3)$$

where $\tau \in T$ and $T = \{Reg, Cls\}$, $y_\tau = y$ for $\tau = Reg$ and $y_\tau = m$ for $\tau = Cls$.

In the next section, we report our numerical results and detail the effects of multitask learning, patch size, length of the shared feature extractor, number of unlabeled samples, and task loss weighting scheme on the Chl-a estimation performance through ablation studies.

3. Results and Discussion

The experiments presented in this section aimed to test our hypothesis of whether unlabeled data can contribute to the Chl-a estimation performance via multitask learning.

3.1. Data Split

Contrary to the state-of-the-art, where single random splits are common [16,21,27,31,32,42], a 10-fold cross-validation strategy was adopted to avoid overfitting. Subsequently, the training data is further split in order to produce a validation set with the same size as the test set to be used for hyperparameter tuning. Consequently, in each fold we have 256 samples (80%) for training, 32 samples (10%) for validation, and 32 samples (10%) for testing. After the hyperparameters are tuned, the model is trained without generating the validation set, since the size of the dataset is relatively small. Thus, the final split has 288 samples (90%) for training and 32 samples (10%) for testing.

3.2. Setup

The proposed multitask learning model was compared against several alternative approaches from the state-of-the-art, including its single-task regression version [29], support vector regression (SVR) [15] as a traditional machine learning method using a radial basis function kernel, multilayer perceptrons (MLP) with various numbers of hidden layers [30] as nonconvolutional neural networks, where the total number of trainable parameters was adjusted to be equal to the proposed model, a recently introduced morphology-based spectral-spatial approach (EFAL) [22], a convolutional neural network (CNN), as proposed in [28] adapted for regression, and of course a statistical spectral band characteristic known to be effective for Chl-a description [16] (band ratio): R_{705}/R_{665} , $(R_{665}^{-1} - R_{705}^{-1}) \times R_{740}$, R_{490}/R_{443} .

3.3. Training

Since the sizes of the labeled and unlabeled training sets differ by several orders of magnitude, it was not a plausible approach to set the batch sizes to have the same number of iterations. Instead, batches of the labeled training set are reused to ensure the same number of iterations on both tasks. The learning rate is chosen to be 10^{-4} . To optimize the multitask network, the root mean square propagation (RMSProp) algorithm is used. Training lasts for 200 epochs as determined through experimentation. The batch size is chosen as 64 for

the labeled set and as 12,318 for the unlabeled set. The same parameters/hyperparameters determined by the validation set were used in all experiments to ensure consistency. Finally, the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used for performance measurement.

3.4. Results

According to the results presented in Table 2, the band ratio approach [16] provides results close to those of the DL models, as it was expertly derived to obtain Chl-a values in water bodies from Sentinel-2 images. As expected, all deep learning models outperformed SVR [15], confirming the reported superiority of deep learning with respect to classic machine learning. EFAL [22] provides the second-best result thanks to its ability to represent pixels with a hierarchy of connected components, which makes it suitable for Chl-a estimation, since it provides a multiscale pixel description and adaptability in the spatial domain. CNN [28] performs similarly to MLP and single-task CNN models in that its performance is limited with the samples of the labeled dataset. For the MLP models [30], even though they have approximately the same number of trainable parameters with the proposed architecture, the estimation performance increases up to four layers and then decreases, possibly due to the model starting to overfit. The multitask learning model trained on both labeled and unlabeled data outperforms the single-task regression approach [29]. Despite the lack of Chl-a measurements in the unlabeled data, capturing seasonal appearances through the auxiliary task's month classification appears to help Chl-a estimation performance by almost 8 percentile points.

Table 2. Mean Chl-a estimation results across 10 folds of cross-validation with the best results in bold, where R^2 , MAE, and RMSE denote the coefficient of determination, mean absolute error, and root mean square error, respectively.

Method	Dataset	$R^2 \times 100$	MAE	RMSE	Model Parameters
Band ratio [16]	Labeled	71.37	0.147	0.199	-
SVR [15]	Labeled	51.82	0.193	0.27	-
EFAL [22]	Labeled	77.12	0.151	0.221	-
CNN [28]	Labeled	73.06	0.146	0.197	340 K
MLP (2 layers) [30]	Labeled	64.99	0.176	0.233	267 K
MLP (3 layers) [30]	Labeled	72.47	0.152	0.204	267 K
MLP (4 layers) [30]	Labeled	75.18	0.139	0.194	267 K
MLP (5 layers) [30]	Labeled	72.11	0.144	0.206	267 K
Single-task CNN [29]	Labeled	76.09	0.142	0.19	192 K
Multitask CNN (ours)	Labeled+unlabeled	83.89	0.107	0.154	267 K

In addition to the quantitative results, Chl-a estimation maps were generated via the proposed method for four dates in 2017 (one from each season) by training with the remaining years' data (Figure 4).

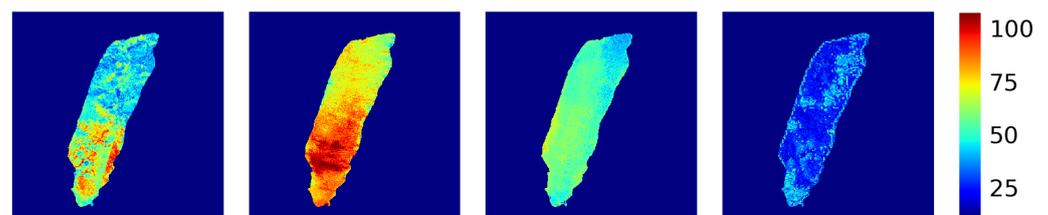


Figure 4. Chl-a estimation maps ($\mu\text{g/L}$) for spring, summer, fall, and winter 2017, respectively. They were generated using the proposed multitask model.

3.5. Ablation Study

Furthermore, the effects on performance of crucial design parameters were studied and are reported in this section: the number of shared layers within the architecture, the size of the input image patch, the number of unlabeled samples, and the task loss weighting scheme.

Since the number of layers in the proposed architecture is six, we experimented with the number of shared layers ranging from one to five by training five multitask models, starting with the first layer. The results obtained are reported in Table 3. The best performance is obtained by splitting the architecture into two task-specific branches after the fourth layer. This suggests that the multitask model benefits from sharing all of its convolutional features between the main regression task and the auxiliary classification task.

Table 3. Chl-a estimation levels of the proposed multitask model across 10-fold cross-validation with different number of shared layers among the two task branches, with the best results in bold, where R^2 , MAE, and RMSE denote the coefficient of determination, mean absolute error, and root mean square error, respectively (patch size = 3×3).

Shared Layers	1	2	3	4	5
$R^2 \times 100$	83.57	82.81	83.38	83.89	83
RMSE	0.156	0.159	0.158	0.154	0.16
MAE	0.108	0.11	0.109	0.107	0.113

For the sample patch size, experiments were conducted with sizes ranging from 1×1 to 9×9 pixels. Based on the results of Table 4, one can observe the contribution of spatial information with respect to pixelwise analysis. However, as the spatial resolution is low, larger patch sizes hardly improve it.

Table 4. Chl-a estimation levels of the proposed multitask model across 10-fold cross-validation with various patch sizes, with the best results in bold, where R^2 , MAE, and RMSE denote coefficient of determination, mean absolute error, and root mean square error, respectively (number of shared layers = 4).

Patch Size	1×1	3×3	5×5	7×7	9×9
$R^2 \times 100$	75.44	83.89	81.48	82.19	83.45
RMSE	0.194	0.154	0.167	0.163	0.158
MAE	0.134	0.107	0.114	0.112	0.105

In addition, to measure the effect of the number of unlabeled samples on performance, four multitask learning models were trained with progressively more unlabeled samples, ranging from none all the way to the entire lake surface. The intermediate lake masks with the employed unlabeled image locations are shown in Figure 2a–c, and the results are reported in Table 5. As expected, training the model with the utmost possible number of unlabeled samples by using the entire lake surface achieves the best performance. This is a promising result for lakes that are geographically difficult to access, and it obtains a large number of in situ samples.

The effect of the number of unlabeled samples on the auxiliary classification task was also examined. Similarly to the previous ablation for Chl-a estimation, four models were trained with an increasing number of unlabeled samples using the same lake intermediate masks as in Figure 2a–c. Likewise, the multitask learning model trained using all possible samples achieved the best performance, and their confusion matrices are shown in Figure 5.

Table 5. Chl-a estimation results of the proposed multitask model across 10-fold cross-validation with different number of unlabeled samples used for its auxiliary task, with the best results in bold, where R^2 , MAE, and RMSE denote coefficient of determination, mean absolute error, and root mean square error, respectively.

Unlabeled Samples	0	28160	503040	2364928
$R^2 \times 100$	76.63	82.41	83.03	83.89
RMSE	0.187	0.161	0.159	0.154
MAE	0.137	0.118	0.118	0.107

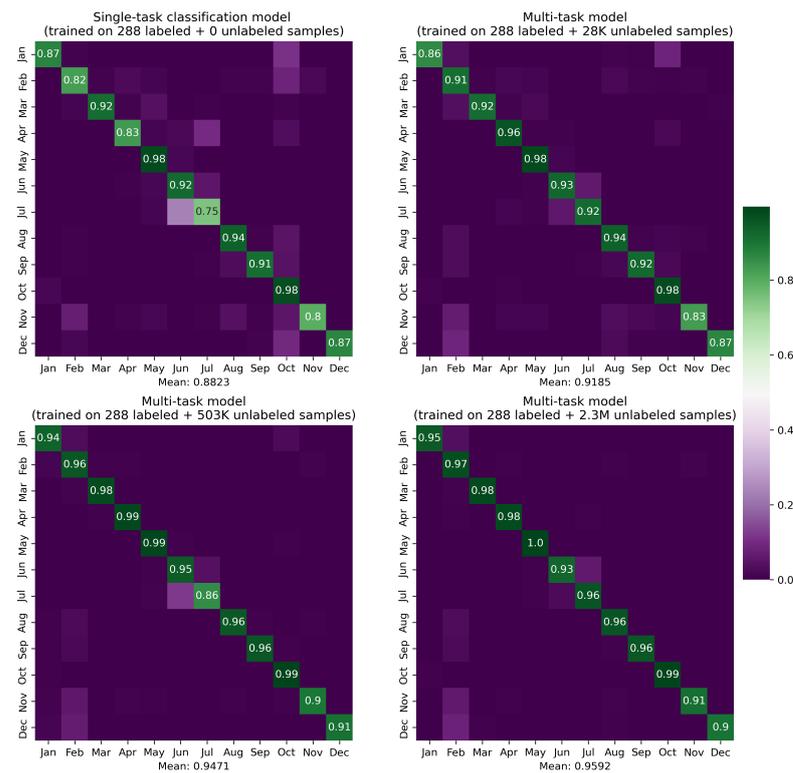


Figure 5. The confusion matrices devoted to the prediction of months.

Finally, a couple of experiments focusing on task-loss weighting were conducted, where a direct sum of task-specific losses was compared against a task-weighting scheme learned by the model. According to Table 6, the latter is superior, but with only a small margin.

Table 6. Chl-a estimation scores of the proposed multitask model across 10-fold cross-validation with two task loss weighting schemes: direct sum and the weighting learned by the model during training, with the best results in bold.

Task Loss Weight	Sum	Weight Learning [41]
$R^2 \times 100$	83.45	83.89
RMSE	0.156	0.154
MAE	0.113	0.107

3.6. Discussion and Limitations

The results presented so far provide relatively clear indications of the proposed method's practical interest with respect to its counterparts. The multitask network outperformed its single-task-regression-oriented alternative with a significant margin, and exhibited increasing performance as the number of unlabeled samples increased. Moreover, one can also safely presume that the large unlabeled dataset probably has an additional

regularization effect. In other words, using such a large dataset helps the model learn a better representation of the data. Thus, the auxiliary task satisfies the notorious data hunger of the deep learning model. However, neither the method nor the undertaken study are without limitations.

First of all, even though utmost care was taken to ensure the reliability of the estimation results and to avoid overfitting, this still does not change the fact that all of the samples originate from a single water body. Hence, it is unknown whether the estimation performance of the method under study can be reproduced with samples of a geographically distinct lake.

Furthermore, even though the samples acquired from Lake Balik demonstrated a significant deviation (from 11.04 $\mu\text{g/L}$ to 108.35 $\mu\text{g/L}$), the lake is still in overall hypereutrophic state. Consequently, the validity of the results obtained, across distinct mesotrophic or oligotrophic lakes, remains uncertain.

It is planned to rectify both limitations in the future, via studies in new lakes, pending financial support.

4. Conclusions

In this article, we proposed the first neural network architecture for water quality estimation that can exploit unlabeled satellite imagery. We addressed the scarcity of labeled data with a multitask CNN, where the main task is Chl-a regression and the auxiliary task is month classification for the unlabeled samples. We thus aimed to exploit the semiregular seasonal fluctuations of Chl-a in order to improve the computed features.

The experiments that were conducted on Lake Balik showed that the proposed multitask learning model increased the coefficient of determination by 7.8 percentile points compared to the single-task model, asserting the contribution potential of unlabeled data via the proposed technique. In fact, the progressive improvement of performance relative to the amount of employed unlabeled data was also shown.

Consequently, the proposed method renders possible the development of a Chl-a estimation model with a modest amount of labeled data, thus minimizing the effort required in terms of the number in situ samples, while enabling the exploitation of unlabeled remote sensing images that are abundantly available.

Moreover, since the only assumption was the semiregular seasonal fluctuations of Chl-a levels, the auxiliary month classification task, in our opinion, holds the potential of benefiting other multitemporal remote sensing analysis contexts with similarly behaving variables.

Nevertheless, despite the strong potential of these findings, and the rigorous experimental protocol to avoid overfitting, this study is limited all the same by the use of a single water body. Future work will concentrate on the validation of this approach with water bodies of distinct trophic indexes at alternative geographical locations. Furthermore, we will also focus on how to exploit multitemporal unlabeled data via month classification for land-cover/land-use map estimation.

Author Contributions: Conceptualization, E.A.; methodology, M.I. and E.A.; software, M.I.; validation, M.I. and E.A.; formal analysis, M.I. and E.A.; investigation, M.I. and E.A.; resources, E.A. and S.A.; data curation, S.A.; writing—original draft preparation, M.I.; writing—review and editing, M.I. and E.A.; visualization, M.I.; supervision, E.A.; project administration, E.A.; funding acquisition, E.A. and S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the TUBITAK Grant 118E258.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results

References

1. Caballero, I.; Steinmetz, F.; Navarro, G. Evaluation of the first year of operational sentinel-2A data for retrieval of suspended solids in medium- to high-turbidity waters. *Remote Sens.* **2018**, *10*, 982. [[CrossRef](#)]
2. Kane, D.D.; Conroy, J.D.; Richards, R.P.; Baker, D.B.; Culver, D.A. Re-eutrophication of Lake Erie: Correlations between tributary nutrient loads and phytoplankton biomass. *J. Great Lakes Res.* **2014**, *40*, 496–501. [[CrossRef](#)]
3. Brooks, B.W.; Lazorchak, J.M.; Howard, M.D.; Johnson, M.V.V.; Morton, S.L.; Perkins, D.A.; Reavie, E.D.; Scott, G.I.; Smith, S.A.; Stevens, J.A. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environ. Toxicol. Chem.* **2016**, *35*, 6–13. [[CrossRef](#)] [[PubMed](#)]
4. Ritchie, J.C.; Zimba, P.V.; Everitt, J.H. Remote sensing techniques to assess water quality. *Photogramm. Eng. Remote Sens.* **2003**, *69*, 695–704. [[CrossRef](#)]
5. Laliberté, J.; Larouche, P.; Devred, E.; Craig, S. Chlorophyll-a Concentration Retrieval in the Optically Complex Waters of the St. Lawrence Estuary and Gulf Using Principal Component Analysis. *Remote Sens.* **2018**, *10*, 265. [[CrossRef](#)]
6. Cui, T.W.; Zhang, J.; Wang, K.; Wei, J.W.; Mu, B.; Ma, Y.; Zhu, J.H.; Liu, R.J.; Chen, X.Y. Remote sensing of chlorophyll a concentration in turbid coastal waters based on a global optical water classification system. *ISPRS J. Photogramm. Remote Sens.* **2020**, *163*, 187–201. [[CrossRef](#)]
7. Johansen, R.; Beck, R.; Nowosad, J.; Nietch, C.; Xu, M.; Shu, S.; Yang, B.; Liu, H.; Emery, E.; Reif, M.; et al. Evaluating the portability of satellite derived chlorophyll-a algorithms for temperate inland lakes using airborne hyperspectral imagery and dense surface observations. *Harmful Algae* **2018**, *76*, 35–46. [[CrossRef](#)]
8. Pahlevan, N.; Smith, B.; Schalles, J.; Binding, C.; Cao, Z.; Ma, R.; Alikas, K.; Kangro, K.; Gurlin, D.; Hà, N.; et al. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. *Remote Sens. Environ.* **2020**, *240*, 111604. [[CrossRef](#)]
9. Dörnhöfer, K.; Oppelt, N. Remote sensing for lake research and monitoring—Recent advances. *Ecol. Indic.* **2016**, *64*, 105–122. [[CrossRef](#)]
10. Huang, C.; Chen, X.; Li, Y.; Yang, H.; Sun, D.; Li, J.; Le, C.; Zhou, L.; Zhang, M.; Xu, L. Specific inherent optical properties of highly turbid productive water for retrieval of water-quality after optical classification. *Environ. Earth Sci.* **2015**, *73*, 1961–1973. [[CrossRef](#)]
11. Giardino, C.; Oggioni, A.; Bresciani, M.; Yan, H. Remote sensing of suspended particulate matter in Himalayan lakes. *Mt. Res. Dev.* **2010**, *30*, 157–168. [[CrossRef](#)]
12. Lan, X.; Guo, Z.; Tian, Y.; Lei, X.; Wang, J. Retrieval of water quality parameters by neural network and analytical algorithm in Guanting Reservoir in Hebei Province in China. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy, 26–31 July 2015; pp. 723–726.
13. Verrelst, J.; Rivera, J.P.; Gitelson, A.; Delegido, J.; Moreno, J.; Camps-Valls, G. Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 554–567. [[CrossRef](#)]
14. Vilas, L.G.; Spyarakos, E.; Palenzuela, J.M.T. Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of Galician rias (NW Spain). *Remote Sens. Environ.* **2011**, *115*, 524–535. [[CrossRef](#)]
15. Wang, X.; Ma, L.; Wang, X. Apply semi-supervised support vector regression for remote sensing water quality retrieving. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Honolulu, HI, USA, 25–30 July 2010; pp. 2757–2760.
16. Ansper, A.; Alikas, K. Retrieval of Chlorophyll a from Sentinel-2 MSI Data for the European Union Water Framework Directive Reporting Purposes. *Remote Sens.* **2019**, *11*, 64. [[CrossRef](#)]
17. O'Reilly, J.E.; Maritorena, S.; Mitchell, B.G.; Siegel, D.A.; Carder, K.L.; Garver, S.A.; Kahruand, M.; McClain, C. Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.* **1998**, *103*, 24937–24953. [[CrossRef](#)]
18. Freitas, F.H.; Dierssen, H.M. Evaluating the seasonal and decadal performance of red band difference algorithms for chlorophyll in an optically complex estuary with winter and summer blooms. *Remote Sens. Environ.* **2019**, *231*, 111228. [[CrossRef](#)]
19. Matthews, M.W.; Bernard, S.; Robertson, L. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* **2012**, *124*, 637–652. [[CrossRef](#)]
20. Pahlevan, N.; Sarkara, S.; Franza, B.A.; Balasubramanian, S.V.; Hec, J. Sentinel-2 multispectral instrument (MSI) data processing for aquatic science applications: Demonstrations and validations. *Remote Sens. Environ.* **2017**, *201*, 47–56. [[CrossRef](#)]
21. Niroumand-Jadidi, M.; Bovolo, F.; Bruzzone, L. Novel Spectra-Derived Features for Empirical Retrieval of Water Quality Parameters: Demonstrations for OLI, MSI, and OLCI Sensors. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 10285–10300. [[CrossRef](#)]
22. Aptoula, E.; Ariman, S. Hierarchical Spatial-Spectral Features for the Chlorophyll-a Estimation of Lake Balik, Turkey. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1500405. [[CrossRef](#)]
23. Neil, C.; Spyarakos, E.; Hunter, P.D.; Tyler, A.N. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. *Remote Sens. Environ.* **2019**, *229*, 159–178. [[CrossRef](#)]

24. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [[CrossRef](#)]
25. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 1, pp. 1097–1105.
26. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
27. Peterson, K.T.; Sagan, V.; Sloan, J.J. Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. *GIScience Remote Sens.* **2020**, *57*, 510–525. [[CrossRef](#)]
28. Pu, F.; Ding, C.; Chao, Z.; Yu, Y.; Xu, X. Water-Quality Classification of Inland Lakes Using Landsat-8 Images by Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 1674. [[CrossRef](#)]
29. Aptoula, E.; Ariman, S. Chlorophyll-a Retrieval From Sentinel-2 Images Using Convolutional Neural Network Regression. *IEEE Geosci. Remote Sens. Lett.* **2021**, 1–5. [[CrossRef](#)]
30. Syariz, M.; Lin, C.; Nguyen, M.; Jaelani, L.; Blanco, A. WaterNet: A Convolutional Neural Network for Chlorophyll-a Concentration Retrieval. *Remote Sens.* **2020**, *12*, 1966. [[CrossRef](#)]
31. Cho, H.; Choi, U.; Park, H. Deep Learning Application to Time Series Prediction of Daily Chlorophyll-a Concentration. *WIT Trans. Ecol. Environ.* **2018**, *215*, 157–163.
32. Barzegar, R.; Aalami, M.; Adamowski, J. Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [[CrossRef](#)]
33. General Directorate for the Protection of Natural Assets. *Samsun Kizilirmak Deltasi Sulak Alan Ve Kus Cenneti Dogal Sit Alanlari Yonetim Plani*; Technical Report; Ministry of Environment and Urban Planning: Ankara, Turkey, 2017.
34. Arar, E.J.; Collins, G.B. *Method 445.0 In Vitro Determination of Chlorophyll-a and Pheophytin-a in Marine and Freshwater Algae by Fluorescence*; Technical Report; U.S. Environmental Protection Agency: Washington, DC, USA, 1997.
35. Organization for Economic Co-Operation and Development. *Eutrophication of Waters: Monitoring, Assessment and Control*; Organization for Economic Co-Operation and Development: Washington, DC, USA, 1982.
36. Steinmetz, F.; Deschamps, P.Y.; Ramon, D. Atmospheric correction in presence of sun glint: Application to MERIS. *Opt. Express* **2011**, *19*, 9783–9800. [[CrossRef](#)]
37. Pereira-Sandoval, M.; Ruescas, A.; Urrego, P.; Ruiz-Verdu, A.; Delegido, J.; Tenjo, C.; Soria-Perpinya, X.; Vicente, E.; Soria, J.; Moreno, J. Evaluation of Atmospheric Correction Algorithms over Spanish Inland Waters for Sentinel-2 Multi Spectral Imagery Data. *Remote Sens.* **2019**, *11*, 1469. [[CrossRef](#)]
38. Son, S.; Wang, M. Water properties in Chesapeake Bay from MODIS-Aqua measurements. *Remote Sens. Environ.* **2012**, *123*, 163–174. [[CrossRef](#)]
39. Choi, J.; Kim, J.; Won, J.; Min, O. Modelling Chlorophyll-a Concentration using Deep Neural Networks considering Extreme Data Imbalance and Skewness. In Proceedings of the IEEE International Conference on Advanced Communication Technology (ICACT), PyeongChang, Korea, 17–20 February 2019; pp. 631–634.
40. Kendall, A.; Gal, Y.; Cipolla, R. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7482–7491.
41. Liebel, L.; Körner, M. Auxiliary Tasks in Multi-task Learning. *arXiv* **2018**, arXiv:1805.06334.
42. Batur, E.; Maktav, D. Assessment of Surface Water Quality by Using Satellite Images Fusion Based on PCA Method in the Lake Gala, Turkey. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 2983–2989. [[CrossRef](#)]