

Received 9 April 2024, accepted 7 May 2024, date of publication 15 May 2024, date of current version 23 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3401416

## RESEARCH ARTICLE

# Land Cover Classification From RGB and NIR Satellite Images Using Modified U-Net Model


WON-KYUNG BAEK<sup>1,2</sup>, MOUNG-JIN LEE<sup>3</sup>, AND HYUNG-SUP JUNG<sup>1,2,4,5</sup>, (Senior Member, IEEE)<sup>1</sup>Korea Ocean Satellite Center, Korea Institute of Ocean Science and Technology, Busan 49111, South Korea<sup>2</sup>Department of Geoinformatics, University of Seoul, Seoul 02504, South Korea<sup>3</sup>Division for Environmental Planning, Korea Environment Institute, Sejong 30147, South Korea<sup>4</sup>Department of Smart Cities, University of Seoul, Seoul 02504, South Korea<sup>5</sup>Department of Earth Sciences, Southern Methodist University, Dallas, TX 75205, USA

Corresponding author: Hyung-Sup Jung (hsjung@uos.ac.kr)

This research was supported by 1) the Institute of Civil Military Technology Cooperation, the Defense Acquisition Program Administration, and the Ministry of Trade, Industry and Energy of Korea (22-CM-EO-02) and 2) the National Research Foundation of Korea(NRF) grant funded by the Korea Government (MIST) (No. 2023R1A2C1004395). The work of Won-Kyung Baek was supported by the funding of the Korea Institute of Marine Science & Technology (KIMST), backed by the Ministry of Oceans and Fisheries (RS-2023-00254717). The work of MounG-Jin Lee was supported by Development of Optimization Techniques for Reducing Heat Wave Considering Urban Environment (2023-014(R)), conducted by the Korea Environment Institute (KEI).

**ABSTRACT** Multi-spectral satellite imagery has been widely used for land cover classification, because it provides meaningful spectral information for Earth's objects that are difficult to be described by using visible band images. The near-infrared image enables us to classify in the fields of agriculture, forestry, and geology/natural resources. However, the classification performances obtained from deep learning approaches using both red-green-blue (RGB) and near-infrared (NIR) images were not significantly superior to the classification performances using the RGB image, because the spectral information may not be appropriately applied to the deep learning methods. In most deep learning approaches, the convolution operation does not separate the pixel values in the band direction, but rather mixes all the pixel values. This mixing can lead to the loss of information, particularly when dealing with multi-band images (like satellite imagery), as important spectral information might be obscured, affecting the model's accuracy and generalization capability. To overcome the drawback, this study presents an efficient model, which is the separated-input-based U-Net (SiU-Net), via modifying the U-Net model based on the separation of RGB and NIR images. To show the performance improvement of land cover classification from the SiU-Net, the performance of SiU-Net was compared with those of the DeepLabV3+ and U-Net models. We utilized a 2020 satellite-derived land cover dataset, consisting of 300 patches in total. These patches were extracted from Sentinel-2 images, including both RGB and NIR bands, with a resolution of 10 meters, and each patch was sliced into  $512 \times 512$  pixel segments. The entire set of 300 patches was selected without overlap, adhering to a distribution ratio of approximately 64% (192 patches) for training, 16% (48 patches) for validation, and 20% (60 patches) for testing. The final performance evaluations were ultimately conducted using the test data. The F1 score obtained from SiU-Net were about 0.797, and it was superior to about 0.541 from DeepLabV3+ and 0.720 from U-Net. Moreover, the F1 scores of SiU-Net (0.589) was more accurate than DeepLabV3+ (0.051) and U-Net (0.455) in the small training data, and the performance degradation due to data imbalance was reduced in the SiU-Net model. This means that the SiU-Net model may be most suitable when the training data are small and unbalanced.

**INDEX TERMS** Land cover, land cover classification, DeepLabV3+, separated input, SiU-Net, U-Net, red, green, and blue (RGB), near-infrared (NIR).

The associate editor coordinating the review of this manuscript and approving it for publication was Stefania Bonafoni .

## I. INTRODUCTION

Multi-spectral satellite imagery has salient advantages for land cover classification, since it can provide meaningful

spectral information, which is impossible to obtain from visible band data alone [1], [2], [3], [4]. Especially NIR images are widely used for land cover classification as below reasons. 1) it can give us additional information about the land cover, specifically vegetation characteristics such as its types, health, and water contents. 2) the spectral range of NIR (750-1100 nm) is within the solar radiation peak and thus high-quality data of resolution and signal-to-noise (SNR) can be anticipated [5]. Thus, NIR image has been widely used for land cover classification in the fields of agriculture, forestry, and geology/natural resources [1], [2], [3], [6], [7], [8], [9], [10], [11].

Building on the well-established benefits of the NIR band, recent advances in deep learning have led to numerous applications incorporating the NIR band for land cover classification. Mirpulatov et al. introduced a pseudo-labeling approach tailored for multi-spectral satellite data to mitigate the challenge of limited training data [12]. Baek et al. demonstrated performance enhancements through the application of data augmentation techniques to Sentinel-2 RGB and NIR band data [13]. Additionally, studies by Cuypers et al., Yu et al., and Lee et al. have confirmed that combining multi-temporal RGB+NIR data with multimodal data leads to significant improvements in land cover classification performance [14], [15], [16].

However, it is known that the effectiveness of NIR imagery was not appreciable in the deep-learning-based approach [8]. In Defence Science & Technology Laboratory (DSTL) satellite Imagery Feature Detection challenge from the Kaggle, it was not reported that the significant performance improvement by using NIR and short-wave infrared (SWIR) in addition to the visible band [8], [17]. Similarly, Jónsson compared agricultural field classification performances between RGB and multi-spectral (including NIR) band images from unmanned aerial vehicle (UAV) [18]. From Jónsson (2018), although accuracy improvement of 0.14 was maximally archived, in most cases it was only less than 0.05 [18]. Moreover, Gani et al. compared the deep-learning-based object detection performance according to spectral band compositions of the input data such as RGB + thermal bands and RGB + thermal + NIR bands. And they showed degradation of mean average precision when the NIR band is additionally used [19].

This is because the spectral information of input data was not appropriately extracted from the deep learning models. Numerous previous studies have applied encoder-decoder architecture with a single input node and an encoder to RGB and NIR satellite imagery [1], [2], [7], [8], [9], [10], [11], [20]. In the model architecture of a single input node, a bunch of convolution filters from a single encoder cannot separate the pixel values in the band direction, but rather mixes all the pixel values. In other words, independent spectral information of each RGB and NIR data are lost [21], [22], [23].

To overcome this drawback and exploit independent spectral information each bands has, the separated input and dual-encoder architecture can be valid [23]. Hou et al.

has solved the similar problem of us [23]. They proposed dual-encoder architecture for the bi-temporal remote sensing image change detection, which has separated input nodes and extract independent contextual information from them [23]. Like Hou et al. [23], to extract independent spectral-spatial information from multi-spectral data, the encoder should be separated with two- or more- branch architectures according to the correlation of RGB and NIR data. Recently, Yu and Jung [16] developed a model with four encoders for land cover classification from multi-seasonal UAV and Lidar data. However, this study could not showcase the performance improvements of separated input because they utilized spectral indices as inputs, which led to pre-mixed information across RGB and NIR bands, compounded by a lack of sufficient training data [16].

In this study, we showed performance improvement of land cover classification through input separation of RGB and NIR data. For that, we implemented a modified U-Net architecture with two separated input nodes for RGB and NIR respectively, which called separated-input-based U-Net (SiU-Net). To evaluate our model's performance, we conducted comparison tests with DeepLabV3+ and U-Net, which are among the most renowned semantic segmentation models [24], [25]. We utilized 2020 satellite-derived land cover dataset for the training and test data, which contains 4 bands of red, green, blue and NIR, from artificial intelligence (AI) Hub, National Information Society Agency (NIA), Korea [26], [27].

This study offers valuable insights into the effective utilization of the NIR band, in conjunction with RGB, for land cover classification through deep learning methodologies. Additionally, by examining the class proportions, we have identified scenarios where the proposed approach demonstrates optimal effectiveness. We anticipate that our strategy will serve as a guideline for developing land cover classification models utilizing multi-spectral satellite imagery.

## II. METHODOLOGY

In this study, our primary objective is to assess whether our proposed approach surpasses other models in land cover classification on the same dataset. Accordingly, we have organized our methodology into two principal sections: the characteristic architectures of the conventional (DeepLabV3+ and U-Net) and proposed (SiU-Net) deep learning models, followed by the evaluation of their performance using test data. In the section on deep learning models, we primarily discuss the structure of the network. Specifically, for SiU-Net, we describe how the number of separated input nodes and encoders are determined. To comprehensively validate the superiority of our model, we conducted not only numerical comparisons among the models utilized but also a detailed analysis of classification performance from both qualitative and quantitative perspectives. This analysis encompasses an in-depth comparison of class-specific features and the distinctive architectural traits

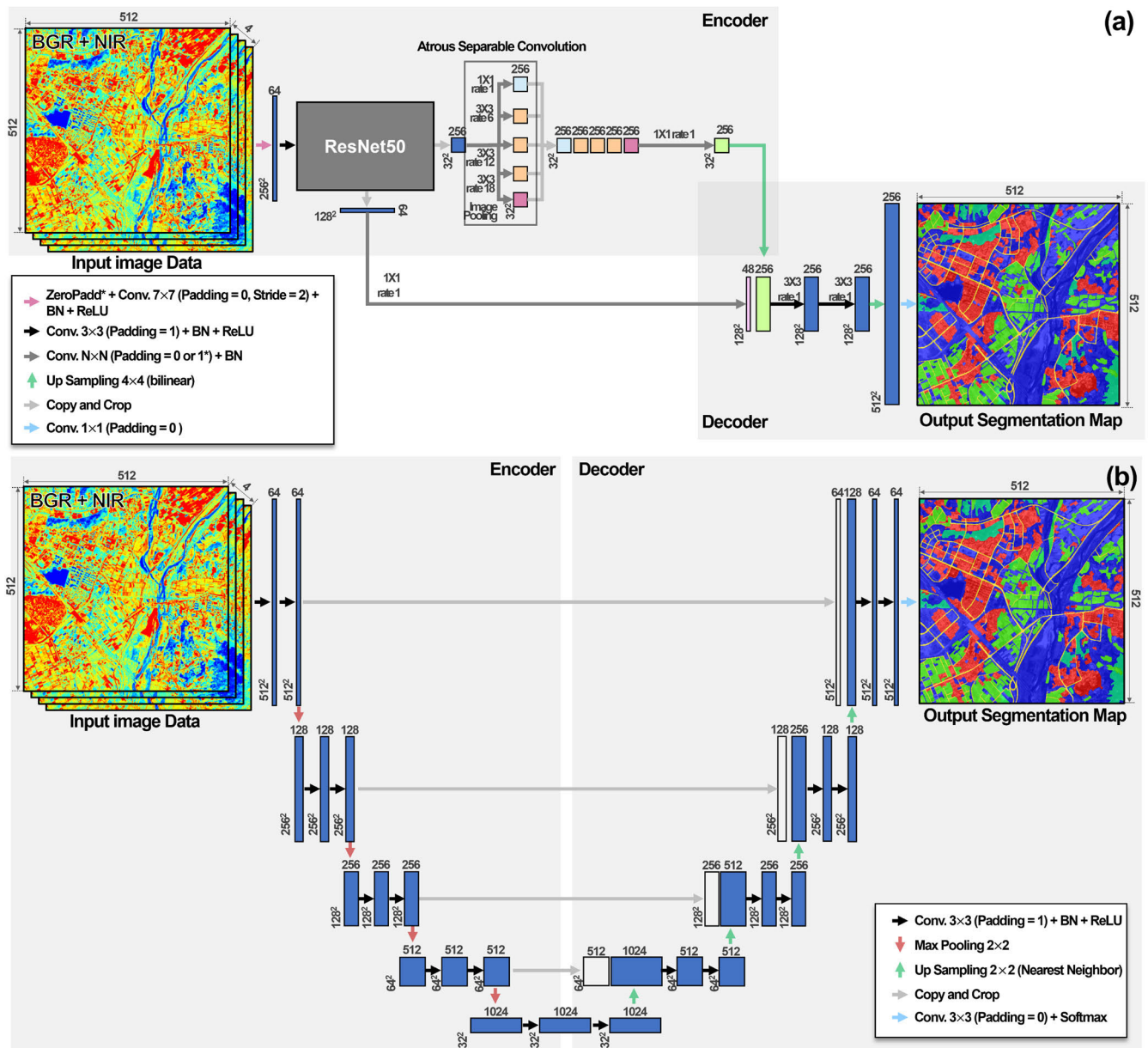


FIGURE 1. The structure of DeepLabV3+ (a) and U-Net (b) used in this study.

of each model. The subsequent portion of this section will detail the architecture of the deep learning models employed in this research and the methods used for evaluating their performance.

### A. DEEP LEARNING MODEL

To validate the effectiveness of our proposed model, well-applied deep learning models are adopted, which are DeepLabV3+ and U-Net [24], [25]. DeepLabV3+ and U-Net are deep learning models, that are constructed in the encoder-decoder structure. Those models have been widely used for semantic segmentation since they have the advantages of localization and recovering boundary

information. Proposed model was designed based on U-Net. Every model used in this study mainly consists of repetitions of  $3 \times 3$  convolutional operations, batch normalization (BN) and a rectified linear unit (ReLU) [28]. The more detailed information about each model is provided in the following sections:

#### 1) DeepLabV3+

DeepLabV3+ is a representative deep learning structure for semantic segmentation, that has been successfully applied in various fields. Figure 1 (a) showed the structure of DeepLabV3+. Image data combined by four channels (blue, green, red and NIR bands) are used as input image data. The

backbone, ResNet-50 generates two bunch of feature maps, which are  $32 \times 32 \times 256$ , and  $128 \times 128 \times 64$  [29]. The spatial resolution of feature maps for the  $32 \times 32 \times 256$  and  $128 \times 128 \times 64$  are 16 times and 4 times lower than that of the original input data respectively.

In the encoder, the atrous separable spatial pyramid pooling (ASPP) module extracts multi-scale contextual information from feature maps of  $32 \times 32 \times 256$  and a  $4 \times 4$  bilinear upsampling operation recovers the size of feature maps as  $128 \times 128 \times 256$ . For the feature maps of  $128 \times 128 \times 64$ , which have low-level information,  $1 \times 1$  convolution was applied to reduce the channels of feature maps ( $128 \times 128 \times 48$ ). By concatenating the feature map of  $128 \times 128 \times 256$  (multi-scale contextual information from the ASPP module) and  $128 \times 128 \times 48$  (low-level information), the edge information is restored. Finally, the full resolution of classification map is generated by applying 2 times convolution and another  $4 \times 4$  bilinear upsampling. However, since the final classification map is recovered from the data whose resolutions are reduced by 4 and 16 times than input data, detailed edge information can be lost.

## 2) U-Net

U-Net, a convolutional neural network with a distinctive encoder-decoder architecture, was initially created for biomedical image segmentation [25]. The key of U-Net is combining feature maps generated in deep layers with those generated in relatively shallow layers by the skip connection. Through this simple skip connection structure, contextual information from deep layers and edge information from shallow layers can be considered together [25]. Eventually, the decoder can generate a detailed segmentation map in spatial.

The U-Net is symmetrical as given in Figure 1(b) [25]. The left part is the encoder for extracting contextual information, and the right part is the decoder for recovering the final classification map. Same as DeepLabV3+, four channels of image data are used as input image data. The basic unit of convolution block is twice  $3 \times 3$  convolution operations, each followed by batch normalization and ReLU. In the encoder, a  $2 \times 2$  max-pooling operation is applied after the convolution block. At each max-pooling operation, the number of convolutional filters is doubled. In the decoder, a concatenate process with skip connection from encoder's feature maps is preceded. And  $2 \times 2$  nearest neighbor upsampling operation is applied instead of max-pooling operations with halving the number of convolutional filters. Totally 4 times of max-pooling and upsampling operations are included in U-Net model of this study. As a result, the dimension of the lowest feature map is  $32 \times 32 \times 1024$ . Unlike DeepLabV3+, it enables to maintain the edge information of the final classification map by skip connection structures of U-Net. However, the independent spectral information is lost because the feature maps are mixed in a single encoder architecture.

## 3) SiU-Net

General CNN-based deep learning model structure consists of a single input node and a single encoder. In the single input node structure, all bands of the input data are stacked and inputted. And in the convolution operations, one feature map is generated by summing the convolution results of each channel. By the serial convolutional operation within a single encoder, multiple input feature maps are mixed and output feature maps are generated. In this operation, the independent spectral features from multi-band input data are mixed, which results in the loss of useful information for reconstructing the final classification map. Accordingly, dual-branch separated input U-Net (W-Net) was proposed [23].

The design of SiU-Net focuses on effectively deciding the number of separate input nodes based on the data. A prior study showed that distinct input nodes prevent the mixing and loss of independent information [23]. Therefore, aligning the number of input nodes and encoders with the number of bands enables the efficient extraction of each band's unique land cover information. However, it would increase the number of weight parameters, and be require large computing powers.

To optimize trade-offs, the numbers of input nodes and encoders should be determined by similarities between each band. Comparing correlation coefficients of each band can be used for assessing band similarity. High correlation indicates high similarity, while low suggests containing unique information. Similar bands should be grouped; less similar ones are separately inputted. Setting threshold or clustering analysis of a correlation coefficient allows for spectral band grouping and designing a SiU-Net structure based on these groups. Grouped bands allow for the determination of separated input nodes and encoders based on these groups. By designing SiU-Net based on the correlation coefficients of each spectral band, we ensure that the data generated by each encoder is as distinct as possible.

It is generally accepted that correlation between RGB and NIR bands is low. From that, in this study, we adopted modified U-Net architecture with two separate input nodes and encoders for RGB and NIR bands. This separation ensures that the unique features of both RGB and NIR are independently processed, enhancing the model's performance in tasks requiring nuanced spectral differentiation.

Figure 2 illustrates the more detailed architecture of SiU-Net. The SiU-Net used in this study incorporates dual encoders, allowing for the independent generation of feature maps from both RGB and NIR bands respectively. Every convolution and max-pooling operation in each encoder are same as U-Net encoder architecture. The  $3 \times 3$  convolution block are applied and then  $2 \times 2$  max-pooling operation is following. This process is repeated four times in each encoder. At every max-pooling operation, the number of convolutional filters is doubled. The decoder is also designed to be the same as that of U-Net except for the skip connection with feature maps generated by two encoders. Due to the concatenation of feature maps from two encoders, the initial



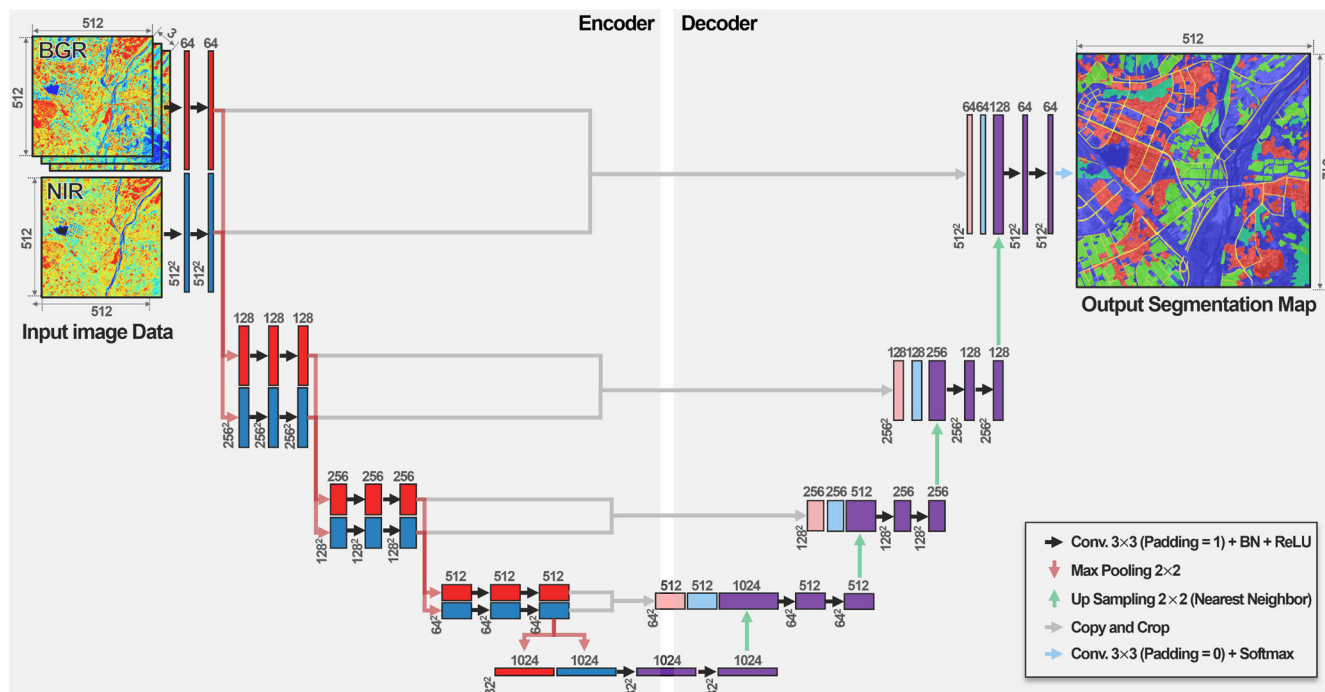


FIGURE 2. The structure of separated-input-based U-Net used in this study.

number of skip-linked feature maps at the decoder is 4/3 times greater than in the same layer of U-Net. Additionally, the initial feature maps dimension of the lowest layer is  $32 \times 32 \times 2048$  (which is  $1024 + 1024$ ), doubling the dimension of U-Net. From the above, we can expect that SiU-Net has two key advantages: 1) It can separately extract unique information from RGB and NIR bands, and 2) It generates more feature maps than U-Net, enabling the extraction of more information for the land cover classification [23].

**B. PERFORMANCE EVALUATION**

Pixel accuracy, precision, recall, and F1 score are used to evaluate classification performance of models [3], [30], [31]. Pixel accuracy is calculated by the percentage of accurately classified pixel in the images. It is the simple and powerful index but has a tendency of being biased in data imbalance case. Precision, recall and F1 score can fulfil shortcomings of pixel accuracy. Precision is the percentage of correctly classified pixels among the pixel predicted as true. Recall can be calculated by dividing the number of correctly classified pixels by the pixel annotated as true. And F1 score is a harmonic average of precision and recall. Precision, recall and F1 score are suitable for not only applying to imbalanced dataset, but also evaluating performance for each class [32]. In addition, precision-recall curve and average precision were adopted to assess performances of models themselves. Precision-recall curve can graphically show the relationship between precision and recall for different threshold [33]. And average precision is the area under the precision-recall curve, which can quantify the

relationship between precision and recall. Average precision is in the range from 0 to 1. It is regarded as the better the performance when average precision close to 1. In this paper, we intend to employ these quantitative metrics to evaluate the classification performance of each model as well as individual classes. Moreover, we will examine the trends in performance variations according to the proportion of the classes, thus demonstrating how the proposed model can be effectively applied.

**III. STUDY DATASET**

In this study, we utilized 2020 satellite-derived land cover dataset generated using Sentinel-2 bands of blue (490nm), green (560nm), red (665 nm), and NIR (842 nm) [27]. 2020 satellite-derived land cover dataset were produced for Seoul, Gyeonggi-do, Gangwon-do, and Chungcheong-do, Korea. The study dataset was constructed using Sentinel-2 LIC data captured over the area during the periods of September to November 2019, March to June 2020, March to August 2019, and December 2019 to February 2020. Only images with cloud coverage of less than 25% were selected [27]. The data are constructed in 300 pairs of input images and the annotation data from those area. The size of input images is the  $512 \times 512$  with 4 bands, which are blue, green, red and NIR bands from Sentinel-2 data, which spatial resolution is 10 m as same as that of Sentinel-2. The classes of the land cover dataset are annotated by totally 6 classes, which are building, road, paddy field, dry field, forest, and unclassified area.

**TABLE 1. Annotation criteria for 2020 satellite-derived land cover dataset.**

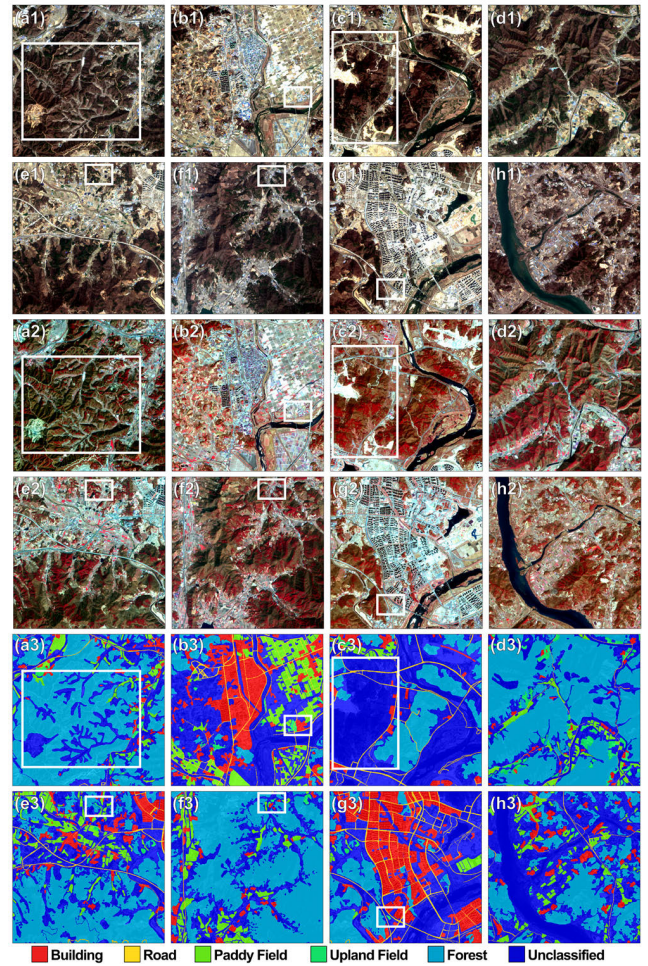
Land cover	Description
Building	- Including every building area - Not including facilities such as facility cultivation area, solar panel, sewage area - The area of 10,000 m <sup>2</sup> or more
Road	- Roads accessible to vehicles - the width of 36 m or more
Paddy Field	- A flooded field for growing rice - The area of 50,000 m <sup>2</sup> or more
Field	- The field for plant without water - Not including orchards, facility cultivation area. - The area of 50,000 m <sup>2</sup> or more
Forest	- The vegetation area of broad-leaved trees, coniferous trees, mixed tree - The area of 100,000 m <sup>2</sup> or more
Unclassified	- The area that does not satisfy the above criteria

Table 1 showed the brief information of the annotation criteria [26]. Every class was annotated only if they met the annotation criteria except for unclassified class. Unclassified class indicates the areas where the conditions (Table 1) are not satisfied. There is also an annotation criterion about minimum area or width of the objects. The minimum area for building, paddy field, upland field and forest classes were 10,000, 50,000, 50,000, and 100,000 m<sup>2</sup> respectively. And the minimum width of the road class was 36 m. If the minimum areas or width are not satisfied, the areas were labeled as unclassified class. Thus, some unclassified areas may include actual building, road, paddy field, upland field, and forest areas. In other words, the well-trained deep learning model may detect actual classes, though those were labeled as unclassified in annotation data. And it can affect the final classification performance.

Figure 3 showed the examples of 2020 satellite-derived land cover dataset included in test dataset. Figure 3 (a1~h1) and (a2~h2) illustrate the true color and color-infrared imagery of some input data. And the label data paired with the input images as shown in Figure 3 (a3~h3). Road network could be recognized in the white box on Figure 3 (a). However, it is annotated as the unclassified class since the width does not exceed 36 m. Most of class can be found on Figure 3 (b, e, g, h). Especially, the linear pattern and the angular boundary are conspicuous, as artificial structures are distributed on wide area such as buildings and roads. Thus, we expected that those can be used to qualitatively analyze the edge-preserving classification performance of the models. On Figure 3 (c), there is the annotation error that forest area of the white box was not annotated. The forest area is widely distributed in Figure 3 (d) and (f). And Upland field class are exquisitely rarely identified, which are located in the white box of Figure 3 (b), (e), (f), and (g).

**IV. RESULTS AND DISCUSSION**

Table 2 showed the pixelwise proportions of each class. The proportions for building, road, paddy field, upland field, forest, and unclassified classes were about 5.38%,



**FIGURE 3. Study dataset: (a1-h1) True color visualization (RGB map) of input images; (a2-h2) color infrared visualization of input images; (a3-h3) annotation data, which refers to a false-color composition using NIR, red, and green spectral bands mapping to RGB respectively.**

**TABLE 2. Pixel-wise proportion of each class.**

Land cover	Proportion (%)
Building	5.38
Road	1.38
Paddy Field	2.62
Field	0.84
Forest	63.78
Unclassified	26.00

1.38%, 2.62%, 0.84%, 63.78% and 26.00% respectively. Approximately 90% of the pixels are annotated into the forest or unclassified classes, as well as only 0.84% of pixels are annotated into upland field class. These data imbalance can lead to classification performance difference for each class.

Correlation coefficients of each band are principal criteria to design SiU-Net structure. To clarify our approach, we compared the correlation coefficients of each band before training deep learning models, although it is known that the correlation between the visible and the NIR bands is low.



TABLE 3. Correlation coefficient of each band.

BAND	BLUE	GREEN	RED	NIR
BLUE	1.00	0.98	0.92	0.64
GREEN	0.98	1.00	0.96	0.71
RED	0.92	0.96	1.00	0.73
NIR	0.64	0.71	0.73	1.00

Table 3 showed the correlation coefficients of each band, which were calculated from the study dataset. Among the visible bands, which are blue, green, and red bands, it showed a high correlation of 0.9 or more as given by Table 3. On the other hand, the correlation coefficient in the NIR and visible bands are 0.64, 0.71, and 0.73 for blue, green, and red bands respectively. From these correlation coefficients, we verified NIR band of study data also contains independent information against visible bands. Based on our hypothesis, we anticipated that applying separated input nodes and encoders for the RGB and NIR bands respectively would enhance land cover classification performance.

To train each model and to compare the classification performance, the training, validation, and test data were randomly sampled at the rate of about 64% (192 patches), 16% (49 patches), and 20% (59 patches) respectively, which do not overlap each other. As a forementioned at section III, forest and unclassified classes are dominant. This data imbalance can influence final classification performance. However, the additional data balancing and augmentation was not performed. Thus, we expected that the classification performance of each class also skewed on the forest and unclassified classes.

TABLE 4. The principal hyperparameters (kernel initializer, loss function, optimizer, learning rate, and epoch) for training DeepLabV3+, U-Net, and SiU-Net.

MODELS	DEEPLABV3+	U-NET	SIU-NET
KERNEL INITIALIZER	HE NORMAL	HE NORMAL	HE NORMAL
LOSS FUNCTION	CROSS ENTROPY	CROSS ENTROPY	CROSS ENTROPY
OPTIMIZER	NADAM	ADAM	ADAM
LEARNING RATE	0.00001	0.0001	0.0001
EPOCH	800	800	800

Table 4 showed the principal hyperparameters of DeepLabV3+, U-Net, and SiU-Net. He normal initializer was adopted since the activation function of the models is ReLU. For the loss function, cross entropy loss was adopted because it is the most commonly used function for the task of semantic classification. After several iterative trials, we empirically determined the optimizers and learning

rates, adopting NAdam with a learning rate of 0.00001 for DeepLabV3+, and Adam with a learning rate of 0.0001 for U-Net and SiU-Net, respectively. This decision was guided by a rule of thumb, reflecting our practical experience and judgment in optimizing model performance [34], [35]. And we trained the models using training data from the scratch by repeating 800 epochs. After training of 800 epochs, the pixel accuracies of validation data were converted into a second decimal place, which were 0.86, 0.91, and 0.96 for DeepLabV3+, U-Net and SiU-Net respectively. U-Net and SiU-Net converged well when applied with the same optimizer and learning rate. However, DeepLabV3+ did not exhibit the same behavior. This discrepancy is understandable given that U-Net and SiU-Net structures are akin, whereas DeepLabV3+ significantly diverges from the other two models. These structural differences justify the variation in hyperparameters for effective training. Consequently, U-Net and SiU-Net were optimally trained with identical hyperparameters, whereas DeepLabV3+ required different settings for optimal performance.

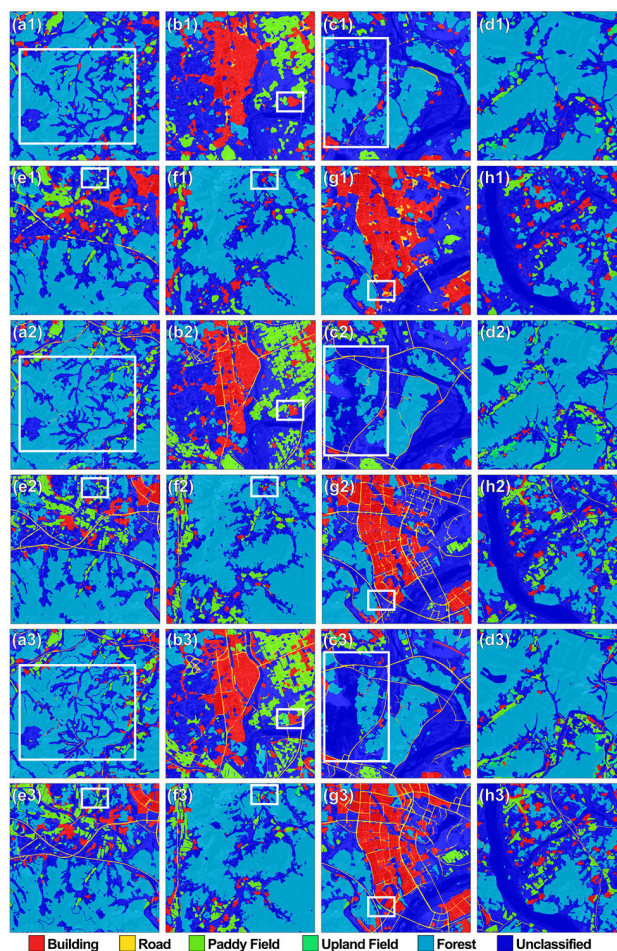


FIGURE 4. Land cover classification maps using (a1-h1) DepLabV3+, (a2-h2) U-Net, and (a3-h3) SiU-Net.

The final models for DeepLabV3+, U-Net and SiU-Net structures were selected by comparing pixel accuracy of

every epoch. Figure 4 illustrated the land cover classification map from test data of a 2020 satellite-derived land cover dataset using the best fitted models for DeepLabV3+, U-Net and SiU-Net. From the Figure 4, we can find that the models well localized each class commonly. Nevertheless, the most significant difference of classification results for each model was the performance on the semantic edge (the boundaries between the classes). DeepLabV3+ could not clearly segment semantic edge and blurred at the boundaries. As shown in Figure 4 (a1-c1, e1-g1), the roads were scarcely classified if the road is narrow. Similarly, it is not well classified that the boundaries of the angular artificial objects such as building, paddy fields and upland fields. On the other hand, the localization and classification of the region were well performed. For example, although the detailed edge information was lost in the Figure 4 (a1-h1), The objects distributed in a wide area such as forests, buildings, paddy fields, and unclassified could be distinguished well. Even the misannotated area shown in the white box of Figure 4 (c1) could be classified accurately. It indicates that the encoder of DeepLabV3+ can extract sufficient information for distinguishing the objects. However, the decoder fails to restore the original spatial resolution of the input data since the final classification map is generated by using the feature maps with 4 and 16 times degraded spatial resolution.

Compared to the classification maps of DeepLabV3+, U-Net (Figure 4 (a2-h2)) and SiU-Net (Figure 4 (a3-h3)) clearly distinguish the boundaries of each class. U-Net and SiU-Net properly restored the linear pattern of the road networks, even the roads that were not annotated due to their narrow width could be detected. In addition, it can be identified that the artificial structures, such as buildings, road, paddy fields, and upland fields, is more clearly distinguished than those of DeepLabV3+ (Figure 4 (a2-h2), (a3-h3)). It indicates that the edge information of final classification maps was well maintained due to the decoder utilizing full-resolution feature maps to generate them. Especially, SiU-Net surpassed U-Net in restoring edge information and detecting the objects. At Figure 4 (b2), and (g2), False alarms of the road class can be found. The road objects classified by U-Net showed the noisy and disconnected pattern. Besides the boundaries of paddy fields was not classified clearly. From those, we can recognize that its angular edge information was lost. On the other hand, the false alarm and misclassification of roads classified by SiU-Net are lower than those of U-Net in the qualitative comparison. The angular characteristics of paddy fields also well recovered in the final classification maps of SiU-Net. Furthermore, SiU-Net could detect the objects which U-Net cannot localize from input data. Upland field area was classified by SiU-Net in the small white boxes on the Figure 4 (b3), (e3), (f3), and (g3), even though they are mis-detected on the final classification maps of U-Net (white boxes on the Figure 4 (b2), (e2), (f2), and (g2)). It indicates that two encoder of SiU-Net extracted independent information from visible and NIR band respectively.

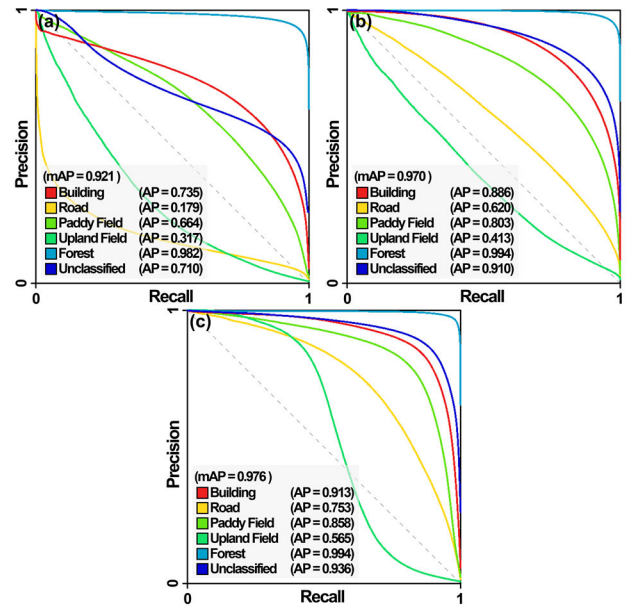


FIGURE 5. Precision-recall curves of (a) DeepLabV3+, (b) U-Net, and (c) SiU-Net; AP and mAP mean average precision and mean average precision respectively.

The precision-recall curve and average precision (AP) score were adopted for comparing the model ability to classify each class. The higher AP, which is area under the precision-recall curve, means the contrast of dividing the class is bigger. Thus, it is related to the performance of segmenting the semantic edge. Precision-recall curves and AP for each model and class were generated by comparing annotation data and final classification maps of test dataset as given by Figure 5. The mean APs (mAP), the weighted mean of each AP, for the DeepLabV3+, U-Net and SiU-Net were about 0.921, 0.970 and 0.976 respectively. The proportion of forest class was overwhelmingly high. Thus, the mAPs of each model were similar each other, which were affected by AP of forest class. The AP of road class in DeepLabV3+ (Figure 5 (a)) was about 0.179, which were more than 3 times lower than those of U-Net (Figure 5 (b)) and SiU-Net (Figure 5 (c)). The AP differences except for the road and forest class were approximately reached to 0.2 comparing U-Net and SiU-Net. Only forest class of DeepLabV3+ showed similar performance with U-Net and SiU-Net. The low APs of DeepLabV3+ trained by 2020 satellite-derived land cover dataset were due to its poor classification performance on the semantic edge. The APs of SiU-Net were about 0.026 to 0.155 higher than those of U-Net except for the forest class. The APs of upland field for SiU-Net and U-Net were about 0.565 and 0.413 respectively, which showed the largest AP difference of about 0.152. The AP differences of road and paddy field were 0.133 and 0.055 respectively, showing the next largest differences. Meanwhile the differences of AP for building, forest, and unclassified area are only 0.027, 0.000 and 0.026 respectively. From the AP difference between U-Net and SiU-Net, we can

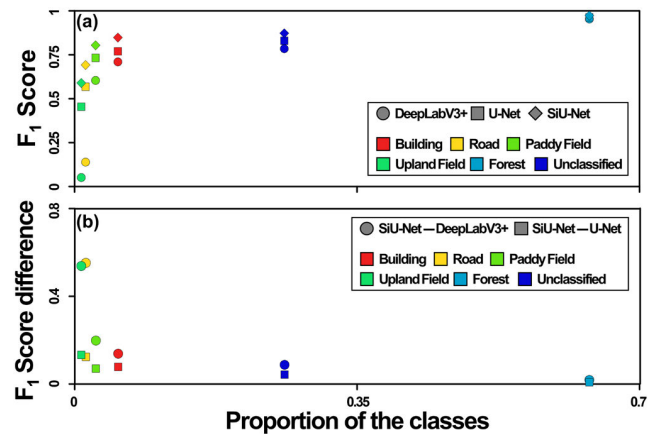


find the relation between the difference and the pixel-wise proportion of the classes, which is that AP differences are low as the pixel-wise proportion of the classes increases. And it can be identified that the efficiency of SiU-Net was higher when the proportion of the classes was lower.

**TABLE 5. Performance evaluation indices of the DeepLabV3+, U-Net, and SiU-Net using pixel accuracy (Acc.), precision (Prec.), recall (Rec.), F1 score, and mean F1 score.**

model	Class	Acc.	Prec.	Rec.	F1 score	Mean F1 score
DEEPLABV3+	Building	0.878	0.721	0.700	0.710	0.541
	Road		0.376	0.085	0.139	
	Paddy Field		0.713	0.525	0.605	
	Field		0.029	0.228	0.051	
	Forest		0.943	0.965	0.954	
	Unclassified		0.773	0.798	0.785	
U-NET	Building	0.905	0.879	0.685	0.770	0.720
	Road		0.639	0.511	0.568	
	Paddy Field		0.689	0.782	0.733	
	Field		0.503	0.415	0.455	
	Forest		0.960	0.969	0.964	
	Unclassified		0.818	0.842	0.830	
SIU-NET	Building	0.930	0.858	0.840	0.849	0.797
	Road		0.741	0.648	0.692	
	Paddy Field		0.786	0.823	0.804	
	Field		0.768	0.478	0.589	
	Forest		0.966	0.979	0.973	
	Unclassified		0.879	0.867	0.873	

An additional quantitative analysis was performed to verify the effect of input separation of RGB and NIR data in deep learning model structure for land cover classification. Table 5 summarized performance evaluation indices for DeepLabV3+, U-Net, and SiU-Net, which are pixel accuracy, precision, recall, F1 score, and Mean F1 score. Same as precision-recall curves, performance evaluation indices were calculated by comparing annotation data and final classification maps of test dataset. The pixel accuracy for DeepLabV3+, U-Net, and SiU-Net were 0.878, 0.905, and 0.930 respectively. Since the class of the data is largely skewed to the forest and unclassified areas, the pixel accuracies depend on the pixel accuracy of those classes. Thus, pixel accuracies of each model were similar to each other. Unlike pixel accuracy, the F1 score, which is a harmonic average of precision and recall, can be utilized to evaluate quantitative performance on the imbalanced data. The best F1 score can be found at the SiU-Net. All F1 scores of SiU-Net for each class were higher than that of DeepLabV3+ and U-Net. While DeepLabV3+ showed worst F1 score for each class. Mean F1 scores for DeepLabV3+, U-Net, and SiU-Net were 0.541, 0.720, and 0.797 respectively. SiU-Net and DeepLabV3+ showed best and worst performance respectively.



**FIGURE 6. The relationship between the proportion of each class and F1 score; (a) The variation of F1 score according to proportion of the classes; (b) The variation of F1 score difference against SiU-Net according to proportion of the classes.**

Data balance has the most important influence on the unbiased performance of the deep learning model. From the AP difference we found the relation between the difference and the pixel-wise proportion of the classes. Similar pattern can be found on F1 score of each class. Figure 6 showed the relationship between the proportion of each class and F1 score, which were referred on Table 2 and 5. The shapes and the colors of symbols indicates models and class respectively (Figure 6). From Figure 6 (a), F1 scores of upland fields, which has lowest proportion, were especially lowest among them of the classes in common for all models. Besides, the F1 score of each class commonly showed an increasing trend along to the proportion, although the trend variations are different in each model. This ascending trend is influenced by data imbalance of each class. The F1 score differences between forest and upland field classes were about 0.903, 0.509, and 0.384 for DeepLabV3+, U-Net and SiU-Net respectively. In addition, standard deviations of F1 scores for DeepLabV3+, U-Net and SiU-Net are about 0.333, 0.167 and 0.125 respectively. SiU-Net showed not only the highest F1 score, but also the most stable performance for each class.

Based on the stable performance for each class, it would be seemed that SiU-Net may reduce skewed classification performance according to classes due to data imbalance. Figure 6 (b) showed the variation of F1 score difference against SiU-Net for DeepLabV3+ and U-Net according to proportion of the classes. As given by Figure 6 (b), the F1 score difference between SiU-Net and other models showed a descending trend as the proportion of the class increased except for the road class of DeepLabV3+. DeepLabV3+ has lost its detailed information, and it is difficult to extract spatially linear information of the road. Thus, the road class of the model is out of the declining trend. Meanwhile, F1 score differences of upland field against SiU-Net were about 0.538 and 0.134 for DeepLabV3+ and U-Net respectively. F1 score differences of the forest, the most dominant class, were converged to nearly zero, which are 0.019 and

0.009 for DeepLabV3+ and U-Net respectively. In other words, the performance improvement of SiU-Net seems to be greatly affected by the number of training data. The number of training data increases, the magnitude of performance improvement by input separation decreased. Consequently, it is expected that, SiU-Net can be a good alternative model if the data imbalance is severe and insufficient quantity of the training data.

## V. CONCLUSION

Satellite based RGB and NIR data have been successfully used for mapping land cover. CNN-based deep learning approach accelerate the pace of the improvement. Most recent studies have been adopted the single encoder-decoder structure to segment land cover from the multi-spectral data, which leads to losing independent spectral information from each multi-spectral band.

In this study, we implemented separated-input-based U-Net in the consideration of correlation between RGB and NIR bands. We trained the models using 2020 satellite-derived land cover dataset, which contains RGB and NIR bands, from AI Hub, NIA, Korea. And we verified the effectiveness of input separation by comparing DeepLabV3+, and U-Net. Precision-recall curve, average precision, precision, recall, F1 score were adopted as performance evaluation indices. The mean APs for DeepLabV3+, U-Net, and SiU-Net were about 0.921, 0.970, and 0.976 respectively. The AP differences except for the forest class were approximately 0.15 or more comparing U-Net and SiU-Net. The AP differences, which calculated from AP of SiU-Net and U-Net, of upland field and forest were about 0.152, and 0.000 respectively. Mean F1 scores for DeepLabV3+, U-Net, and SiU-Net were 0.541, 0.720, and 0.797 respectively, and SiU-Net showed the best performance in AP and F1 score. The F1 score differences between forest and upland field classes were respectively about 0.903, 0.509, and 0.384 for DeepLabV3+, U-Net and SiU-Net. And standard deviations of F1 scores for DeepLabV3+, U-Net and SiU-Net are about 0.333, 0.167 and 0.125 respectively. SiU-Net showed most stable performance for each class. In other words, SiU-Net might reduce data imbalance problem. F1 score differences of upland field against SiU-Net were about 0.538 and 0.134 for DeepLabV3+ and U-Net respectively. And F1 score differences of the forest, the most dominant class, were 0.019 and 0.009 for DeepLabV3+ and U-Net respectively. From that, we can conclude the effectiveness of SiU-Net is more emphasized when the training data are not sufficient.

Finally, we can summarize as (1) DeepLabV3+ is not suitable for segmenting land cover from satellite-based RGB and NIR data due to the low spatial resolution of the final classification map; (2) SiU-Net is effective to segment land cover from satellite-based RGB and NIR data with 0.930, 0.976, and 0.797 of pixel accuracy, mAP, and mean F1 score; (3) The efficiency of SiU-Net is more emphasized as the number of data is smaller; (4) SiU-Net can reduce the biased classification performance caused by data imbalance.

Furthermore, this research has certain limitations. Typically, the independence between RGB and NIR, a well-established concept in remote sensing, guided the qualitative design of the two encoders. However, in this study, we only separated input node with a qualitative criterion of correlation coefficients. Thus, it is still unsolved that a qualitative level of the correlation coefficient can effectively enhance the classification performance by input separation. Future studies that quantitatively determine the number of encoders based on the correlation coefficient of spectral bands are expected to enhance the effectiveness of land cover classification using multi-spectral satellite data.

## ACKNOWLEDGMENT

This study was performed by using a 2020 satellite-derived land cover dataset. The study data can be download on AI Hub (<https://aihub.or.kr>).

## REFERENCES

- [1] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sens.*, vol. 12, no. 15, p. 2495, Aug. 2020, doi: [10.3390/rs12152495](https://doi.org/10.3390/rs12152495).
- [2] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016, doi: [10.1109/TGRS.2016.2584107](https://doi.org/10.1109/TGRS.2016.2584107).
- [3] J.-W. Yu, Y.-W. Yoon, W.-K. Baek, and H.-S. Jung, "Forest vertical structure mapping using two-seasonal optic images and LiDAR DSM acquired from UAV platform through random forest, XGBoost, and support vector machine approaches," *Remote Sens.*, vol. 13, no. 21, p. 4282, Oct. 2021, doi: [10.3390/rs13214282](https://doi.org/10.3390/rs13214282).
- [4] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. Amsterdam, The Netherlands: Elsevier, 2006.
- [5] J. B. Campbell and R.-H. Wynne, *Introduction to Remote Sensing*, 5th ed. New York, NY, USA: Guilford Press, 2011.
- [6] W.-K. Baek, S. H. Park, N. K. Jeong, and S. Kwon, "A study for the techniques and applications of NIR remote sensing based on static analyses of NIR-related papers," *Korean J. Remote Sens.*, vol. 33, nos. 3–5, pp. 889–900, Aug. 2017, doi: [10.7780/kjrs.2017.33.5.3.11](https://doi.org/10.7780/kjrs.2017.33.5.3.11).
- [7] M. Liu, B. Fu, S. Xie, H. He, F. Lan, Y. Li, P. Lou, and D. Fan, "Comparison of multi-source satellite images for classifying Marsh vegetation using DeepLabV3 plus deep learning algorithm," *Ecolog. Indicators*, vol. 125, Jun. 2021, Art. no. 107562, doi: [10.1016/j.ecolind.2021.107562](https://doi.org/10.1016/j.ecolind.2021.107562).
- [8] K. Yuan, X. Zhuang, G. Schaefer, J. Feng, L. Guan, and H. Fang, "Deep-learning-based multispectral satellite image segmentation for water body detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7422–7434, 2021, doi: [10.1109/JSTARS.2021.3098678](https://doi.org/10.1109/JSTARS.2021.3098678).
- [9] J. V. Solórzano, J. F. Mas, Y. Gao, and J. A. Gallardo-Cruz, "Land use land cover classification with U-Net: Advantages of combining Sentinel-1 and Sentinel-2 imagery," *Remote Sens.*, vol. 13, no. 18, p. 3600, Sep. 2021, doi: [10.3390/rs13183600](https://doi.org/10.3390/rs13183600).
- [10] A. Stoian, V. Poulain, J. Inglada, V. Poughon, and D. Derksen, "Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems," *Remote Sens.*, vol. 11, no. 17, p. 1986, Aug. 2019, doi: [10.3390/rs11171986](https://doi.org/10.3390/rs11171986).
- [11] S. Lee, W.-K. Baek, H.-S. Jung, and S. Lee, "Susceptibility mapping on urban landslides using deep learning approaches in Mt. Umyeon," *Appl. Sci.*, vol. 10, no. 22, p. 8189, Nov. 2020, doi: [10.3390/app10228189](https://doi.org/10.3390/app10228189).
- [12] I. Mirpulatov, S. Illarionova, D. Shadrin, and E. Burnaev, "Pseudo-labeling approach for land cover classification through remote sensing observations with noisy labels," *IEEE Access*, vol. 11, pp. 82570–82583, 2023, doi: [10.1109/ACCESS.2023.3300967](https://doi.org/10.1109/ACCESS.2023.3300967).
- [13] W.-K. Baek, M.-J. Lee, and H.-S. Jung, "The performance improvement of U-Net model for landcover semantic segmentation through data augmentation," *Korean J. Remote Sens.*, vol. 38, nos. 2–6, pp. 1663–1676, Dec. 2022, doi: [10.7780/kjrs.2022.38.6.2.8](https://doi.org/10.7780/kjrs.2022.38.6.2.8).

- [14] S. Cuypers, A. Nascetti, and M. Vergauwen, "Land use and land cover mapping with VHR and multi-temporal Sentinel-2 imagery," *Remote Sens.*, vol. 15, no. 10, p. 2501, May 2023, doi: [10.3390/rs15102501](https://doi.org/10.3390/rs15102501).
- [15] E.-R. Lee, W.-K. Baek, and H.-S. Jung, "Mapping tree species using CNN from bi-seasonal high-resolution drone optic and LiDAR data," *Remote Sens.*, vol. 15, no. 8, p. 2140, Apr. 2023, doi: [10.3390/rs15082140](https://doi.org/10.3390/rs15082140).
- [16] J.-W. Yu and H.-S. Jung, "Forest vertical structure mapping using multi-seasonal UAV images and LiDAR data via modified U-Net approaches," *Remote Sens.*, vol. 15, no. 11, p. 2833, May 2023, doi: [10.3390/rs15112833](https://doi.org/10.3390/rs15112833).
- [17] Kaggle. (Dec. 15, 2016). *DSTL Satellite Imagery Feature Detection*. [Online]. Available: <https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection>
- [18] S. B. Jónsson, "RGB and multispectral UAV image classification of agricultural fields using a machine learning algorithm," M.S. thesis, Dept. Phys. Geogr. Ecosyst. Sci., Lund Univ., Lund, Sweden, 2019.
- [19] M. O. Gani, S. Kuiry, A. Das, M. Nasipuri, and N. Das, "Multispectral object detection with deep learning," in *Proc. CICBA*, Santiniketan, India, 2021, pp. 105–117.
- [20] I. Ahmed, M. Ahmad, F. A. Khan, and M. Asif, "Comparison of deep-learning-based segmentation models: Using top view person images," *IEEE Access*, vol. 8, pp. 136361–136373, 2020, doi: [10.1109/ACCESS.2020.3011406](https://doi.org/10.1109/ACCESS.2020.3011406).
- [21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: [10.1109/TPAMI.2016.2644615](https://doi.org/10.1109/TPAMI.2016.2644615).
- [22] G. L. Oliveira, "Encoder–decoder methods for semantic segmentation: Efficiency and robustness aspects," Ph.D. dissertation, Dept. Fac. Tech., Albert-Ludwigs-Universität Freiburg, Fahrenbergplatz, Germany, 2019.
- [23] B. Hou, Q. Liu, H. Wang, and Y. Wang, "From W-Net to CDGAN: Bitemporal change detection via deep learning techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1790–1802, Mar. 2020, doi: [10.1109/TGRS.2019.2948659](https://doi.org/10.1109/TGRS.2019.2948659).
- [24] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, Munich, Germany, Sep. 2018, pp. 801–818.
- [25] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Munich, Germany, Oct. 2015, pp. 234–241.
- [26] S.-H. Lee and M.-J. Lee, "A study of establishment and application algorithm of artificial intelligence training data on land use/cover using aerial photograph and satellite images," *Korean J. Remote Sens.*, vol. 37, nos. 1–5, pp. 871–884, Aug. 2021, doi: [10.7780/kjrs.2021.37.5.1.4](https://doi.org/10.7780/kjrs.2021.37.5.1.4).
- [27] AI Hub. (Jun. 18, 2021). *Satellite-Derived Landcover Dataset*. [Online]. Available: <https://www.aihub.or.kr/aihubdata/data/list.do?pageIndex=1&currMenu=115&topMenu=100&dataSetSn=&srchdataCICode=DATA001&searchKeyword=>
- [28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [30] W.-K. Baek and H.-S. Jung, "Performance comparison of oil spill and ship classification from X-band dual- and single-polarized SAR image using support vector machine, random forest, and deep neural network," *Remote Sens.*, vol. 13, no. 16, p. 3203, Aug. 2021, doi: [10.3390/rs13163203](https://doi.org/10.3390/rs13163203).
- [31] S.-H. Lee and M.-J. Lee, "A study on deep learning optimization by land cover classification item using satellite imagery," *Korean J. Remote Sens.*, vol. 36, nos. 2–6, pp. 1591–1604, Dec. 2020, doi: [10.7780/kjrs.2020.36.6.2.9](https://doi.org/10.7780/kjrs.2020.36.6.2.9).
- [32] W.-K. Baek, H.-S. Jung, and D. Kim, "Oil spill detection of Kerch strait in November 2007 from dual-polarized TerraSAR-X image using artificial and convolutional neural network regression models," *J. Coastal Res.*, vol. 102, no. sp1, pp. 137–144, Dec. 2020, doi: [10.2112/si102-017.1](https://doi.org/10.2112/si102-017.1).
- [33] W.-K. Baek, Y.-S. Lee, S.-H. Park, and H.-S. Jung, "Classification of natural and artificial forests from KOMPSAT-3/3A/5 images using deep neural network," *Korean J. Remote Sens.*, vol. 37, nos. 3–6, pp. 1965–1974, Dec. 2021, doi: [10.7780/kjrs.2021.37.6.3.5](https://doi.org/10.7780/kjrs.2021.37.6.3.5).
- [34] D. Choi, C. J. Shallue, Z. Nado, J. Lee, C. J. Maddison, and G. E. Dahl, "On empirical comparisons of optimizers for deep learning," 2019, *arXiv:1910.05446*.
- [35] G. Joo, C. Park, and H. Im, "Performance evaluation of machine learning optimizers," *J. IKEEE*, vol. 24, no. 3, pp. 766–776, Sep. 2020, doi: [10.7471/ikeee.2020.24.3.766](https://doi.org/10.7471/ikeee.2020.24.3.766).



**WON-KYUNG BAEK** was born in Daegu, South Korea, in 1991. He received the M.S. and Ph.D. degrees in remote sensing and deep learning from the University of Seoul, Seoul, South Korea, in 2017 and 2022, respectively. He is currently a Senior Research Scientist with Korea Ocean Satellite Center, Korea Institute of Ocean Science and Technology. His research interests include the improvement of surface deformation mapping methods based on synthetic aperture radar (SAR); applying multi-temporal SAR interferometry (InSAR) to measure gradual surface deformation; two- or three-dimensional surface displacements mapping by combining InSAR, multiple-aperture InSAR (MAI), and offset tracking measurements; target detection, classification, segmentation, and prediction from satellite and geographic information system (GIS) data; and multi-temporal satellite based wetland topography analysis.



**MOUNG-JIN LEE** received the M.S. degree in environmental sciences and the Ph.D. degree in remote sensing and climate change from Yonsei University, Seoul, South Korea, in 2003 and 2012, respectively. He is currently a Senior Research Fellow with the Water and Land Research Group/Division of Environmental Planning, Korea Environment Institute. His research interests include the development of algorithms related to deep learning network for remote sensing applications; data mining with geospatial information; analysis of climate change data (mitigation and adaptation); and multi-climate change data processing and fusion, thermal remote sensing, and multi-temporal optical and thermal sensing. He has been developing algorithms for remote sensing applications related to remote sensing applications using deep learning; time series analysis of climate change scenarios; and multi-data fusion by the integration of big data and AI training data.



**HYUNG-SUP JUNG** (Senior Member, IEEE) received the M.S. degree in geophysics and the Ph.D. degree in remote sensing from Yonsei University, Seoul, South Korea, in 1998 and 2007, respectively.

He is currently a Professor with the Department of Geoinformatics, University of Seoul, Seoul. His primary research interests include the development of algorithms related to synthetic aperture radar (SAR), SAR interferometry (InSAR), multiple-aperture InSAR (MAI), and multitemporal InSAR (MTInSAR); automated geometric correction of multisensor images; and multisensor image processing and fusion, thermal remote sensing, and multitemporal optical and thermal sensing. He has been developing algorithms for remote sensing applications related to 3-D deformation mapping by combining MAI and InSAR; 2-D surface velocity estimation by combining MAI and along-track interferometry; MAI-based ionospheric correction of radar interferograms; multisensor fusion by the integration of optic and SAR, SAR and thermal, and optic and thermal images; automated geometric correction for optic and SAR images; and Earth's surface variation monitoring, such as urban subsidence monitoring, glacier monitoring, volcano monitoring, deforestation monitoring, forest mapping, forest fire mapping, and snow depth estimation.

• • •