A New Deep Learning approach based on Bilateral Semantic Segmentation Models for Sustainable Estuarine Wetland Ecosystem Management

Hanh Nguyen Pham[a], Kinh Bac Dang[b*], Thanh Vinh Nguyen[a], Ngoc Cuong Tran[a], Xuan Quy Ngo[a], Duc Anh Nguyen[c], Thi Thanh Hai Phan[b], Thu Thuy Nguyen[d], Wenshan Guo[d], Huu Hao Ngo[d*]

[a] Nature and Biodiversity Conservation Agency, Vietnam Environment Administration, Ministry of Natural Resources and Environment, 10 Ton That Thuyet, Nam Tu Liem, Hanoi, Vietnam

[b] VNU University of Science, Vietnam National University, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam

[c] SKYMAP High Technology Co., Ltd., No.6, 40/2/1, Ta Quang Buu, Hai Ba Trung, Hanoi, Vietnam

[d] Center for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydney, NSW 2007, Australia

***** Corresponding author: Huu Hao Ngo, *E–mail:* ngohuuhao121@gmail.com;
Kinh Bac Dang, *E-mail*: dangkinhbac@hus.edu.vn;

Abstract:

Nowadays, estuarial areas have been strongly affected by the construction of electrical power dams from upstream, downstream urbanization and many types of hazards along the coastal regions. It has resulted in significant changes in estuarine wetland ecosystems between rainy and dry seasons. To avoid estuary vulnerability, monitoring and evaluation of the estuarine ecosystems are very critical tasks. The main goal of this research is to propose and implement a novel deep learning method in monitoring various ecosystems in estuarine regions. The processing speed and accuracy of common neural networks is improved more than ten times through spatial and context paths integrated into a novel Bilateral Segmentation Network (BiSeNet). The multi-sensor and multi-temporal satellite images (including Sentinel-2, ALOS-DEM, and NOAA-DEM images) served as input data. As a result, four BiSeNet models out of 20 trained models achieved a greater than 90% accuracy, especially for interpreting estuarine waters, intertidal forested wetlands, and aquacultural lands in subtidal regions. These models outperformed Random Forest and Support Vector Machine approaches. The best one was used to map estuarine ecosystems from 12 satellite images over a five-year period in the largest estuary in northern Vietnam. The ecosystem changes between dry and rainy seasons were analyzed in detail to assess the ecological succession in estuaries. Furthermore, this model can potentially update new estuarine ecosystem types in other estuarine areas across the world, making possible real-time monitoring and assessing estuarine ecological conditions for sustainable management of wetland ecosystem.

Keywords: Deep learning; Sustainability; Bilateral semantic segmentation; Estuary; Wetland.

1. Introduction

Around the world, more than 60% of the large estuaries are located within 25 kilometers of metropolitan areas with a population of more than 100,000 (Lee and Yeh, 2009; Pye and Blott, 2014). Nearly two-thirds of all estuarine mangrove forests are situated near urbanized coastal regions (Sannigrahi et al., 2019). People rely on coastal fisheries as their main income; the total value of these fisheries is estimated to be worth at least $34 billion yearly (Luisetti et al., 2014). Due to increasing urbanization and climate change, estuaries and coral reefs have been assessed as the most endangered ecosystems in coastal regions (Heywood, 2011; de Groot et al., 2016; Oiry and Barillé, 2021). Poor watershed management often results in the deterioration of the estuary environment. Particularly in the Southeast Asian countries, constructing too many hydroelectric dams leads to a large reduction of alluvium in the river, not only causing a shortage of sediment in the middle region, but also increasing the possibility of coastal erosion in the estuary area (Pereira et al., 2016). With nearly 40 hydropower projects built on the upper Red River in China and Vietnam, the amount of flow on this river in 2019 fell by about 40% compared to the average in the last century, causing a reduction in the amount of water in the river (Swainson, 2008; Le et al., 2020). The river bottom elevation dropped to nearly 3m, especially in the last few months of 2019. Therefore, monitoring changes occurring in estuarial regions, especially with the wetland area has become an important job for coastal managers.

More than 1,200 estuaries in the world have been identified and mapped, totaling roughly 500,000 square kilometers of digitized land (Basset et al., 2013; Li et al., 2021). The changes in the area of estuarine ecosystems in many countries, particularly in Vietnam, have been monitored poorly based on investigation surveys every five years (Le et al., 2020; Nguyen et al., 2021). It requires lots of works and money for sampling, interview surveys, and satellite data processing at province and national levels (Cannizzaro et al., 2019). A machine understanding and learning all

experience of interpreters in estuarine ecosystem classification can become a better solution for coastal managers, to quickly observe the changes of the estuary. Accordingly, they can provide timely sufficient decisions for economic development, as well as natural conservation.

The semantic segmentation method is now starting to be used in machine learning models, especially in separating the coastal and wetland land cover regions for each overlay feature on the remote sensing image into a separate layer (Quoc Vo et al., 2015; Guo et al., 2020; Hu et al., 2021). Some networks use semantic segmentation methods such as the Convolutional neural network (CNN) which encodes high-level semantic information with a separating process of sample data (Abubakar and Boukari, 2018; DeLancey et al., 2020). However, this method does not preserve the most detailed spatial information of the input image. Some models such as PSPNet, DeepLab v2 and Deeplab v3 have been integrated with more convolutional layers to preserve spatial information throughout the process (Krestenitis et al., 2019; Yao et al., 2019; Hu et al., 2021). In particular, the U-shaped network model can recover the spatial information of input data after the encoding and decoding process (Garg et al., 2019; Stoian et al., 2019; Diakogiannis et al., 2020a). These two stages are continuously and multidimensionally connected, helping to retain all spatial and attribute information. However, these networks often encode data quite deeply, generate many parameters, consume memory space, and it is difficult to recover both spatial and attribute information during the decoder process.

To overcome the problem of losing spatial information on images, many studies have used U-shaped network structures for image analysis (Zhang et al., 2018; Garg et al., 2019; Stoian et al., 2019; Diakogiannis et al., 2020a). Accordingly, the U-shaped structure gradually increases the spatial resolution but this technique still has two shortcomings: firstly, complete U-shaped structure which computes on high resolution objects after encoding, can reduce the data processing speed of the model; and secondly, the separation of the input data will be difficult to

recover on the decoder side, causing much spatial information to be lost (Rakhlin et al., 2018; Giang et al., 2020). Consequently, the U-Net models still need further improvement for large volumes of high-resolution data.

This study developed an advanced U-structured neural network – known as Bilateral Segmentation Network (BiSeNet) models, which helps to increase processing speed and accuracy in interpreting estuarine waters, intertidal forested wetlands, and aquacultural lands in subtidal regions. The processing speed of the BiSeNet models for new prediction is faster than SegNet models more than ten times and than Enet models more than 1.5 times (Yu et al., 2018). Based on the encoding context and spatial information separately in two paths, this network has become a potential method to preserve the data information from original images and minimize the loss of spatial information in the decoding data (Rakhlin et al., 2018; Giang et al., 2020). Data for models testing and validation was collected from two estuaries in the northeast region of Vietnam: Tien Yen estuary (Quang Ninh province) and Ba Lat estuary (Thai Binh and Nam Dinh provinces). In particular, the study uses multi-sensor remote sensing data including Sentinel-2, ALOS and NOAA to build the model. The resulting model is compared with previous traditional models to evaluate the effectiveness of well-known machine learning models which include Support Vector Machine and Random Forest.

2. Material and methods

2.1. Field data collection

Field data is collected in the Ba Lat estuary area, with a coastline of about 35 km from the south of Thai Binh province to the north of Nam Dinh province (Duc et al., 2012). Additionally, to collect more spectral information from the estuarine area, the authors chose the Tien Yen estuarine, Quang Ninh province (Figure 1).

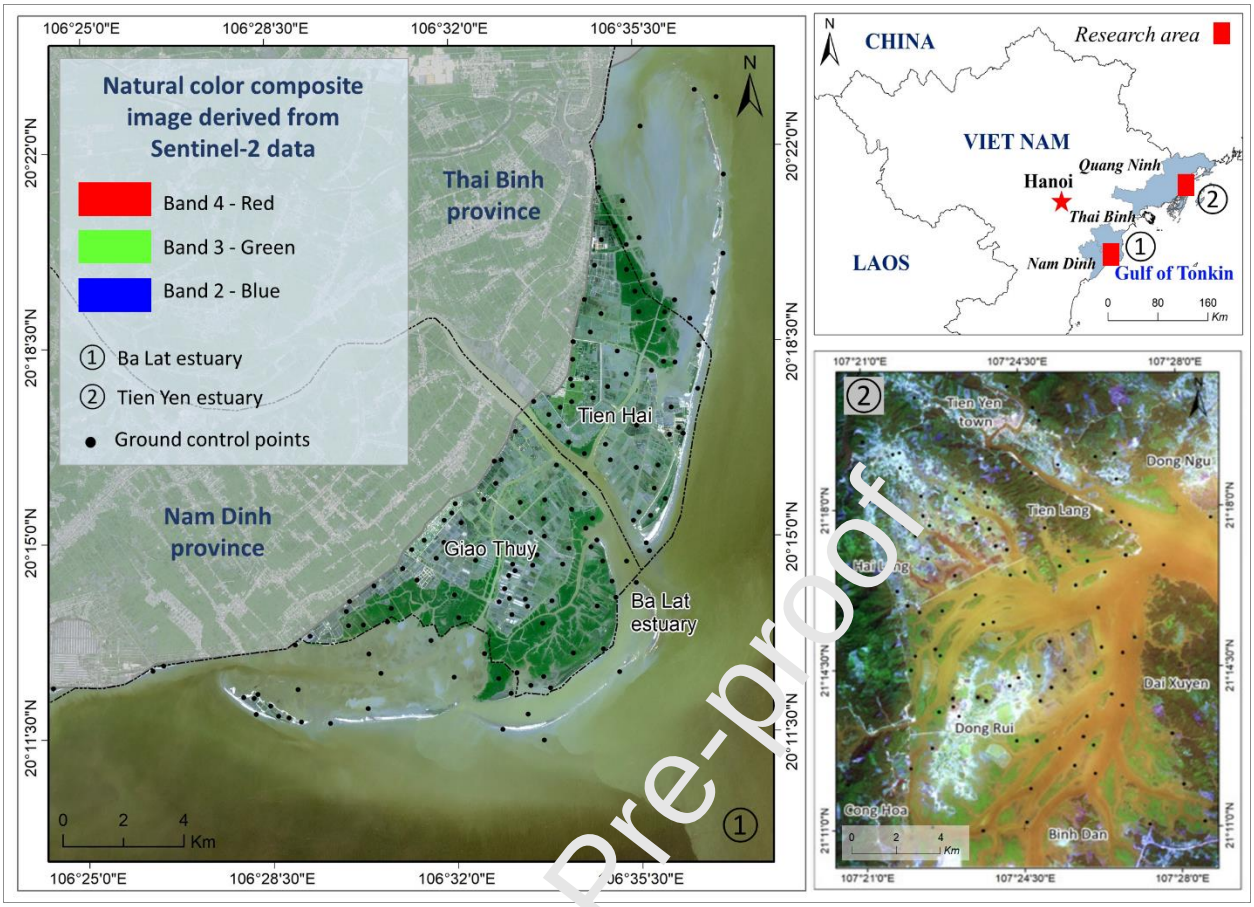

Figure 1. Location of samples collected in Ba Lat and Tien yen estuaries, Vietnam.

Regarding the distribution of ecosystems, based on the MONRE/VEA/BCA/JAICA-NBDS project implemented in 2012 and 2013 and the Ramsar Convention's wetland classification, the authors identified the main wetland ecosystems in the study area that includes: (1) Intertidal forested wetlands, (2) intertidal marshes, (3) farm ponds, (4) sand, shingle or pebble shores outside the river mouth (prominent at Con Lu, Con Xanh, and Du Mo), (5) tributaries/tidal creeks, (6) estuary waters; and (7) seasonally flooded agricultural land (No.07/2020/TT-BTNMT, 2020).

Due to the shoreline of the Ba Lat estuary changes complicatedly. Its river flow dynamics prevail in the Ba Lat estuary area due to the large alluvial discharge from the Red River (about 81 million tons of suspended sediment annually). Many geomorphological studies have analyzed

shoreline changes in this area. For example, (Vu, 2018) calculated that the erosion rate increased by 20% in the 1990s, although the accretion trend still prevails from the 1960s to the present. However, (Le *et al.*, 2020) clarified that the northern dunes in Tien Hai district, Thai Binh province move towards the inland from 150 to 250 m in the 2010s, equivalent to equivalent displacement speed from 30 to 50 m/year. The southern dunes in Giao Thuy district, Nam Dinh province move towards the shore from 180 to 280 m, equivalent to the average displacement speed from 40 to 60 m/year.

2.2. BiSeNet Set-up for Estuarine Ecosystem Classification

2.2.1 Network Architecture

In this study, the BiSeNet structure is developed with two parts: Spatial Path (SP) and Context Path (CP), which differ from the U-Net model structure with two decoder and encoder paths (Yu et al., 2018; Neupane et al., 2021). The two BiSeNet paths are devised to have two separate tasks, which help to reduce the loss of spatial information and reconstructing information in corresponding Receptive Fields (RF) (Figure 2). The BiSeNet parameters convert original data into new states depending on predefined characteristics in each layer (Dai et al., 2020). The SP and CP architectures have been refined using five consecutive kinds of layers including:

(1) Input layers: are created to enter the raw image pixels of input data into the preprocessing step. The raw Sentinel-2 images (with four bands: red, green, blue, and near-infrared bands) were combined with the DEM data in this research.

(2) Convolutional layers (CONV): CONV layers compute new outputs using filtering layers that are narrower than the input values (Albawi et al., 2018). The selected filter size in this research is 3x3. The filter moves over the objects, connecting them to local areas. 52 CONV layers were

chosen to build the BiSeNet architecture. Filters of 64, 128, 256, and 512 were utilized for the 52 CONV layers to reduce training and validation time.

(3) Batch Normalization layers: normalize the results from the CONV layer to the same size as the previous measurement (Kim and Jeong, 2019). Two parameters can be modified for the outcomes during the normalization process including the mean (β) and variance (or standard deviation - γ) parameters as follows:

$$y_i = \gamma \hat{x_i} + \beta \quad (1)$$

where $\hat{x_i}$ is determined based on the calculation of mean (μ_M) and variance (σ_M^2) of mini-batch M = {x1…n} as follows:

$$\mu_M \leftarrow \frac{1}{n}\sum_{i=1}^{n} x_i \quad (2)$$

$$\sigma_M^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_M)^2 \quad (3)$$

$$\hat{x_i} \leftarrow \frac{x_i - \mu_M}{\sqrt{\sigma_M^2 + \varepsilon}} \quad (4)$$

As batch normalization produces two parameters for each convolutional layer, the batch normalization layer comprises the following number of parameters:

$$P_{batch} = 4 * D_{Conv2D} \quad (5)$$

Where D_{Conv2D} is the depth of the convolutional layer. For example, the first batch normalization layer has a parameter count of 2x64 = 128.

(4) Activation Function layers: commonly put immediately after the CONV layers, and each matrix data after going through the convolution layer will go through the Activation Function layer such as ReLU or Sigmod layer to become non-linear (Nwankpa et al., 2018).

(5) Global average pooling (GAP) layers: Instead of creating a complex of feature matrices with multiple target layers, using the GAP helps to convert the target map of a layer into an average of all those class values (Albawi et al., 2018). This method helps to curtail the number of computational parameters in the model, reducing the spatial size of the output matrices of each class.

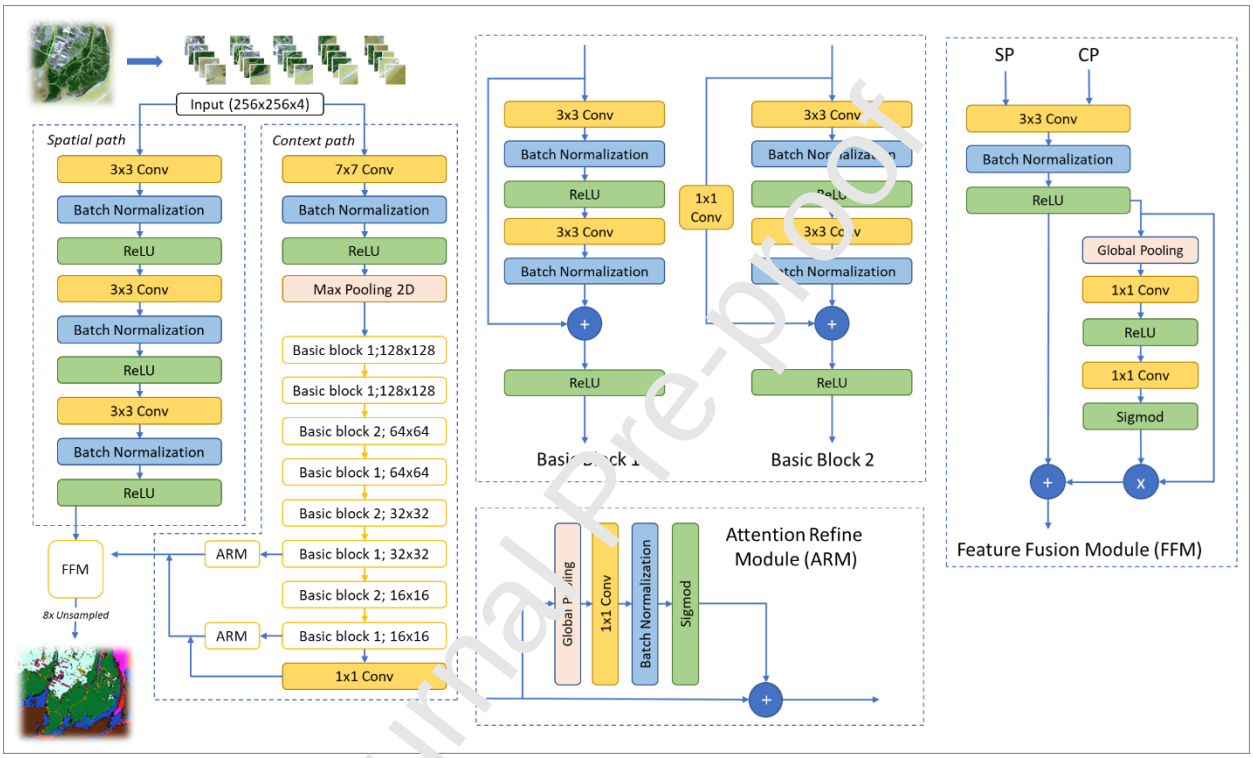

Figure 2. The structure of BiSeNet network for classification of estuary ecosystems with the input size of 256x256x4.

In the BiSeNet structure, the SP path was established with the aim of preserving spatial information at high resolution. This information will be encrypted and paired with the result to restore the spatial structure of the data. For the CP path, the model establishes a field that receives rapid sampling information, parallel to the SP path. Figure 2 shows the full structure of these two components. In order to ensure accuracy without reducing model processing speed, the two paths are integrated with a screening step for the final results through Feature Fusion Modules (FFM)

and Attention Refinement Modules (ARM) (Feng et al., 2019). The FFM and ARM modules contain a system of GAP layers so that the overall context of both the original image is understood, and recover this context after the computation, instead of applying up-sampling layers like conventional U-Nets. It is important to note that these modules help to reduce computational costs in model building.

2.2.2. Spatial Path (SP)

The SP path is used in the BiSeNet to retain as much spatial information as possible while encoding the detailed data in images as a whole for final prediction. Both the depth of models and receptive fields are not considered given the independent extraction procedure in this path. After the input layer of Sentinel-2 image, the data was run through three blocks. Each block includes three layers: a convolution layer with stride equal to two, a batch normalization layer, and an activation function (ReLU). After three blocks, the output data is eight times smaller than the original image.

2.2.3. Context path (CP)

Coding models of spatial and contextual data together often require computationally complex and memory-intensive processes, and reduced modeling speed. However, Yu et al. (2018) explained that the SP paths often have a shallow structure. The separation of spatial and context paths can potentially reduce memory and data processing speed.

The CP path generated in this study is based on the ResNet-18 structure. Therefore, it will include two types of basic blocks in its structure. To improve the overall processing time, the first convolutional layer of the CP path applies a filter with a size of 7x7, while the SP path utilize filters with a size of 3x3. The CP path uses lightweight modeling and GAP layers to provide the best receptive information. In this study, the computational model narrows down the sample size quickly (up to 16 times, from 256 to 128, 64, 32, and 16), to obtain the maximum information

about the features of the data classes, and to encode the correct contextual information. The down-sampling process can greatly reduce the sampling frequency, and sometimes generate aliasing issues in encoded data (Yu et al., 2018; Dai et al., 2020). Consequently, the information after being down-sampled to 32x32 and 16x16 will be run through the ARM to refine the encoded information (Hao et al., 2020). The input data of the module will be used to build a GAP instruction vector, before arriving at the main structure of the ARM block consisting of 1 CONV layer, 1 Batch Normalization layer and 1 Activation Function layer (with the use of the Sigmod function, instead of the ReLU function as in spatial path). This contextual information will be stored and used to refine the output in the next calculation of down-sampling rounds in the CP path. Based on this method without using up-sampling steps, the CP path of BiSenet can recover contextual information more easily and accurately than conventional U-Net networks.

2.2.4. Integrating spatial and contextual information

All this contextual information will be returned to the spatial information field to recreate the original context of the data before the final output is produced. However, the information contained in the SP and CP paths is completely different. The SP path carries information linking objects between those at the shallow encoding layer, while CP paths carry highly encoded attribute data information. Therefore, their integration will become more complex than with traditional U-Net models. The Feature Fusion Module (FFM) has been developed by Yang et al. (2019) for the purpose of connecting data from these two paths together. By integrating RF and SF information, the BiSeNet model creates an incomplete U-shaped neural network.

Based on the five layers described in section 2.3.1, the FFM module connects information between the SP and CP paths. In this case the "batch normalization" layer helps to balance two input data types and normalize data to the same computational size. The spatial and informational attributes after being connected will be weighted in separate vectors. Some important vectors

were learned to balance the valuable features of objects between the two paths through GAP computation, two Convolutional layers (1x1) and two different activation functions including the ReLU and Sigmod. Thereby the spatial and context data will be converted to the final weight representing the estuary ecosystem types.

2.3. Model development process

The deep learning models for mapping estuarine ecosystems are prepared in three steps: zoning estuarine regions, preparing input data, and training models. Figure 3 depicts the model development process for estuarine ecosystem categorization.

2.3.1. Zoning estuarine regions and preparing data

Estuaries, or transitional waterways are affected by both freshwater and marine ecosystems (Bournaris *et al.*, 2015; Kapetas *et al.*, 2019). Therefore, the estuarine zone along the coast will be identified not only in salty ocean and freshwater river, but also in brackish water (Pinto et al., 2010; Basset et al., 2013). From coastal regions to places below -6 m elevation, wetland habitats may be recognized according to the RAMSAR classification system (Northern Ireland Environment Agency, 2003; RAMSAR, 2010). Although the tidal level in the research areas is 2.8m, high dykes along all coasts in the northern part of Vietnam can prevent the salt water from sea getting into the inland areas (Tong et al., 2020). Therefore, as well as employing -6 m offshore depth as outer boundary, the dykes were used as inter-boundary sections in inland areas. In areas without a dyke, inland elevation data along the coastline was collected with every 2.5-meters contour lines at 1:5.000 scale for both Ba Lat and Tien Yen estuaries.

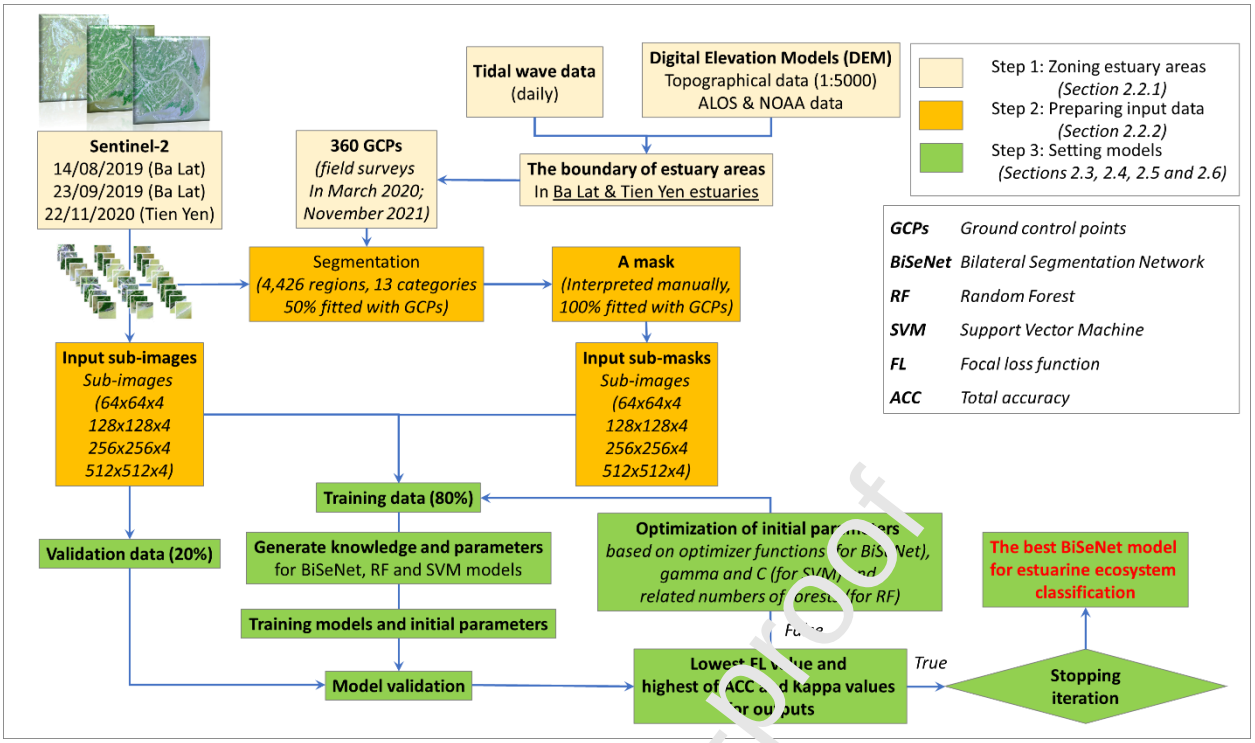

Figure 3. The structure to develop BiSeNet models for estuarine ecosystem classification.

Digital elevation models (DEM) are vital for input data regarding the first step and new prediction after completing the training process (sections 2.4 and 2.5). In addition to the topographical maps collected by Vietnam Academy of Science and Technology (VAST), the offshore and inland elevation at medium resolution was gathered from Advanced Land Observing Satellite (ALOS) and National Centers for Environmental Information (NOAA) satellite data. ALOS (Ghuffar, 2018; Dang et al., 2020a) retrieved 30-meter inland DEMs from the Google Earth Engine system produced by the Panchromatic Remote Sensing Instrument for Stereo Mapping (PRISM). However, ALOS satellite data only give inland elevation. The ALOS DEM's lowest value is zero, so for this reason the sea-land boundary was established at the value '0' inland border. The offshore DEM was obtained from the NOAA from Global Relief Data (Ozesmi and Bauer, 2002; NOAA, 2004). The DEM data was re-projected to WGS84/UTM horizontal datum – 48N and downscaled to a 30 meters resolution raster. The authors then used

ArcGIS software to integrate inland ALOS-DEM data with offshore NOAA-DEM data to create a full DEM.

2.3.2. Preparing input data

Different remote sensing data has been used for analyzing water environment such as MODIS-Aqua for phytoplankton analysis (Sun *et al.*, 2019), Landsat for water transparency assessment (Zhang *et al.*, 2021), and Sentinel-3A for organic carbon quantification (Xu *et al.*, 2020). To collect land cover information in the estuarine area, the Sentinel-2 (level-2 correction with spatial resolution of 10 meters) images were selected. Based on the medium-resolution satellite images, permanent and temporary wetlands, mangrove swamps, and others covered by saltwater or influenced by tides (Comittee, 2012; de Groot et al., 2016) will be differentiated. Also, in 2019 Sentinel-2 images collected twice in Ba Lat estuary (14/08/2019 and 23/09/2019) and once in Tien Yen estuary (22/11/2019) were used for training the BiSeNet model. The satellite images were obtained at 2.8 m tide. Additionally, the authors utilized 12 Sentinel-2 images taken from 2017 to 2021 to evaluate wetland changes in the Xuan Thuy National Park once the BiSeNet was completed (for more details see section 2.6).

Based on two field investigations, the Tien Yen estuarine ecosystems were validated in March 2020, while the Ba Lat estuarine ecosystems were validated in November 2021. Excepting inland areas, 12 ecosystems in total were investigated in the two estuarine regions (Figure 4) according to the international RAMSAR and Vietnamese wetland classification systems (Hoang and Le, 2006; Ramsar Convention Secretariat, 2013; Banks et al., 2019; Dang et al., 2020b). Here the heterogeneous rural areas and protection forests were added to the estuary area because of the specific communities and their locations in Vietnam (Hoang and Le, 2006; Guo et al., 2017). Local people always live near their farms and behind the protection forests to reduce the damage

done by extreme storms and coastal erosion (Duc et al., 2012; Takagi et al., 2015). For this reason, two additional ecosystems were added to the prediction model.

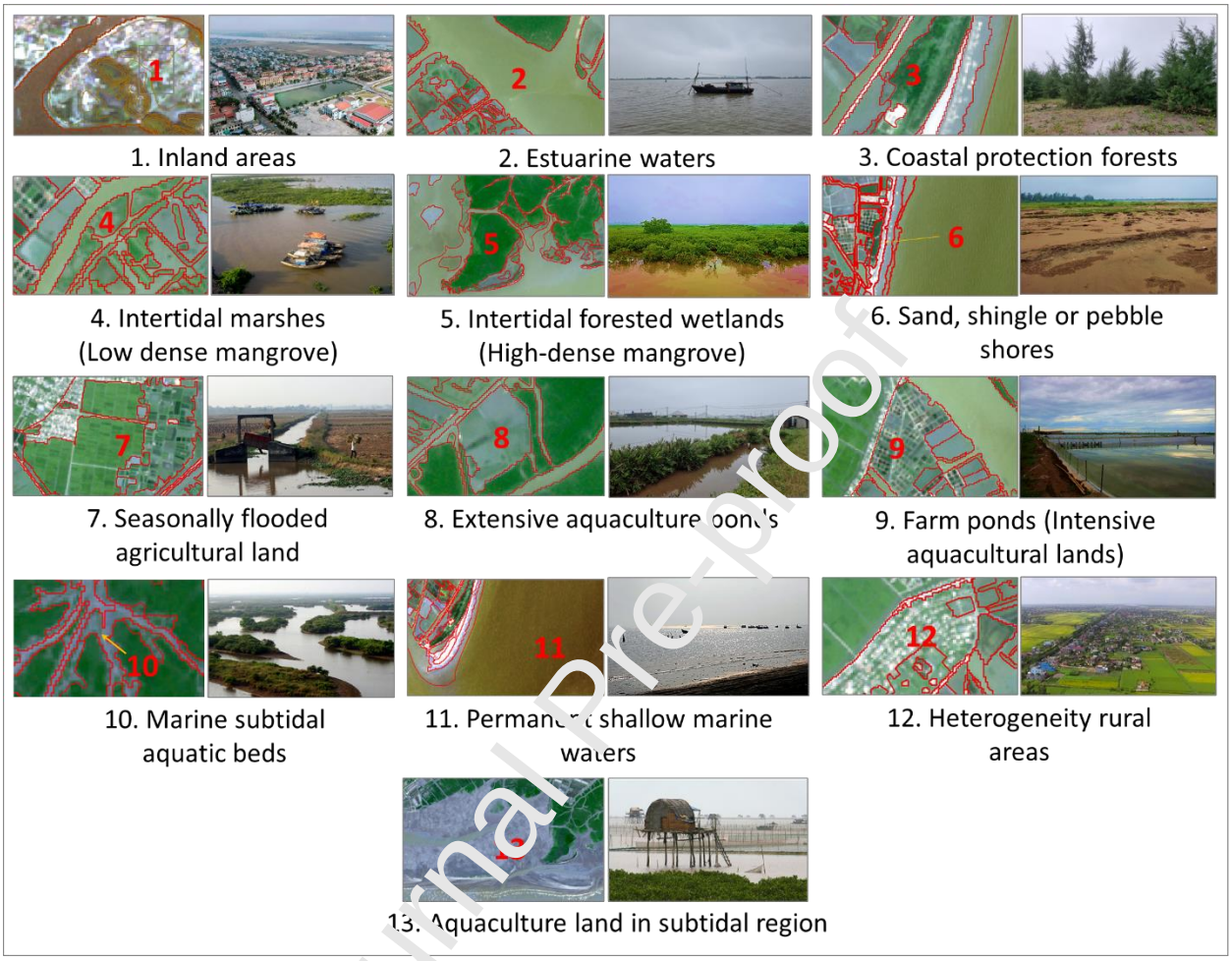

Figure 4. Field data collection in Ba Lat (November 2021) and Tien Yen (March 2020) estuaries.

The main aim of three field survey was to identify ground control points (GCPs) for estuarine ecosystem types. To begin, the authors used the SAGA 7.6.3 software to classify unsupported three Sentinel-2 images into polygons. Many tiny patches with the same tones, structures, and forms were given distinct object classes. Thus, visual interpretation and field interpretation samples utilizing standard GCPs helped to minimize automated image partition error. 30 GCPs were selected randomly for each estuarine ecosystem category. The total number of standard

plots in the research regions was 390 (30 GCPs x 13 types). On the fields, the authors created 50m-radius circular plots to check the GCPs. After identifying GCPs from the source images, field surveys are used to evaluate and assess their accuracy. Because the automated segmentation was made before the field operations, the accuracy is less than 50% compared to GCPs. In the post-processing step, the input samples for training models are 4,426 regions split into 13 groups. This data was converted into a mask for the BiSeNet development in the subsequent steps.

2.4. Alternatives to BiSeNet model development

2.4.1. Performance checking methods

As the training and validation datasets for 13 estuarine ecosystems are not balanced, this research utilized two indices to check the model's performance: total accuracy and loss function values. This loss function analyzes the loss globally rather than microscopically. Other cross-entropy loss is less effective for image-level prediction (Li et al., 2019). The focal loss function (FL) is thus computed as formula 7 to quantify loss between original Sentinel-2 image (S) and the corresponding ground true points:

$$FL = -\frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N} G_{mn}\alpha(1 - S_{mn})^{\gamma} ln(S_{mn}) \quad (6)$$

where N is set at 13 as the number of the estuarine ecosystems, M is the number of observed samples in whole input data, α and γ are weighting factors fluctuating in [0 to 5].

Total accuracy (ACC) is an additional accuracy metric that will be calculated as follows:

$$ACC = \frac{2TP}{2TP+FP+FN} \quad (7)$$

TP, FP, FN, and TN, respectively, are true positive, false positive, false negative, and true negative between prediction and ground truth. The trained model with the lowest values of the

loss function and the highest values of the ACC will be the best model for identifying new estuarine areas.

2.4.2. Optimizer techniques

Neural networks may be built using different optimizer techniques to minimize their cost functions (e.g., loss of data information, training time and uncertainty) using optimization methods (Dang et al., 2020b, 2020a). During the optimization cycles, errors in the trained models (or the loss function) were computed continuously. The weights of all the BiSeNet models are updated after each epoch to minimize the magnitude of the weight loss for the following assessment. The BiSeNet development used seven optimization methods consecutively as shown in Table 1. This table summarizes the optimization methods discussed previously. Therefore, the optimizer method selection becomes the best way to determine a model with the best accuracy and least loss function value.

Table 1. The seven optimization algorithms to train parameters of the BiSeNet architecture for the estuary ecosystem classification (adapted from Gulli and Pal, 2017; Iglovikov et al., 2017; Wang et al., 2017; Alom et al., 2019; Falbel et al., 2019).

Formula	Model	Optimizer method	Algorithms
8	BiSe-Adam	Adam	$p_{t+1} = p_t - \frac{\eta}{\sqrt{\hat{v}_t}+\epsilon}\hat{m}_t$
9	BiSe-Adamax	Adamax	$p_{t+1} = p_t - \frac{\eta}{u_t}\hat{m}_t$
10	BiSe-Adagrad	Adagrad	$p_{t+1} = p_t - \frac{\eta}{\sqrt{G_t}+\epsilon} g_t$
11	BiSe-Nadam	Nadam	$p_{t+1} = p_t - \frac{\eta}{\sqrt{\hat{v}_t}+\epsilon}(\beta_1\hat{m}_t + \frac{(1-\beta_1)g_t}{1-\beta_1^t})$
12	BiSe-RMS	RMS probability	$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$

			and $p_{t+1} = p_t - \frac{\eta}{\sqrt{E[g]^2{}_t + \epsilon}} g_t$
13	BiSe-SGD	SGD	$p_{t+1} = p_t - \eta_t . \nabla_p Q\left(p_t; x^{(i)}; y^{(i)}\right)$

where p is parameter value; η is the learning rates; t is time step; ϵ = 10-8; g_t is the gradient; E[g] — moving average of squared gradients; m, v are estimates of first and second moments; u_t - the max operation; β - moving average parameter (good default value — 0.9); η – step size.

2.5. Model comparison

Integrating the four types of input sub-image sizes and five optimizer options, 20 BiSeNet options were trained. The accuracy and loss function values of the best four BiSeNet models are compared to those of two benchmark models (RF and SVM) after the training process. Overall accuracy (ACC) and kappa coefficients are selected as evaluation values. The RF and SVM models were produced and their details are documented below.

2.5.1. Random Forest (RF)

RF is a non-parametric machine learning method proposed by Breiman (2001) and Berhane et al. (2018). With reference to RF development, after analyzing data in different branches of decision trees, the last step is determined by majority vote (Chen et al., 2019). The training dataset was split into two parts: a bootstrap sample (consists of 80% of database) for each decision tree; and an out-of-bag sample (consists of 20% of database) for each RF model. The RF randomly selects a subset of variables and groups the training data. So the forest's decision trees were modified to prevent overfitting (Breiman, 2001). The number of trees, variables, and training data may be changed. The last RF with the highest total accuracy will be chosen and utilized for new

prediction and categorization. The research examined 10, 100, 500, and 1000 trees. Finally, 100 trees delivered the best accuracy in this study.

2.5.2. Support Vector Machine (SVM)

Various models have used SVM as a supervised machine learning method for image classification in real-time (Karatzoglou et al., 2006; Wang et al., 2019). For classification purposes, SVM models produce a hyperplane or plane with large gaps between categories. As an example, Hassan and Sadek (2017) proposed a SVM model for image classification with a two-dimensional hyperplane so that the data could be categorized. To improve SVM models, two parameters were sought and optimized: "gamma" as a kernel coefficient and "C" as a error penalty parameter. These two parameters were examined to get the greatest ACC and kappa. Alternative kernel functions such as linear, polynomial, sigmoid, and radial basis functions influence the SVM precision (Karatzoglou et al., 2006). Even a minor error may cause over-fitting issues in training a SVM. The optimum "gamma" value was chosen at 0.25 while the optimal "C" value was chosen at 100.

2.6. Using BiSeNet models to classify estuary ecosystems

After selecting the best BiSeNet model, the predictive models must forecast the distribution of estuarine ecosystems and their changes using 20 Sentinel-2 images in Xuan Thuy National Park from 2017 to 2021. To identify the final type values for each data point, the trained BiSeNet model accesses the trained parameters in the BiSeNet's layers to translate new image data into the approperiate ecosystem types. In numerical classes the names of the estuary ecosystems will be assigned to the FC layer. The estuary ecosystem classification and their changes in the regions are addressed in section 4 and compared with predictions made in other previous studies.

3. Results

3.1. BiSeNet model performance

Based on the input mask generated from two fieldworks (Figure 5), 20 BiSeNet models were trained and compared with each other based on the training and testing accuracies and loss function values. With the same BiSeNet structure, the input data was separated into sub-images with different sizes. According to the training and testing accuracy, the input sub-images with larger sizes can provide higher accuracy values. While the training accuracy rose slightly, the testing accuracy increased significantly when the training size was upgraded from 64x64 to 512x512 (Figure 6). In contrast, the lost function values diminished linearly to nearly zero. It shows that the size of the input sub-images should be used at 512x512. However, the time cost to train a BiSeNet model with input 512x512 sub-images requires more than 300 minutes with a computer Intel(R) Core (TM) i7-9750H CPU 2.60GHz, 16.0 GB RAM, and NVIDIA GeForce RTX 2060. Especially, the BiSe-Nadam models took more than 650 minutes for the training process.

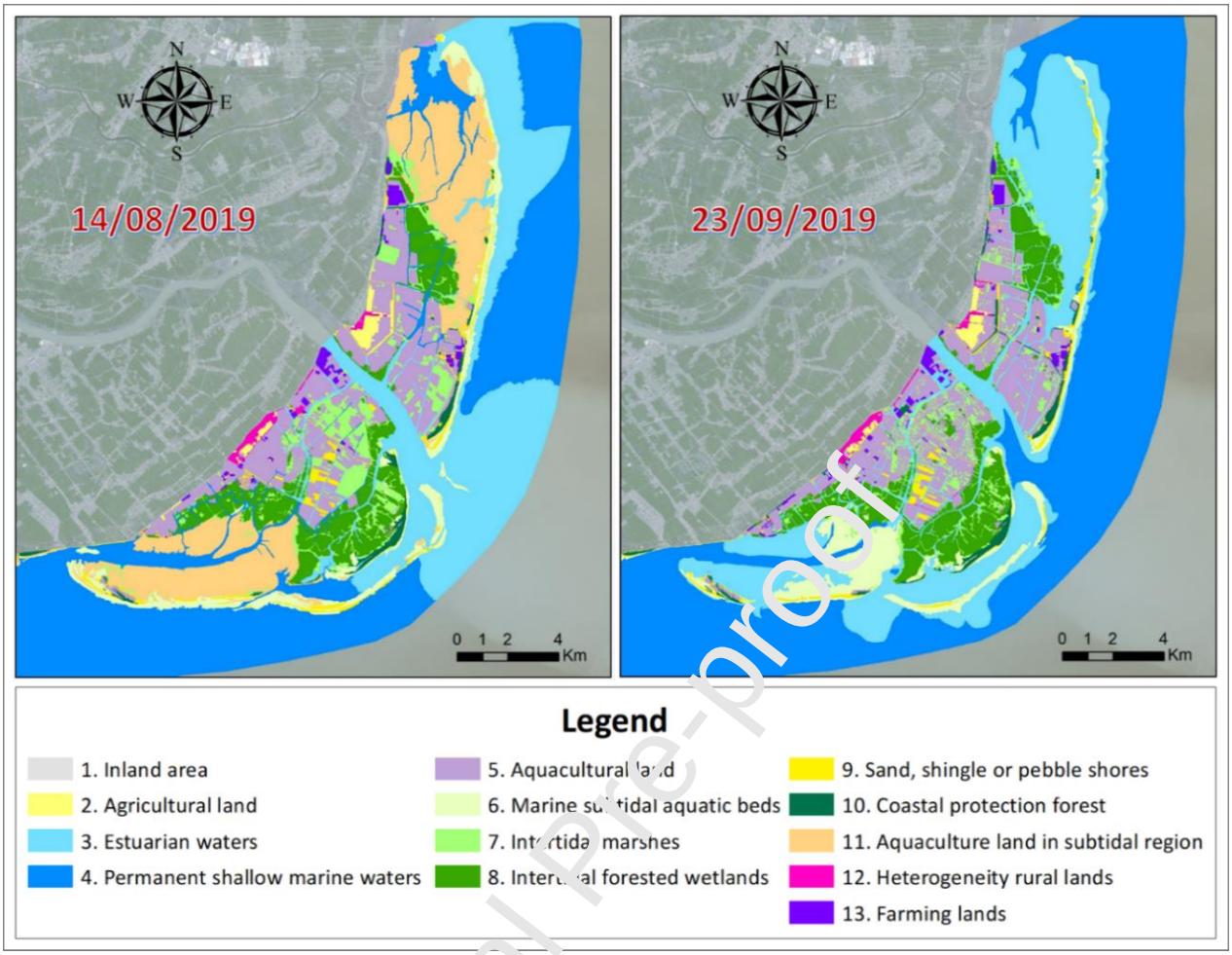

Figure 5. The input mask in the Ba Lat estuary generated based on visual interpretation, combined with field interpretation samples using standard GCPs.

Based on 390 GCPs collected from fieldworks and high-resolution Google Earth images, the input masks for the BiSeNet training process were generated (Figure 5). Compared to the Tien Yen estuary, the estuarine ecosystems in the Ba Lat region located in a flatter floodplain are affected significantly by river and ocean flows. People in the Ba Lat estuary area develop aquatic breeds in four ecosystems: (1) intensive farming; (2) extensive aquacultural ponds; (3) extensive aquaculture in mangroves; and (4) flooded intertidal zones (around watchtowers). The intertidal marshes containing a large area of low-dense mangroves were also used for aquacultural purposes. It is easy to observe the aquacultural land in subtidal regions in the satellite image

taken in August, 2019. However, three storm events and tropical depressions in September, 2019 clearly changed the land cover of aquacultural lands and narrowed the estuarine water area.

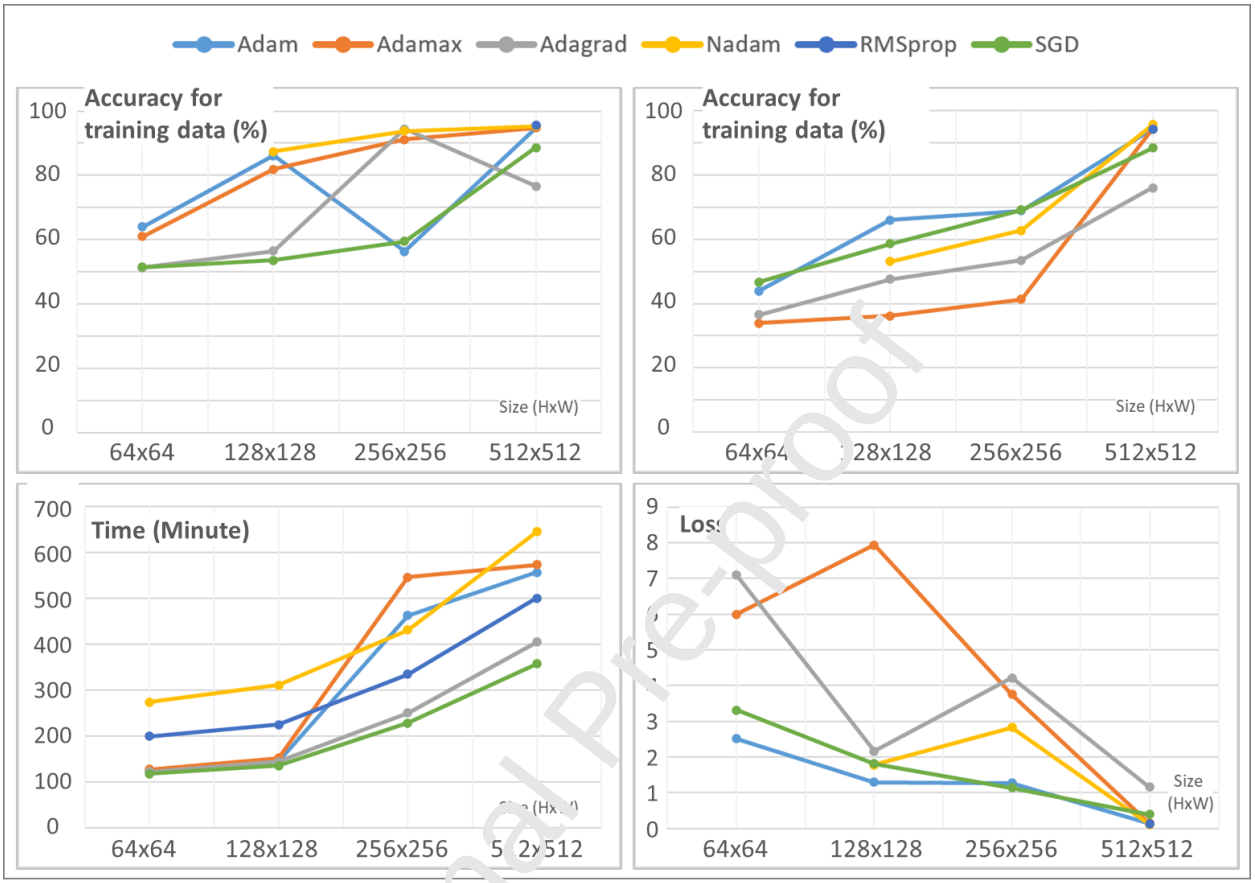

Figure 6. Accuracy, loss function value, time cost information of 20 trained BiSeNet models.

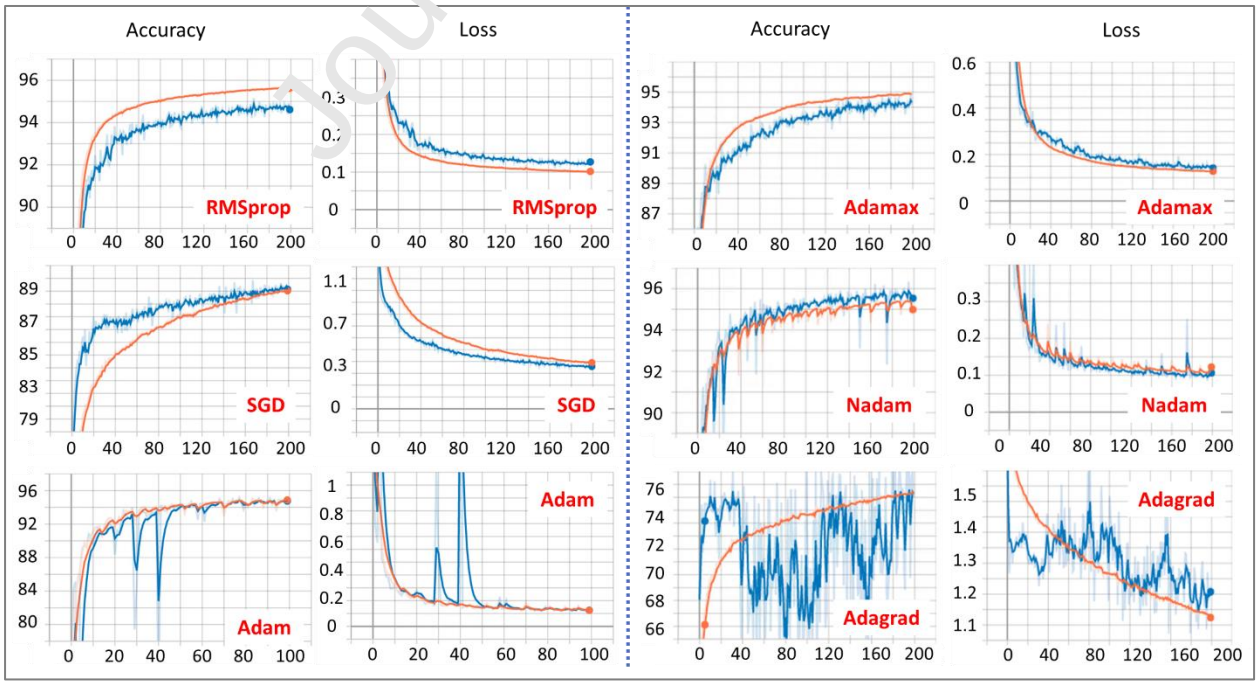

Figure 7. Fluctuation of accuracy and loss function values after 200 epochs of BiSeNet models with the sub-image size of 512x512x4 using six optimizer functions.

Table 2. Accuracy of four BiSeNet models and two benchmark models in classifying 13 ecosystems in the estuarine area.

No.	Estuarian ecosystem types	No. Sample	Aggregated class accuracy of models					
			BiSeNet				SVM	RF
			SGD	RMS	Nadam	Adam		
1	Inland area	100	100.0	100.0	100.0	100.0	91.0	95.9
2	Agricultural land	100	89.2	89.2	92.2	73.5	95.1	64.7
3	Estuarian waters	100	92.3	92.3	90.8	87.7	86.2	73.8
4	Shallow marine waters	100	84.3	84.3	85.3	91.2	2.0	68.6
5	Aquacultural land	100	86.3	84.2	88.5	89.2	87.1	54.7
6	Marine subtidal aquatic beds	100	68.8	68.8	67.5	68.8	85.0	37.5
7	Intertidal marshes	100	82.4	82.4	81.4	84.3	55.9	45.1
8	Intertidal forested wetlands	100	95.9	95.9	93.8	93.8	94.8	61.9
9	Sand, shingle or pebble shores	100	85.7	85.7	83.1	87.0	40.3	0.0
10	Coastal protection forest	100	83.3	83.3	80.3	81.8	74.2	68.2
11	Aquaculture land in subtidal region	100	94.0	94.0	92.9	90.2	0.0	77.2
12	Heterogeneity rural lands	100	82.5	82.5	81.6	74.8	42.7	73.8
13	Farming lands	100	47.0	47.0	45.8	61.4	50.6	66.3
	Total OA (%)		**85.2**	**84.9**	**84.7**	**84.3**	**57.7**	**62.2**
	Cohen's kappa		**83.8**	**83.5**	**83.3**	**82.9**	**53.8**	**59.0**

In twenty trained BiSeNet models only 16 have a higher than 50% accuracy. Four models with input sub-image sizes of 64x64 and 128x128 were not trained successfully. Only four BiSeNet models had a testing accuracy higher than 90% include BiSe-Adam (94%), BiSe-Adamax (94%), BiSe-RMS (94%) and BiSe-Nadam (96%). The loss function values of these four models were recorded at 0.12 in average (Figure 7). Compared with the GCPs, the highest overall accuracy achieved in the BiSe-Nadam model with 85% with Kappa index of 84%. Three other BiSeNet models achieved overall accuracy and Kappa indices higher than 84% and 82%, respectively.

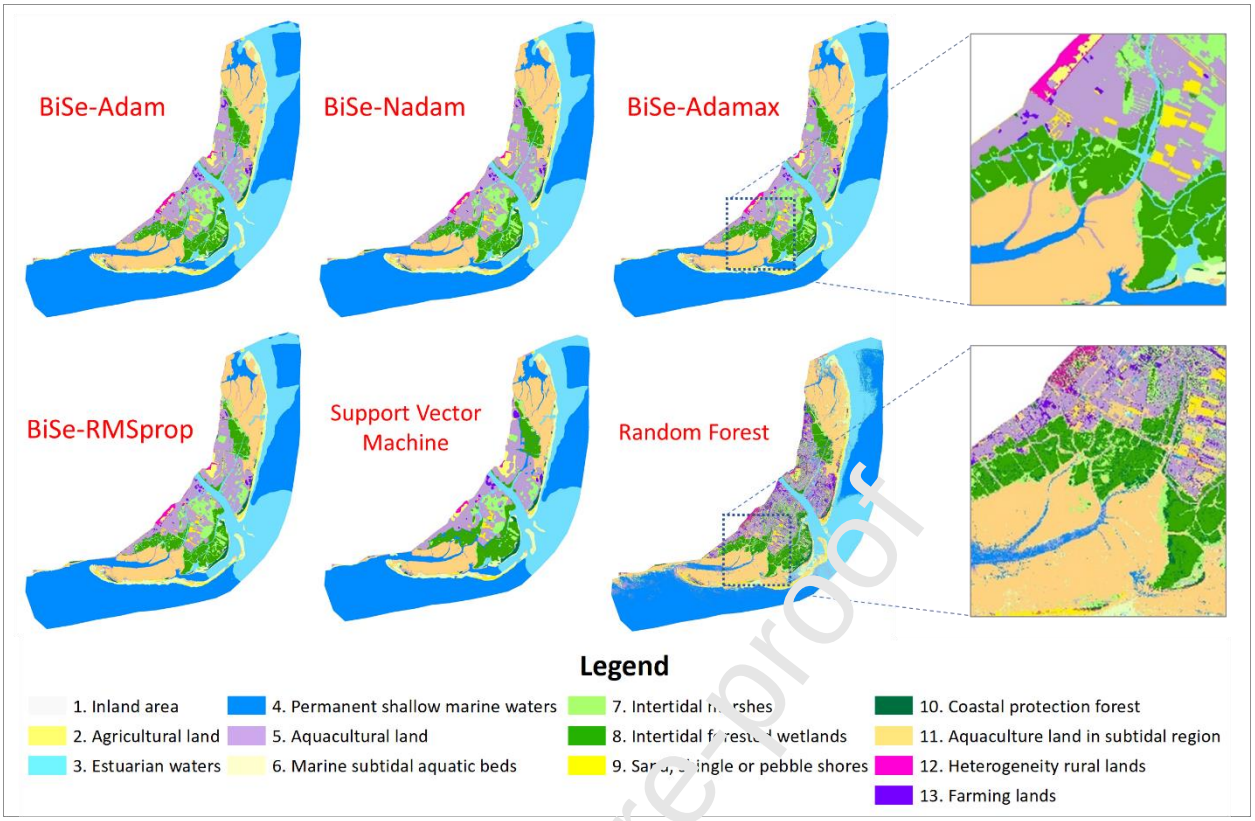

Figure 8. Estuarine ecosystem classification from the Sentinel-2 images and two benchmark models.

According to the prediction from four BiSeNet models for the Sentinel-2 images taken on 14/08/2019 (Figure 8), the accuracy of each estuarine ecosystem type was assessed. Accordingly, almost all ecosystems can be identified with an accuracy better than 80%. Especially, the BiSeNet models can identify exactly the estuarine waters, intertidal forested wetlands and aquacultural land in a subtidal region with an accuracy of higher than 90% (Table 2). In contrast, the accuracy of the marine subtidal aquatic beds and the farming lands in the outcomes ranges only from 50% to 70%. Comparing to two benchmark models, four BiSeNet models proved their performance in the ecosystem classification. The accuracy and Kappa indices of four models are higher than those of two benchmark models. Especially, three types of aquacultural lands were mixed with each other in the outcomes of the SVM and RF models. The sandy areas and intertidal marshes could not be identified well in both models.

3.2. Wetland cover changes in Xuan Thuy National Park

Twelve interpretation results were sourced from twenty new Sentinel-2 images taken during five years based on the BiSe-Nadam model (Figure 9). Apart from the shallow marine waters and inland area, the area of the other eleven wetland ecosystems changed according to a natural rule. In six images taken in rainy seasons (Figure 10), the wetland ecosystems area expands to more than 9,000 ha due to the increase in the estuarial material from Red River. In particular, the estuarine waters enlarged to shallow-sea water and marine subtidal aquatic bed areas during the rainy season (April to November). It leads to suitable conditions for aquacultural activities on canals and rivers. In the center of the National Park, the extensive aquacultural land takes up a large area. Meanwhile, the intensive farming land was found in the buffer zone of the National Park without any evident arial change. In the dry season (December to March), the area of wetland ecosystems shrinks to approximately 8,500 ha. However, the area comprising the mangroves, agriculture, protection forest and rural regions in general has not changed after five years.

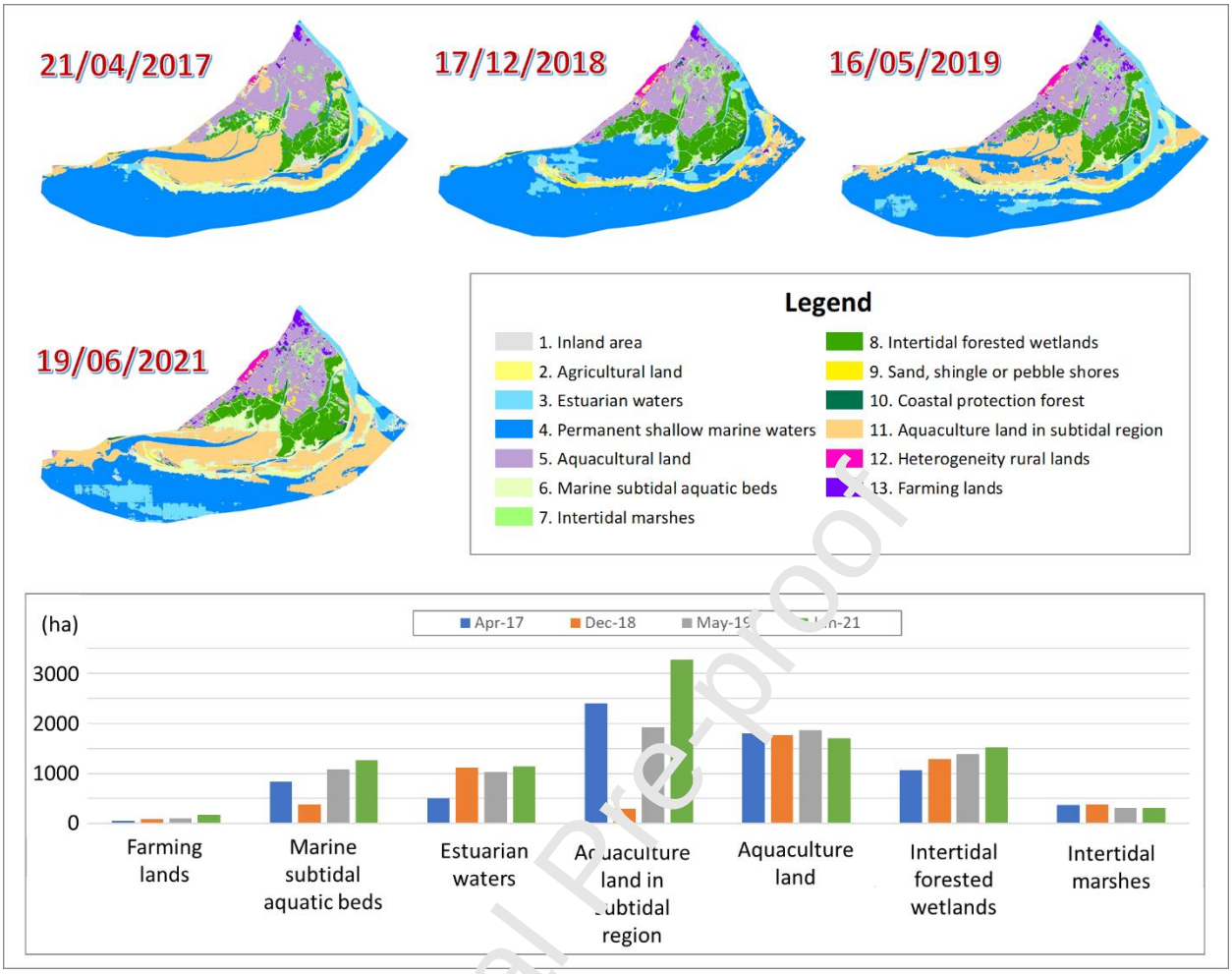

Figure 9. Wetland cover distribution and their changes in Xuan Thuy National Park, Vietnam from 2017 to 2021.

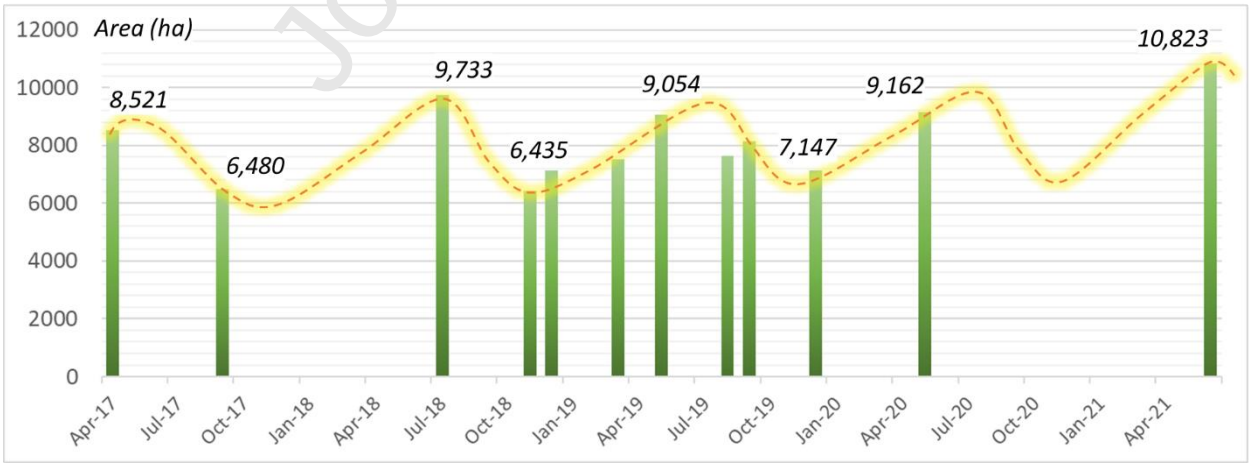

Figure 10. Arial variation of wetland ecosystem from 2017 to 2021 in Xuan Thuy National Park (excluding deep water ecosystems).

According to Figure 10, the area of intensive aquacultural land tripled after five years but it only took about 1% of the wetland area in the National Park. The subtidal area, shallow-sea, low dense mangrove and subtidal area mixed with aquaculture were strongly affected by river flow and tidal in two seasons. With the forest planting strategies implemented by the managers and local households, the area's high dense mangrove has increased by about 150% since 2017.

4. Discussion

4.1. Network comparison

The use of BiSenet model limits the loss of information and format in the received characteristics and provides outcomes with high speed at the same time. These benefits are often traded off with each other in traditional neural-network models. The "semantic segmentation" technique necessitated a fast rate of data inference/analysis in order to give immediate and intuitive feedback about the outcome representation. In this study, this technique was employed effectively in the trained BiSeNet models for estuarine ecosystem classification in contrast to "instance segmentation," which divides each comparable item into several layers (Garg et al., 2019; Diakogiannis et al., 2020b). As a consequence, the model can preserve the original image's quality of medium-resolution remote sensing data as the Sentinel-2 image. It retains adequate spatial information through convolutional layers integrated into the SP path acquiring the RF field using pyramidal spatial pooling layers with a high kernel value (Breiman, 2001; Lary et al., 2016). Models such as CNN or U-Net are unlikely to satisfy both of these requirements at the same time (Abubakar and Boukari, 2018; Kim and Jeong, 2019). Based on the results, a small-sized input image loses most of the spatial information included in the original image, leading to much less total accuracy (Figure 6).

In all trained BiSeNet models, the CP and SP paths were set up at the beginning of the main framework. They were combined with the FFM to balance processing speed and outcome accuracy of two paths. Therefore, it significantly solves the information loss issue, compared with traditional real-time algorithms. The outputs of these two pathways are then merged to provide the deciphered result's final forecast. Through this computation, the output results may take 30 seconds to a minute to achieve while maintaining a high level of accuracy. Commonly, lightweight models may cause channel truncation to distort spatial information. Accordingly, the proposed model utilizes a lightweight path as done in the CP path to compress the image and accelerate the computing process. As a result, this procedure is not computationally demanding, even with big satellite data. Therefore, the proposed model has potential as a tool for monitoring changes in the estuarine ecosystems based on high resolution satellite images. Especially with unmanned aerial vehicles, the effects of alluvial materials from rivers to the estuarine ecosystems can be monitored quickly monthly if managers are equipped with this tool.

4.2. Uncertainties in sampling

Firstly, it is difficult to separate the "farm ponds" and the "aquaculture ponds" in the input Sentinel-2 images, despite the fact these estuarine ecosystem types are easily accessible for the purposes of fieldwork. The aquaculture ponds have been utilized for intense farming without modern technology, whereas farm ponds are frequently planned for high-technology shrimp farming. Aquaculture ponds are usually bigger than farm ponds, while farm ponds are distributed uniformly across a wide region. Farming ponds cover an area less than 9 ha, and can be recognized with a small plot layout, dark green hue, and bordered by a thin bank. Due to the specific characteristics, two ecosystems could not be identified by a pixel-based analysis. In contrast, the object-based analysis integrated in the BiSeNet network effectively categorized the differentiation between them (Table 2). There are other deep learning models using U-Net

structures, which have been effectively used for land-cover classification. These models need to be developed for the estuarine wetland ecosystem classification to compare with the trained BiSeNet models. It can provide more information to select a classification model for "farm ponds" and the "aquaculture ponds" ecosystems in the Sentinel-2 images in real-time.

Secondly, one estuarine ecosystem type could be interpreted in different colors in a spectral composite image. For example, the color of the aquaculture ecosystems varies from dark green to black for deep ponds, while being light green for algal cover or shallow ponds. The same is the case for sandy dunes, which are dark gray when inundated, and light in color at low tide. In particular, the distribution of sandy shores does not retain a fixed shape. To minimize the uncertainties arising from this issue, the author had to check the location of these objects and check the land cover on the field works and Google Earth images, and correct the sample information for the training model. Due to these similarities between the estuarine ecosystems causing errors, this led to the outcome models being less accurate.

Thirdly, the estuarine ecosystems and especially those in Xuan Thuy National Park, are affected significantly by tidal, river and ocean currents between two seasons of the year. Some fixed ecosystems such as sandy lands, intertidal marshes, and subtidal aquatic beds are sometimes replaced by estuarine waters and shallow marine waters, once the alluvial materials flow from inland. It is easy to separate these in the middle of two seasons but becomes more difficult in the transition time between two seasons (November, December, March and April in Xuan Thuy National Park). To eliminate this problem, all ecosystems should be observed more frequently so that the location of stable ecosystems can be fixed. The multi-temporal satellite images could be used to provide more samples for ecosystems that have rapid seasonal changes. The deep learning models need to function so as the information of ecosystems from old images can be updated and identify ecological succession rules in the estuarine ecosystems.

5. Conclusion

The completed BiSe-Nadam model based on semantic-segmentation operation can be used to interpret 13 ecosystems in estuarine regions from Sentinel-2 images in less than one minute. The implementation of the BiSeNet model may help coastal managers with monitoring dynamic estuarine ecosystems monthly, instead of every five years as ecologists have historically done. The analysis of spatial and contextual information in two stages helps the network to well record the spectral, shape and structure of each ecosystem type. The Nadam optimizer maximized the model's accuracy to more than 96% and minimized the loss of data to less than 0.12 times higher than benchmark models. Additionally, the trained BiSe-Nadam model successfully monitored the dynamic of the estuarine ecosystems in Xuan Thuy National Park in Vietnam. Accordingly, the area of these has fluctuated in terms of a seasonal rule. The estuary and shallow marine waters experience considerable variation due to the influence of rivers, while the mangroves and protection forests have been maintained since 2017. It will become more interesting for further studies to test this model in monitoring estuarine ecosystems based on the use of higher-resolution satellite images.

Acknowledgement

The paper was conducted under the science and technology project of the Ministry of Natural Resources and Environment: "Research and propose solutions to maintain ecological characteristics and improve community livelihoods at some Ramsar sites in Vietnam", code TNMT.2020.562.08. The authors would like to express special thanks to the support of Centre for Technology in Water and Wastewater, University of Technology, Sydney.

References

Abubakar, F.A., Boukari, S., 2018. A Convolutional Neural Network with K-Neareast Neighbor for Image Classification. Ijarcce 7, 1–7.

Albawi, S., Mohammed, T.A., Al-Zawi, S., 2018. Understanding of a convolutional neural network. Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017 2018-Janua, 1–6.

Alom, M.Z., Taha, T.M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M.S., Hasan, M., Van Essen, B.C., Awwal, A.A.S., Asari, V.K., 2019. A state-of-the-art survey on deep learning theory and architectures. Electronics (Switzerland) 8, 1–67.

Banks, S., White, L., Behnamian, A., Chen, Z., Montpetit, B., Brisco, B., Pasher, J., Duffe, J., 2019. Wetland Classification with Multi-Angle/Temporal SAR Using Random Forests. Remote Sensing 11, 670.

Berhane, T.M., Lane, C.R., Wu, Q., Autrey, B.C., Anenkhonov, O.A., Chepinoga, V. V., Liu, H., 2018. Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. Remote Sensing 10.

Bournaris T, Papathanasiou J, Manos B, Kazakis N, Voudouris K. 2015. Support of irrigation water use and eco-friendly decision process in agricultural production planning. Operational Research, 15(2): 289–306. https://doi.org/10.1007/s12351-015-0178-9

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Chen, X., Wang, T., Liu, S., Peng, F., Tsunekawa, A., Kang, W., Guo, Z., Feng, K., 2019. A New Application of Random Forest Algorithm to Estimate Coverage of Moss-Dominated Biological. Remote Sens. 11, 18.

Comittee, F.G.D., 2012. Coastal and marine ecological classification standard, June 2012. National Oceanic and Atmospheric Administration 343.

Dai, M., Leng, X., Xiong, B., Ji, K., 2020. Sea-land segmentation method for SAR images based on improved BiSeNet. Journal of Radars 9, 886–897. https://doi.org/10.12000/JR20089

Dang, K.B., Dang, V.B., Bui, Q.T., Nguyen, V.V., Pham, T.P.N., Ngo, V.L., 2020a. A Convolutional Neural Network for Coastal Classification Based on ALOS and NOAA Satellite Data. IEEE Access 8, 11824–11839.

Dang, K.B., Nguyen, M.H., Nguyen, D.A., Phan, T.T.H., Giang, T.L., Pham, H.H., Nguyen, T.N., Van Tran, T.T., Bui, D.T., 2020b. Coastal wetland classification with deep u-net convolutional networks and sentinel-2 imagery: A case study at the tien yen estuary of vietnam. Remote Sensing 12, 1–26.

de Groot D., Brander L., Finlayson M. (2016) Wetland Ecosystem Services. In: Finlayson C. et al. (eds) The Wetland Book. Springer, Dordrecht.

DeLancey, E.R., Simms, J.F., Mahdianpari, M., Brisco, B., Mahoney, C., Kariyeva, J., 2020. Comparing deep learning and shallow learning for large-scalewetland classification in Alberta, Canada. Remote Sensing 12.

Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020a. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing 162, 94–114.

Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020b. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing 162, 94–114.

Duc, D.M., Nhuan, M.T., Ngoi, C. Van, 2012. An analysis of coastal erosion in the tropical rapid accretion delta of the Red River, Vietnam. Journal of Asian Earth Sciences 43, 98–109.

Feng, Q., Yang, J., Zhu, D., Liu, J., Guo, H., Bayartungalag, B., Li, B., 2019. Integrating multitemporal Sentinel-1/2 data for coastal land cover classification using a multibranch convolutional neural network: A case of the Yellow River Delta. Remote Sensing 11.

Garg, L., Shukla, P., Singh, S.K., Bajpai, V., Yadav, U., 2019. Land use land cover classification from satellite imagery using mUnet: A modified UNET architecture. VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications 4, 359–365.

Ghuffar, S., 2018. DEM generation from multi satellite Planetscope imagery. Remote Sensing 10, 1–22.

Giang, T.L., Dang, K.B., Toan Le, Q., Nguyen, V.G., Tong, S.S., Pham, V.-M., 2020. U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV Imagery. IEEE Access 8, 186257–186273.

Guo, M., Li, J., Sheng, C., Xu, J., Wu, L., 2017. A review of wetland remote sensing. Sensors (Switzerland) 17, 1–36.

Guo, Q., Jin, S., Li, M., Yang, Q., Xu, K., Ju, Y., Zhang, J., Xuan, J., Liu, J., Su, Y., Xu, Q., Liu, Y., 2020. Application of deep learning in ecological resource research: Theories, methods, and challenges. Science China Earth Sciences 2172.

Hao, S., Zhou, Y., Zhang, Y., Guo, Y., 2020. Contextual Attention Refinement Network for Real-Time Semantic Segmentation. IEEE Access 8, 55230–55240.

Hassan, S.M., Sadek, M.F., 2017. Geological mapping and spectral based classification of

basement rocks using remote sensing data analysis: The Korbiai-Gerf nappe complex, South Eastern Desert, Egypt. Journal of African Earth Sciences 134, 404–418.

Heywood, V., 2011. The hazardous future of island floras. In: Bramwell, D., Caujapé-Castells, J. (Eds.), Island Ecosystems – Priorities for Conservation. Cambridge University Press, 22.

Hoang, V.T., Le, D.D., 2006. Wetland classification system in Vietnam, CRES, Viet. ed. Vietnam Environment Administration, Hanoi.

Hu, X., Ban, Y., Nascetti, A., 2021. Uni-temporal multispectral imagery for burned area mapping with deep learning. Remote Sensing 13.

Iglovikov, V., Mushinskiy, S., Osin, V., 2017. Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition.

Karatzoglou, A., Meyer, D., Hornik, K., 2006. Support Vector Algorithm in R. Journal of Statistical Software 15, 1–28.

Kapetas L, Kazakis N, Voudouris K, McNicholl D. 2019. Water allocation and governance in multi-stakeholder environments: Insight from Axios Delta, Greece. Science of the Total Environment, 695: 133831. https://doi.org/10.1016/j.scitotenv.2019.133831

Kim, H., Jeong, Y.S., 2019. Sentiment classification using Convolutional Neural Networks. Applied Sciences (Switzerland) 9, 1–14.

Krestenitis, M., Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., Kompatsiaris, I., 2019. Oil spill identification from satellite images using deep neural networks. Remote Sensing 11, 1–22.

Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. Geoscience Frontiers 7, 3–10.

Lee, T.M., Yeh, H.C., 2009. Applying remote sensing techniques to monitor shifting wetland vegetation: A case study of Danshui River estuary mangrove communities, Taiwan. Ecological Engineering 35, 487–496.

Li, B., Liu, Y., Wang, X., 2019. Gradient Harmonized Single-Stage Detector. Proceedings of the AAAI Conference on Artificial Intelligence 33, 8577–8584.

Luisetti, T., Turner, R.K., Jickells, T., Andrews, J., Elliott, M., Schaafsma, M., Beaumont, N., Malcolm, S., Burdon, D., Adams, C., Watts, W., 2014. Coastal zone ecosystem services: From science to values and decision making; a case study. Science of the Total Environment 493, 682–693.

Neupane, B., Horanont, T., Aryal, J., 2021. Deep learning-based semantic segmentation of urban features in satellite images: A review and meta-analysis. Remote Sensing 13, 1–41. https://doi.org/10.3390/rs13040808

Pinto, R., Patrício, J., Neto, J.M., Salas, F., Marques, J.C., 2010. Assessing estuarine quality under the ecosystem services scope: Ecological and socioeconomic aspects. Ecological Complexity 7, 389–402

Pye, K., Blott, S.J., 2014. The geomorphology of UK estuaries: The role of geological controls, antecedent conditions and human activities. Estuarine, Coastal and Shelf Science 150, 196–214.

Quoc Vo, T., Kuenzer, C., Oppelt, N., 2015. How remote sensing supports mangrove ecosystem service valuation: A case study in Ca Mau province, Vietnam. Ecosystem Services 14, 67–75.

Rakhlin, A., Davydow, A., Nikolenko, S., 2018. Land cover classification from satellite imagery with U-net and lovász-softmax loss. IEEE Computer Society Conference on Computer

Vision and Pattern Recognition Workshops 2018-June, 257–261.

RAMSAR, 2010. Wetlands: a global disappearing act.

Ramsar Convention Secretariat, 2013. The Ramsar Convention Manual , 6th edition. The Ramsar Convention Manual: A Guide to the Convention on Wetlands (Ramsar, Iran, 1971) 109.

Sannigrahi, S., Chakraborti, S., Joshi, P.K., Keesstra, S., Sen, S., Paul, S.K., Kreuter, U., Sutton, P.C., Jha, S., Dang, K.B., 2019. Ecosystem service value assessment of a natural reserve region for strengthening protection and conservation. Journal of Environmental Management 244, 208–227.

Stoian, A., Poulain, V., Inglada, J., Poughon, V., Derksen, D., 2019. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. Remote Sensing 11, 1–26.

Sun, D., Huan, Y., Wang, S., Qiu, Z., Ling, Z., Mao, Z., He, Y., 2019. Remote sensing of spatial and temporal patterns of phytoplankton assemblages in the Bohai Sea, Yellow Sea, and east China sea. Water Res. 157, 119–133.

Swainson, B., 2008. Rivers at Risk Contents: Dams and the future of freshwater ecosystems, WWF canada.

Takagi, H., Thao, N.D., Esteban, M., Mikami, T., Van Cong, L., Ca, V.T., 2015. Coastal disasters in Vietnam, Handbook of Coastal Disaster Mitigation for Engineers and Planners. Elsevier Inc.

Tong, S.S., Deroin, J.P., Pham, T.L., 2020. An optimal waterline approach for studying tidal flat morphological changes using remote sensing data: A case of the northern coast of Vietnam. Estuarine, Coastal and Shelf Science 236, 106613.

Wang, L., Yang, Y., Min, R., Chakradhar, S., 2017. Accelerating Deep Neural Network Training with Inconsistent Stochastic Gradient Descent. Computer Science, Machine Learning,Cornell University v3, 12.

Wang, X., Gao, X., Zhang, Yuanzhi, Fei, X., Chen, Z., Wang, J., Zhang, Yayi, Lu, X., Zhao, H., 2019. Land-cover classification of coastal wetlands using the RF algorithm for Worldview-2 and Landsat 8 images. Remote Sensing 11, 1–22.

Xu, Jie, Lei, S., Bi, S., Li, Y., Lyu, H., Xu, Jiafeng, Xu, X., Mu, M., Miao, S., Zeng, S., Zheng, Z., 2020. Tracking spatio-temporal dynamics of POC sources in eutrophic lakes by remote sensing. Water Res. 168, 115162.

Yang, W., Wang, W., Zhang, X., Sun, S., Liao, Q., 2019. Lightweight Feature Fusion Network for Single Image Super-Resolution. IEEE Signal Processing Letters 26, 538–542.

Yao, X., Yang, H., Wu, Y., Wu, P., Wang, B., Zhou, X., Wang, S., 2019. Land use classification of the deep convolutional neural network method reducing the loss of spatial features. Sensors (Switzerland) 19.

Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N., 2018. BiSeNet: Bilateral segmentation network for real-time semantic segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11217 LNCS, 334–349.

Zhang, P., Ke, Y., Zhang, Z., Wang, M., Li, P., Zhang, S., 2018. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. Sensors (Switzerland) 18.

Zhang, P., Ke, Y., Zhang, Z., Wang, M., Li, P., Zhang, S., 2018. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite

imagery. Sensors (Switzerland) 18.