remote sensing

MDPI

Article

Application of a Novel Multiscale Global Graph Convolutional Neural Network to Improve the Accuracy of Forest Type Classification Using Aerial Photographs

Huiqing Pei [1,*], Toshiaki Owari [2], Satoshi Tsuyuki [1] and Yunfang Zhong [3]

[1] Department of Global Agricultural Sciences, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan
[2] The University of Tokyo Hokkaido Forest, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Furano 079-1563, Japan
[3] Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Ministry of Education, Hainan University, Haikou 570228, China
* Correspondence: peihq@g.ecc.u-tokyo.ac.jp

Abstract: The accurate classification of forest types is critical for sustainable forest management. In this study, a novel multiscale global graph convolutional neural network (MSG-GCN) was compared with random forest (RF), U-Net, and U-Net++ models in terms of the classification of natural mixed forest (NMX), natural broadleaved forest (NBL), and conifer plantation (CP) using very high-resolution aerial photographs from the University of Tokyo Chiba Forest in central Japan. Our MSG-GCN architecture is novel in the following respects: The convolutional kernel scale of the encoder is unlike those of other models; local attention replaces the conventional U-Net++ skip connection; a multiscale graph convolutional neural block is embedded into the end layer of the encoder module; and various decoding layers are spliced to preserve high- and low-level feature information and to improve the decision capacity for boundary cells. The MSG-GCN achieved higher classification accuracy than other state-of-the-art (SOTA) methods. The classification accuracy in terms of NMX was lower compared with NBL and CP. The RF method produced severe salt-and-pepper noise. The U-Net and U-Net++ methods frequently produced error patches and the edges between different forest types were rough and blurred. In contrast, the MSG-GCN method had fewer misclassification patches and showed clear edges between different forest types. Most areas misclassified by MSG-GCN were on edges, while misclassification patches were randomly distributed in internal areas for U-Net and U-Net++. We made full use of artificial intelligence and very high-resolution remote sensing data to create accurate maps to aid forest management and facilitate efficient and accurate forest resource inventory taking in Japan.

Keywords: deep learning; multiscale global graph convolutional neural network; forest type classification; remote sensing image segmentation; aerial photograph

Citation: Pei, H.; Owari, T.; Tsuyuki, S.; Zhong, Y. Application of a Novel Multiscale Global Graph Convolutional Neural Network to Improve the Accuracy of Forest Type Classification Using Aerial Photographs. *Remote Sens.* **2023**, *15*, 1001. https://doi.org/10.3390/rs15041001

Academic Editors: Pedram Ghamisi, Ce Zhang, Danfeng Hong and Qiqi Zhu

Received: 9 December 2022
Revised: 30 January 2023
Accepted: 4 February 2023
Published: 11 February 2023

1. Introduction

Forest type classification is of fundamental importance for sustainable forest management, i.e., biodiversity modelling [1], management of disturbances [2,3] and fires [4], harvesting [5], biomass evaluation [6], and carbon stock calculations [7]. Unlike urban area classification, forest type classification is challenging given the limitations in training images and ground truth. Expert knowledge is required to prepare training images for convolutional neural networks (CNNs), but there are few open-source datasets for forests [8]. Mountainous terrain is more susceptible to spectral reflectance distortion and canopy shadowing than flat land [9] and has high biodiversity [10]. Moreover, the vegetation is heterogeneous [11–13] because of the effects of climatic and regional parameters [14,15]. In the 1930s, before the advent of aerial photography, forest managers collected information

using ground inventories [16]. The classification of medium-resolution satellite images has been performed since the 1970s [17]. Natural RGB photographs have been used to identify forest tree species [18] and to classify forest types [19]. Although extensive calibration and pre-processing are not required [20], traditional forest mapping and monitoring requires expensive, time-consuming, and potentially unreliable field measurements, along with professional expertise for manual interpretation [19]. Forest labels are assigned to aerial photographs by humans who evaluate textures, shapes, colours, sizes, patterns and associations [21], vegetation composition, forest structural properties (i.e., stem density, tree size, and vertical structure), and surface morphology [22].

Automated approaches for forest type mapping with remote sensing data can be categorised into three types: traditional thresholding methods, classical machine learning methods, and CNN methods. Traditional thresholding methods are limited to colour features (hue, chroma, and lightness) [23,24] and the discrimination of under-story (background) and over-story (tree canopy) signals in forests, given the subtle changes in complicated illumination environments. Classical machine learning methods, i.e., support vector machine (SVM) [25] and random forest (RF) [26], handle classification tasks [27–32] using a small training dataset with Gabor, Canny, Prewitt, Robert, and Gaussian filters applied to aid classification [33,34]. Recently developed CNN methods are effective for representing spatial patterns, such as edges, corners, textures, and abstract shapes, but CNN-based forest type classification with remote sensing data is still a very new field [35]. Most deep learning models are based on classical CNN models (e.g., U-Net [36] and DeepLabv3+ [37]) for mapping non-forest [38,39], forest [40] or tree species [18,37].

U-Net [41] can obtain underlying spatial features using contraction and expansion blocks throughout the encoder and decoder, thereby effectively expanding the horizon of the sensing fields to assess the global contextual environment and to derive detailed information. However, U-Net extracts high-level semantic information using a filter of a certain size and continuously feeds the features to convolutional layers, while ignoring the large amount of low-level information and ignore the correlations between adjacent pixels [42–48]. Redundant information is used many times during the processing flow, which reduces the capacity to represent key features. The skip connection of U-Net++ [49] helps prevent the potential loss of information caused by sampling [36]. However, the number of operation parameters is very large given the dense skip connections [50], which reduces model efficiency and increases the computational load. Also, during feature fusion, upper-layer semantic information is ignored and the network cannot yield effective fine-grained features at the decoding stage, resulting in serious loss of edge and positional information [51]. Principle component analysis (PCA) [52] and vision transformer (ViT) [53] were proposed for extracting minimal features. Research has also focused on using critical local image patches and discarding useless information [47,48,54].

Graph convolutional neural networks (GCNs) have recently been used successfully to analyse irregular data. Graphical spectral, spatial, and geometric data are rich in node and edge information [55,56]. In a previous study, a GCN [57] performed very well when extracting features from irregular graphical data and edge connections to aggregate node information and generate new representations of the nodes. A GCN is limited to shallow layers because of the vanishing gradient. To extract more features during semantic segmentation, several studies [58–60] combined CNNs with GCNs to strengthen the spatial and spectral information and to reduce pixel-level noise by identifying graphical nodes and the spatial relationships between them, as represented by graphical edges. Liu et al. [61] collaborate the Euclidean data-oriented CNN with non-Euclidean data-oriented GCN in a single network to generate complementary spectral-spatial features from the pixel and superpixel levels respectively. Ding et al. [62] and Wang et al. [63] fused two global feature vectors produced by individual CNNs and GCNs; the fused features were optimised. However, robustness may have been low. Peng et al. [64] fused features using MopNet and the parameters of a CNN and GCN; all features were continuously updated via backpropagation guided by joint loss.

Japan has a typical marine climate with abundant rainfall that facilitates forest growth. Nearly 70% of Japan is covered by forests, of which 40% are plantations [65]. Natural mixed forest (NMX) is defined as forest in which broadleaved species account for 25–75% of the coniferous canopies [66]. Plantation and natural forests differ at the leaf scale (i.e., in leaf inclination, morphology, and clumping) [9] and canopy structure scale (i.e., in crown morphology and canopy cover) [67], as well as in spectral characteristics [68–71]. They also differ at the forest stand scale (i.e., in forest composition and diversity and tree distributions and interactions within the forest). It is challenging to segment NMX, natural broadleaved forest (NBL), and conifer plantation (CP) because a mixed forest is more heterogeneous than other forests; it contains both conifers and broadleaved trees. In terms of vertical structure, unlike an NBL with a single-story stand, an NMX has multistorey classes with varying vertical structures within each stand type [72]. Furthermore, the natural forests of Japan are mostly located on steep mountainsides, which makes forest inventory taking and monitoring difficult [19]. Challenges to achieving highly accurate forest type classification include the complexity of natural forest canopy [73], which exhibits multiscale differences [9]. We can easily identify vegetation properties of interest, even with little reference data, in remote sensing data due to the distinct canopy structure or contrasting flowers, while subtle differences [18] and complex relationships require complex algorithms and more ground truth samples to identify specific features [74]. Second, the ambiguity of the boundaries of different forest types [9], which usually have a high chance of land-cover mixing [11,26,30], affect the edge pixels. Finally, redundant content and noise in very high spatial-resolution remote sensing datasets [25] (such as crown shadow effects [75]) and continuous CNN filters [54] lead to intraclass variation and high interclass similarity.

Our main aims were to develop a novel U-shaped deep learning method using a multiscale global graph convolutional neural network (MSG-GCN) to improve forest type classification accuracy. This network has several notable features. First, convolution kernels of different scales are used to capture image features and data from different receptive fields are fused to capture features that reduce the computational complexity when combining semantic information [76,77] from a previous level. Second, a multiscale graph convolution network (MSGCN) module serves as a transitional module between the encoder and decoder. The MSGCN combines the features of the encoding module with its own features to better represent edge and multiscale features [78]. Third, the skip connection is replaced by local attention (LA) [79] that focuses on salient features, thereby reducing the use of redundant information. Fourth, fusion of the decoding layers ensures consideration of both high- and low-level feature information, and improves the attribute decision-making capacity for boundary cells [80]. The main goals of this study are to compare the novel MSG-GCN with other state-of-the-art (SOTA) methods, investigate the specific areas and digital number (DN) of correctly and misclassified forest types, compare U-shaped deep learning models (such as U-Net and U-Net++) and the classical RF machine learning method, and map and classify entire forest areas, as well as determine the spatial distribution of misclassified forest types.

2. Materials and Methods

2.1. Study Area

The study area was the University of Tokyo Chiba Forest (UTCBF) in central Japan (longitude = 140°05′33″ to 10′10″E, latitude = 35°08′25″ to 12′51″N) (Figure 1). The area is in a warm temperate zone with a mean annual temperature of 14.1 °C and mean annual precipitation of 2474 mm [81]. The altitude ranges from 50 to 370 m above sea level. The area exhibits the unique forest landform of the Boso Hills, which merit academic study. The terrain and slopes are generally complex and very steep. The main soil type is brown forest soil and the geological structure comprises marine deposits from the Neogene Period that are partly covered by nonmarine deposits from the Quaternary Period [81].

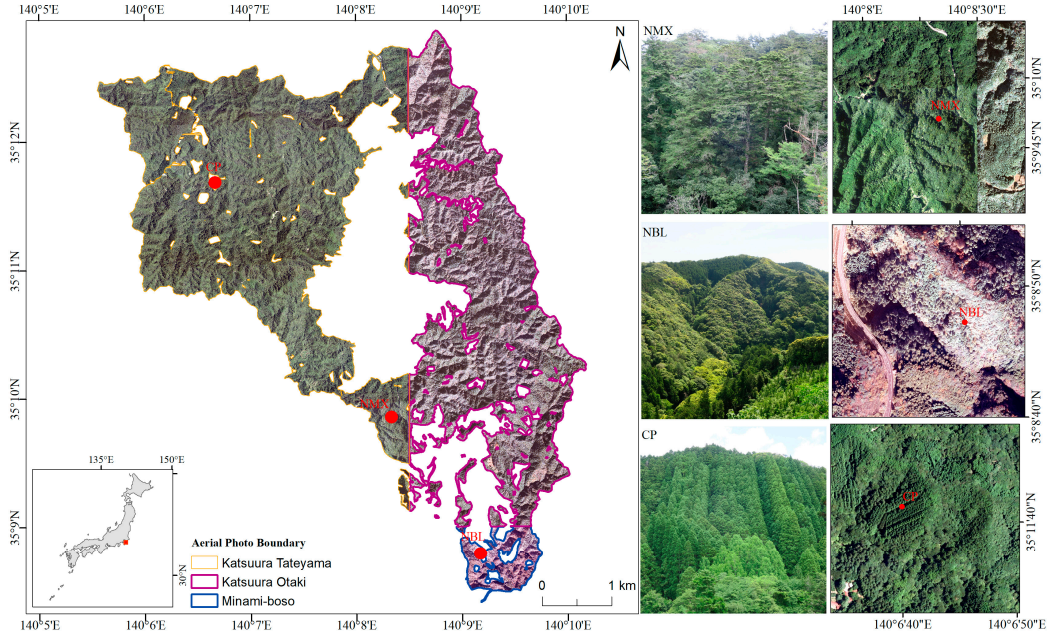

Figure 1. Location of the study area and views of typical forest sites. NMX, NBL, and CP indicate natural mixed forest, natural broadleaved forest, and conifer plantation, respectively. The aerial photographs form three datasets collected at different times in the Katsuura Tateyama, Katsuura Otaki, and Minami Boso districts. The aerial photographs were taken by Geospatial Information Authority of Japan.

The total forest area is approximately 2160 ha and can be divided into three forest types: 267 ha of NMX, 1117 ha of NBL, and 715 ha of CP (Figure 1). NMX areas have complicated horizontal and vertical structures, mixed stands, and various age classes. The main tree species include *Abies firma*, *Tsuga sieboldii*, evergreen *Quercus* spp., and *Castanopsis siedoldii* [81]. Natural forest is composed of many vegetation communities that exhibit natural succession over long periods of time and thus have multivertical layers and a rough texture (Figure 1, NMX). NBL has a complex and diverse crown structure, but the vertical does not exhibit extensive variance. NBL is dominated mainly by *C. siedoldii*, evergreen *Quercus* spp., *Zelkova serrata*, and *Acer* spp. (Figure 1, NBL) [81]. Although NBL areas often have a smoother texture than other areas, NMX areas can be easily misidentified as NBL areas. CP areas are generally monocultures with simple structures characterised by coniferous trees (*Cryptomeria japonica* and *Chamaecyparis obtusa*) that are cone-like in shape with a characteristic crown structure. It is easier to detect this forest type than others. CP established for timber production is manually planted and managed and exhibits well-organised arrangements in columns and rows (Figure 1, CP).

2.2. Data Sources and Preprocessing

We used digitally georeferenced orthorectified aerial photographs taken by the Geospatial Information Authority of Japan (GSI) that were purchased from the Japan Map Center (https://www.jmc.or.jp/, accessed on 18 June 2019). Orthorectification was applied only to the RGB bands; data in the form of $30'' \times 30''$ tiles were available. The orthophoto tiles use the JGD2000 or JDG2011 datum geographical coordinate system. The entire study area is covered by three blocks of GSI aerial photographs collected in different years. Four image tiles were acquired on 25 November 2012 over the Katsuura Tateyama district at a height of 2000 m. The size of each photograph was 3750 × 4650 pixels and the ground sample distance (GSD) was 0.2 m. One hundred image tiles were acquired on 27 October 2017 over the Katsuura Otaki district at a height of 4000 m. The size of each photograph was 1875 × 2325 pixels and the GSD was 0.4 m. Finally, 144 image tiles were acquired on 15 July 2017 over the Minami Boso district at a height of 2350 m. The size of each photo-

graph was 1875 × 2325 pixels and the GSD was 0.2 m. We resampled all photographs to a resolution of 0.2 m using the nearest neighbour algorithm [82], mosaicked them into a single tile, and ensured compatibility of the data with JGD_2011_Japan_Zone_9 projection using ArcGIS software version 10.8 (Esri Inc., Redlands, CA, USA). We used TNTmips for linear raster contrast enhancement of the mosaicked image. Since 1970, the NMXs have not been logged, except to remove obstructive trees [83]. Until the early 1960s, the NBLs were used as fuelwood forests, but logging has been minimal since 1980 [83]. Although the aerial photographs were taken in 2012 and 2017, we assumed that the boundaries between forest types did not change greatly over this period.

The fully annotated ground truth classification map was used as the forest type map of the entire UTCBF (Figure 2). This was produced by UTCBF staff in vector format based on a forest map compilation, forest register, inventory data, and aerial photograph interpretations. The original map had a scale of 1:10,000 and was converted into a raster format tiff file that covered 35,152 × 41,152 pixels using the 'polygon to raster' method in ArcGIS. We reclassified the original classes into three types: NMX, NBL, and CP; other categories, such as the nursery and forest museum [81] used for education and research, covered very small areas and were not considered in this study. The images of these areas were removed, set as non-forest background (BG), and assigned a value of 255. The ground truth map was processed using the raster extraction boundary of UTCBF and then overlapped with the mosaic aerial photograph using the 'extract by mask' routine of ArcGIS to create an aerial photographic image of the same scale and spatial resolution as the ground truth map.

The aerial photograph and ground truth map, both of which are 35,152 × 41,152 in size, were cropped into 512 × 512 tiles without any repetition or overlap to ensure that each image was completely unique. Tiles lacking pixel values were removed. We randomly separated the remaining image and ground truth maps into three parts into training, evaluation, and test datasets at a ratio of 8:1:1 (1981, 248, and 248 image tiles, respectively).

2.3. Model Architecture

2.3.1. Basic Overview

U-Net uses encoders and decoders to model the space of the segmentation target. U-Net is an effective deep learning model for segmentation [41]. However, during modelling, subtle changes in the target may be ignored and sensitivity to fine details can be low. The continuous use of single-scale convolutional filters and pooling operations during encoding makes feature extraction insufficient. Therefore, given the complex content and high-frequency spectrum of a remote sensing image, we developed a locally aggregated MSG-GCN for accurate image segmentation.

Figure 3 shows the overall structure of the segmentation framework. There are three basic components: an encoder with three encoding blocks; a transitional MSGCN module between the contraction and expansion blocks; a decoder with three decoding blocks. The LA [79] integrates encoding and decoding features through an attention mechanism. LA effectively filters out irrelevant information, enhances the capacity to represent key information, and completes the segmentation task. For each encoding block and corresponding decoding block, we use 1 × 1 convolution for compression and 1 × 1 and 3 × 3 convolution layers for excitation. Thus, each encoding and decoding block extracts local and global semantic features by splicing features at different graph scales to capture multiscale information from the segmentation target. Simultaneously, pooling operations at various scales are run using the final decoder, which aids description of the global and contextual semantics [84]. The MSGCN is used by both the encoder and decoder as a transitional layer (strictly speaking, the MSGCN is an encoding block). This accurately identifies multiscale local and global information of the segmentation target. The MSGCN determines interactions between pixels by merging and transferring node information, reduces the intraclass ambiguity caused by different characteristics, and thus aids the modelling of long-term dependence [85].

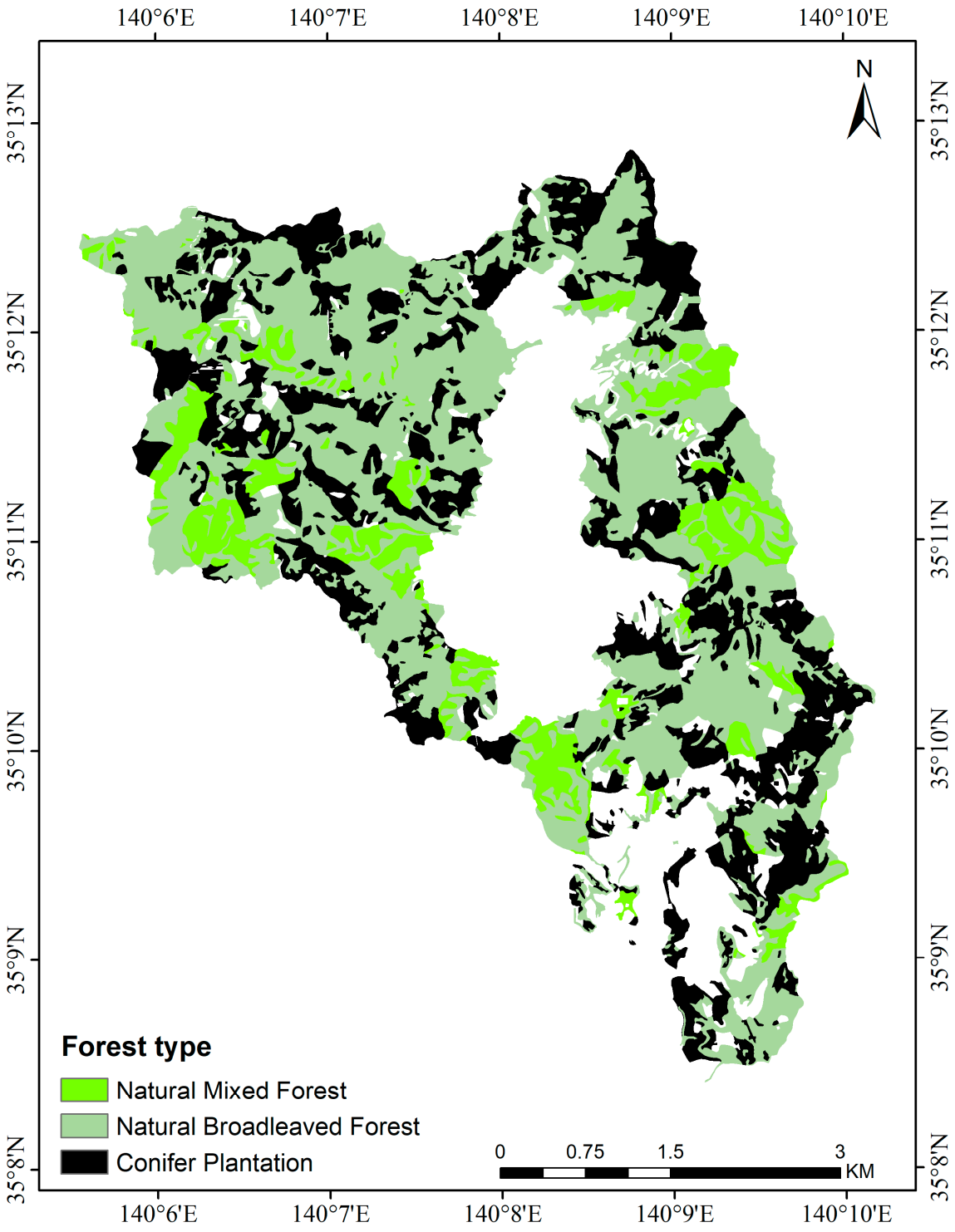

Figure 2. Ground truth map of forest type classification, provided by the University of Tokyo Chiba Forest (UTCBF).

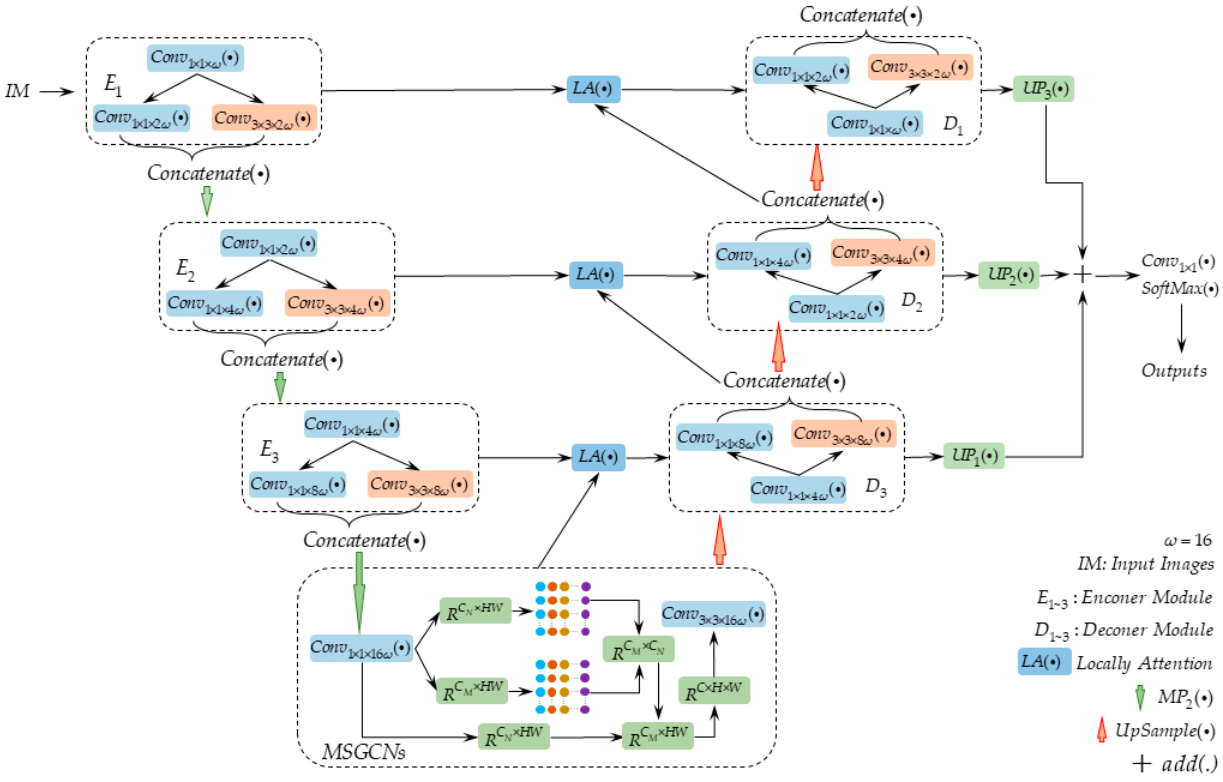

Figure 3. Overall structure of the segmentation framework. *IM* is the original remote sensing image, *E*1,2,3 is the encoder, *D*1,2,3 is the decoder, $Conv1 \times 1 \times \omega$ is the convolution operation with a convolution kernel of 1×1, and ω is the filter. $Conv3 \times 3 \times 2\omega$ is a 3×3 convolution operation with a filter of 2ω, where ω is 16. LA(·) is the local attention module. Concatenate(·) is the matrix concatenation; SoftMax(·) is the classifier; *C*, *H*, and *W* are the channel, height, and width of the feature, respectively. "*UP*" indicates the upsampling operation of bilinear interpolation, where $UP_1(\cdot)$ indicates that the upsampling kernel size is 8×8 and $UP_2(\cdot)$ indicates that it is 4×4. $UP_3(\cdot)$ indicates that the upsampling kernel size is 2×2.

2.3.2. Encoding Module

Convolutional operations at various scales replace the continuous convolution layer of the U-Net model. Specifically, a simple squeeze excitation multiscale [77] is used to model the multiscale information of the target. The number of network parameters and computational complexity are thus reduced and multiscale local semantics are captured more accurately. Each encoding block has an extruded layer, two excitation layers of different scales, and a splicing layer. Then, the coding information is input into the maximum pooling layer for fusion and dilution, which improves multiscale representation. The splicing and pooling process is given by the equation below:

$$\begin{cases} O_{cat}^{k} = Concat([Conv_{1\times1\times2^{\kappa}w}(f_{ex}(x)); Conv_{1\times1\times2^{\kappa}w}(Conv_{3\times3\times2^{\kappa}w}(f_{ex}(x)))]) \\ f_{ex}(x) = Conv_{1\times1\times2^{\kappa-1}w}(x) \\ O_{mp}^{k} = MP(O_{cat}^{k}), k = 1, 2, 3 \end{cases} \quad (1)$$

where O_{cat} is the output of the splicing layer, w is the number of filters, $Conv1 \times 1 \times 2^{k}\omega(\cdot)$ and $Conv3 \times 3 \times 2^{k}\omega(\cdot)$ are convolutions using 1×1 and 3×3 filters, $2^{k}\omega$ is the number of filters, $f_{ex}(x)$ are 'squeeze incentive features', k is the kth encoding module, and $O^{k}{}_{mp}$ is the maximum pooling layer of the kth encoding module.

2.3.3. Multiscale Graph Convolution Network (MSGCN) Module

The CNN described above effectively captures local and global information. To create long-term dependencies and to identify the spatial relationships between different semantic details, the encoded information is sent to a large optimised topological GCN that extracts detailed semantic information and constructs spatial relationships by merging and transmitting node information [59]. A topological graph can be expressed as G = (V,E), where V is the set of nodes and E is the set of edges between nodes. The steps used to construct the spatial relationship are as follows.

Step 1. To improve feature representation, we first send the encoded features to a 1×1 convolutional layer for information conversion. In other words, we embed the data into a unified low-dimension space to compress the features and improve the "expressive" nature of the network; the reconstructed features provide a basis for the subsequent graphical representation. The reconstruction process is described as follows:

$$\begin{cases} O_N = reshape(Conv_{1\times1\times2^{\kappa+1}w}(O_{mp})) \\ O_M = reshape(Conv_{1\times1\times2^{\kappa+1}w}(O_{mp})) \\ O_N \in R^{C\times(H\times W)}, O_M \in R^{(H\times W)\times C} \end{cases} \tag{2}$$

where O_M and O_N are the reconstruction features, O_{mp} is the maximum pooling layer of the convolutional encoding layer, and reshape$(\cdot)$ is the information conversion function.

Step 2. The reconstructed features are used to construct a multiscale graph with dimensions of $C_N \times C_M$, where C_N is a channel feature with N nodes and C_M is a channel feature with M nodes. The Euclidean distance is used to determine the spatial relationships between nodes. The construction of adjacency matrix A is shown in Figure 3 and described by Equation (3). The reconstructed feature map is used to construct the topology map G. If there is a relationship between adjacent feature points, a value of 1 is assigned; if not, the value assigned is 0. Next, graph convolution is used for node learning and optimisation, and node transfer is then employed to gather detailed information on spatial relationships. The multilayer learning graph convolution is given by the following equation:

$$f_G^{(l)} = \sigma(\widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}f_G^{(l-1)}W_G^{(l-1)}) \in R^{C\times C}, \widetilde{A} = A + I \tag{3}$$

where $f_G^{(l)}$ is the output of the lth graph convolution layer, $\sigma(\cdot)$ is the LeakyRelu activation function, $\widetilde{A}$ is the adjacency matrix normalised by the Laplace method, $\widetilde{D}$ is the degree matrix, I is the identity matrix, A is the adjacent scale, and $W_G^{(l-1)}$ is the weight matrix of the $(l-1)$th graph convolution layer.

Step 3. The MSGCM creates effective long-term dependencies between nodes using the node transfer function and efficiently captures the global spatial semantics of target objects using the aggregation function. Given that the initial encoding features contribute to the representation of local semantics, we can reconstruct the features from the third encoding module, fuse them with the multiscale graph convolution features, and input these into the 3×3 convolution layer to further strengthen the representations of local and global features and establish interactions between them.

2.3.4. Decoding Module

In the decoding stage, the decoder module restores the feature maps to the original input size. In U-Net, the expansive path involves upsampling of the feature map followed by a 2×2 convolution that halves the number of feature channels, a concatenation with the corresponding cropped feature map from the contracting path, and two 3×3 convolutions, each followed by a ReLU. A 1×1 convolution at the final layer maps the feature vector to the desired number of classes. Our redesigned expansive path is roughly symmetric to the contracting path with a U-shaped architecture. The parallel 1×1 convolution and 3×3 convolution replaced by two consecutive 3×3 convolutions were used to extract multiscale features. Extrusion and excitation feature learning can better represent target

objects through this process. Furthermore, concatenation of high-dimensional upsampled features can easily lead to confusion regarding the features of objects at different scales, especially when the contour boundaries are blurred and irregular. Therefore, to explicitly learn the position and boundary information from the corresponding encoder blocks and fine-grained feature maps from the previous decoding convolution layer, we designed an LA embedding module to enhance the recognition ability, increase differences among target objects at various scales, and refine the features of different categories of target objects. The LA is calculated using Equations (4) and (5).

$$LA = AttE(O_{mp}, f_D) \tag{4}$$

where LA is the local attention module, $AttE(\cdot)$ is the embedding operation for LA, and O_{mp}, f_D are the maximum pooling output feature of the encoding module and the corresponding position decoding output feature, respectively. LA improves feature representation, eliminates redundant information, and establishes a complementary relationship between high- and low-level features. Thus, when high-level features are insufficiently represented, low-level features are strengthened. To ensure that low-level features do not interfere with high-level ones, a residual module is used to emphasise (or de-emphasise) the importance of low-level features. The LA embedding operation is as follows:

$$LA = \frac{O_{mp} + \alpha_{mp} O_{mp} + f_D + \alpha_D f_D}{2} \tag{5}$$

where α_{mp}, α_D are the attention matrices of the encoding and decoding features, respectively. Low-level detailed feature maps capture rich spatial and concrete information that highlights a target's boundaries and morphology; high-level feature maps capture abstract information, which results in the loss of detailed information. Spatial and size differences make it difficult to fuse the high- and low-level features directly and effectively. As a result, we upscaled each decoding output feature map to the original input size of 512 × 512 using bilinear interpolation. These three feature maps are added directly to the unchanged channels, which increases the amount of information and reduces the model calculations. Adding the decoder stage output can be computed as:

$$\begin{cases} O_{out} = UP_1(f_{D3}) + UP_2(f_{D2}) + UP_3(f_{D1}) \\ Y = SoftMax(Conv_{1\times1}(O_{out})) \end{cases} \tag{6}$$

The output feature map of the different decoding modules f_{D3}, f_{D2}, and f_{D1} and bilinear interpolation upscaling operations of the various filter kernels are represented by $UP_1()$, $UP_2()$, and $UP_3()$, respectively; the kernel sizes are 8 × 8, 4 × 4, and 2 × 2. O_{out} is the total feature plot of the decoded path output.

2.3.5. Loss Function

To ensure the robustness of the feature representations used for learning, we employed Dice loss function [86] to optimise the model and to facilitate attribute decisions for the boundary cells [87]. This effectively suppresses the class imbalance problem [87].

$$L_{dice}(\theta) = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \tag{7}$$

where Y is the prediction label and X is the truth label.

2.3.6. Comparison with SOTA Models and Experimental Settings

We compared the proposed MSG-GCN with two machine learning methods (RF [88] and SVM [25]) and five SOTA methods (U-Net [41], U-Net++ [49], fully convolutional networks (FCN) [89], vision transformer (ViT) [90], and GCN [78]). The methods were implemented using their original codes and trained on the datasets used herein. The MSG-

GCN module had two graph convolutional layers. ResNet101 [91] was used to create the initial feature matrices. AdamW [92] served as the optimiser; the learning rate was 1×10^{-4}, the weight decay was 1×10^{-3}, and the learning rate was adjusted by cosine annealing warm restarts. All experiments involved 100 training epochs. The batch size was 33. All models were implemented using the PyTorch 1.7.1 framework and the experiments were conducted on the Wisteria/BDEC-01 Supercomputer System (FUJITSU Server PRIMERGY GX2570 M6 (FUJITSU, Tokyo, Japan)) at the Information Technology Center (The University of Tokyo, Tokyo, Japan), which is equipped with NVIDIA A100 Tensor Core (NVIDIA, Santa Clara, CA, USA) graphical processing unit (GPU) (40 GB SXM). For the RF and SVM models, the OpenCV 3.4.2 [93], scikit-learn 0.24.2, and scikit-image 0.18.1 [94] Python 3.7.0 libraries were used to implement machine learning. Using the methods of Canny, Robert, Sobel, Scharr, and Prewitt, edge-based features were extracted as grayscale images that identified and highlighted edge information [95]. Gaussian, median, and variance filters were applied to extract noise and to reduce blurring [96].

2.4. Data Analysis

2.4.1. Accuracy and Complexity Evaluation

To identify the best model, we calculated the overall accuracy (OA), mean intersection over union (mIoU), kappa, F1-score, precision (Pre), and recall (Rec) metrics and evaluated classification accuracy using a confusion matrix. In this matrix, true positives (TPs) are correctly classified pixels, false positives (FPs) are incorrectly classified pixels, true negatives (TNs) are correctly predicted failures, and false negatives (FNs) are incorrectly predicted failures. Pre is the ratio of TPs to the sum of the TPs and FPs and indicates the accuracy within the class. Rec is the ratio of TPs to the sum of TPs and FNs; this is a measure of class confusion. The inclusion of Rec in the F1-score mitigates the imbalance between different types. The OA and kappa value were calculated as percentages of correctly classified pixels [97]. Kappa is more accurate and objective when there is an imbalance in dataset types. Kappa denotes the correlation between human-based validation and that of the machine learning classifier. The kappa value [98] ranges from –1 to 1, where values below 0 indicate no agreement, 0–0.2 indicate slight agreement, 0.2–0.4 indicate fair agreement, 0.4–0.6 indicate agreement, 0.6–0.8 indicate substantial agreement, and 0.8–1 indicate near-perfect or perfect agreement. IoU is the ratio of the overlapping area of the sum of the ground truth and predicted area to the total area and is widely used to assess semantic segmentation [2]. We compared the number of floating-point operations per second (FLOPs) and number of parameters [99] between the MSG-GCN and other SOTA models.

2.4.2. Classification Difference Analysis

The MSG-GCN architecture was refined and optimised based on U-shape encoder and decoder models (U-Net and U-Net++). The MSG-GCN, U-Net, U-Net++, and RF classical machine learning methods were used to map the spatial distributions of correctly classified and misclassified areas. A confusion matrix was applied to the test dataset using the raster calculator of ArcGIS version 10.8 (Esri Inc., Redlands, CA, USA). The results are shown as I–J values, where I is the ground truth and J is the predicted forest type. This provides insight not only into the number of errors but also the type [100]. The percentage is the sum of the correctly classified and misclassified areas divided by the total area of each forest type and makes comparisons within and between forest types more intuitive. A DN is associated with the value of a given pixel for each forest type in the different spectral bands. We used ArcGIS software to extract the DNs for each forest type, which were then summed and divided by the total number of pixels for each classification type. To evaluate differences in mean DNs between classification types, we used Fisher's least significant difference (LSD) test [101] and the 'boxplert' function of R software v 4.2.0 (R Development Core Team) to analyse significant differences between pairs of spectral DNs for correctly classified and misclassified forest types.

The spatial positions of correctly classified and misclassified forests over the entire research site were visualised. The forest type tiles predicted by each model were merged into a 35,152 × 41,152 map and overlaid with the ground truth using the ArcGIS raster calculator. The misclassification rates of multiple buffer zones (10–170 m) around the research site were extracted by ArcGIS [102] to analyse the spatial distribution of the misclassified forests. Given the limitations in data availability, all aerial photographs and composites of the three series were collected. Thus, the effects of seasons, solar radiation and weather conditions were included in the dataset. The areas of misclassified forests from the three districts were extracted to determine how the dataset affected classification.

3. Results

3.1. Classification Accuracy Indices of Different Models

Table 1 shows the quantitative evaluation metrics (OA, kappa, IoU, F1-score, Pre, and Rec) for all classification types, which were calculated using the test dataset to allow for quantitative comparisons. The accuracy metric showed that the MSG-GCN model performed best (kappa = 0.7808, OA = 0.8523), while U-Net++ (kappa = 0.7263, OA = 0.8143) and GCN (kappa = 0.7473, OA = 0.8240) performed well; FCN (kappa = 0.5209, OA = 0.6895), U-Net (kappa = 0.6706, OA = 0.8098), and ViT (kappa = 0.6942, OA = 0.7914) performed moderately. RF (kappa = 0.3373, OA = 0.5739) and SVM (kappa = 0.0028, OA = 0.6344) did not perform as well as the other models. In terms of the classification quantitative metrics for the three forest types, the MSG-GCN model performance was competitive relative to the other methods. Of the three forest types, NMX was the least accurate for all models because of dataset imbalance; few NMX areas were correctly classified. Although GCN had the best accuracy for NMX, the accuracies of NBL and CP were lower than that of MSG-GCN. Of the three forest types, CP was most accurately detected by U-Net, U-Net++, ViT, GCN, and MSG-GCN. RF and SVM detected NBL most accurately, followed by CP. Considering the parameter numbers and FLOPs required to process the images, our MSG-GCN architecture combining GCNs with CNNs has the largest model size (88.10 M parameters) and required more operations for processing and slightly fewer FLOPs (104.99 GMac). The FCN has the fewest parameters (3.93 M) and U-Net is optimal in terms of FLOPs (218.94 GMac).

3.2. Area and Digital Number for Each Forest Type Predicted by the Various Models

To further evaluate classification, the confusion matrix was applied to the test dataset. All evaluation metrics (Table 1) showed the same tendencies in terms of the percentages of correctly classified forest areas of each type (Table 2). NMX and CP forests tended to be more susceptible to be classified as NBL forest, whereas the misclassification rates between NMX and CP were very low. Taking MSG-GCN as an example, NMX was more likely to be classified as NBL than CP, while CP was more likely to be classified as NBL than NMX. U-Net and U-Net++ showed the same tendencies as MSG-GCN. There were very few misclassifications of BG points as other forest types; such errors were confined to the edges because of the inevitable errors in raster pixel values. The RF model could not detect NMX forest and tended to misclassify NMX as CP rather than NBL.

Figure 4 shows that NMX, NBL, and CP forests correctly classified by the four models had very similar mean DN values for the three RGB bands. LSD analysis revealed that the mean DN of NBL was higher than that for CP (Figures 4 and 5) and that the DNs for correctly classified and misclassified forest types were very similar. The distributions of the correctly classified and misclassified forest types overlapped greatly in terms of the DNs, as shown in the plot figures (Figure 4).

Table 1. Performance comparison of the seven methods. Accuracy, parameters, and FLOPs are listed as a function of model size on the aerial photo. The best results are in bold.

Models	OA	Kappa	IoU_NMX	IoU_NBL	IoU_CP	F1_NMX	F1_NBL	F1_CP	Pre_NMX	Pre_NBL	Pre_CP	Rec_NMX	Rec_NBL	Rec_CP	FLOPs (GMac)	Params(M)
RF	0.5739	0.3373	0.0596	0.4605	0.1584	0.1124	0.6306	0.2734	0.1017	0.6964	0.2307	0.1257	0.5762	0.3357	-	-
SVM	0.6344	0.0028	0.1245	0.2324	0.1497	0.2214	0.3772	0.2605	0.3416	0.2938	0.3659	0.1637	0.5264	0.2022	-	-
U-Net	0.8098	0.6706	0.344	0.6679	0.6776	0.5119	0.8009	0.8078	0.5151	0.7758	0.8504	0.5088	0.8277	0.7693	**218.94**	31.04
U-Net++	0.8143	0.7263	0.3367	0.6741	0.6992	0.5037	0.8054	0.8229	0.5221	0.7919	0.8358	0.4867	0.8192	0.8105	153.00	47.18
FCN	0.6895	0.5209	0.0007	0.5531	0.4367	0.0014	0.7123	0.6079	0.3857	0.6149	0.6561	0.0007	0.8462	0.5664	102.19	**3.93**
ViT	0.7914	0.6942	0.3499	0.6454	0.6530	0.5184	0.7845	0.7900	0.5913	0.7475	0.8189	0.4615	0.8254	0.7632	22.66	23.28
GCN	0.8240	0.7473	**0.5463**	0.6651	0.6811	**0.7066**	0.7989	0.8103	**0.7232**	0.8154	0.7781	**0.6907**	0.7830	0.8453	57.66	9.18
MSG-GCN	**0.8523**	**0.7808**	0.4374	**0.7341**	**0.7451**	0.6086	**0.8467**	**0.8539**	0.6103	**0.8475**	**0.8510**	0.6069	**0.8459**	**0.8569**	104.99	88.10

Table 2. Areas and percentages of forest types correctly classified and misclassified by the RF, U-Net, U-Net++, and MSG-GCN models.

Classification	Ground Truth		MSG-GCN		U-Net++		U-Net		RF	
	Number of Pixels	Percentage (%)	Number of Pixels	Percentage (%)	Number of Pixels	Percentage (%)	Number of Pixels	Percentage (%)	Number of Pixels	Percentage (%)
BG	11,654,816	100	**11,624,871**	**99.74**	**11,651,030**	**99.97**	**11,650,029**	**99.96**	**11,645,984**	**99.92**
BG-NMX			1138	0.01	343	0	55	0	665	0.01
BG-NBL			18,005	0.16	1597	0.01	2461	0.02	5039	0.04
BG-CP			10,802	0.09	1846	0.02	2271	0.02	3128	0.03
NMX	6,282,528	100	**3,833,885**	**61.02**	**3,279,871**	**52.21**	**3,235,934**	**51.51**	**638,718**	**10.17**
NMX-BG			3691	0.06	2741	0.04	5495	0.09	4416	0.07
NMX-NBL			2,342,218	37.28	2,742,520	43.65	2,633,321	41.91	4,320,615	68.77
NMX-CP			102,734	1.64	257,396	4.10	407,778	6.49	1,318,779	20.99
NBL	30,431,072	100	**25,789,099**	**84.75**	**24,098,604**	**79.19**	**23,609,032**	**77.58**	**21,192,944**	**69.64**
NBL-BG			37,705	0.12	30,617	0.10	51,540	0.17	42,704	0.14
NBL-NMX			2,351,894	7.73	3,308,442	10.87	2,935,850	9.65	2,920,111	9.60
NBL-CP			2,252,374	7.40	2,993,409	9.84	3,834,650	12.60	6,275,313	20.62
CP	16,643,296	100	**14,163,402**	**85.1**	**13,911,171**	**83.58**	**14,153,704**	**85.04**	**3,838,865**	**23.07**
CP-BG			13,638	0.08	9119	0.05	22,016	0.13	17,659	0.10
CP-NMX			129,378	0.78	150,747	0.91	188,197	1.13	1,521,940	9.15
CP-NBL			2,336,878	14.04	2,572,259	15.46	2,279,379	13.70	11,264,832	67.68

BG indicates background areas lacking spectral information; BG-NMX indicates BG areas misclassified as NMX forests; BG-NBL indicates BG areas misclassified as NBL forests; BG-CP indicates BG areas misclassified as CP forests; NMX-BG indicates NMX forests misclassified as BG areas; NMX-NBL indicates NMX forests misclassified as NBL forests; NMX-CP indicates NMX forests misclassified as CP forests; NBL-BG indicates NBL forests misclassified as BG areas; NBL-NMX indicates NBL forests misclassified as NMX forests; NBL-CP indicates NBL forests misclassified as CP forests; CP-BG indicates CP forests misclassified as BG areas; CP-NMX indicates CP forests misclassified as NMX forests; CP-NBL indicates CP forests misclassified as NBL forests.

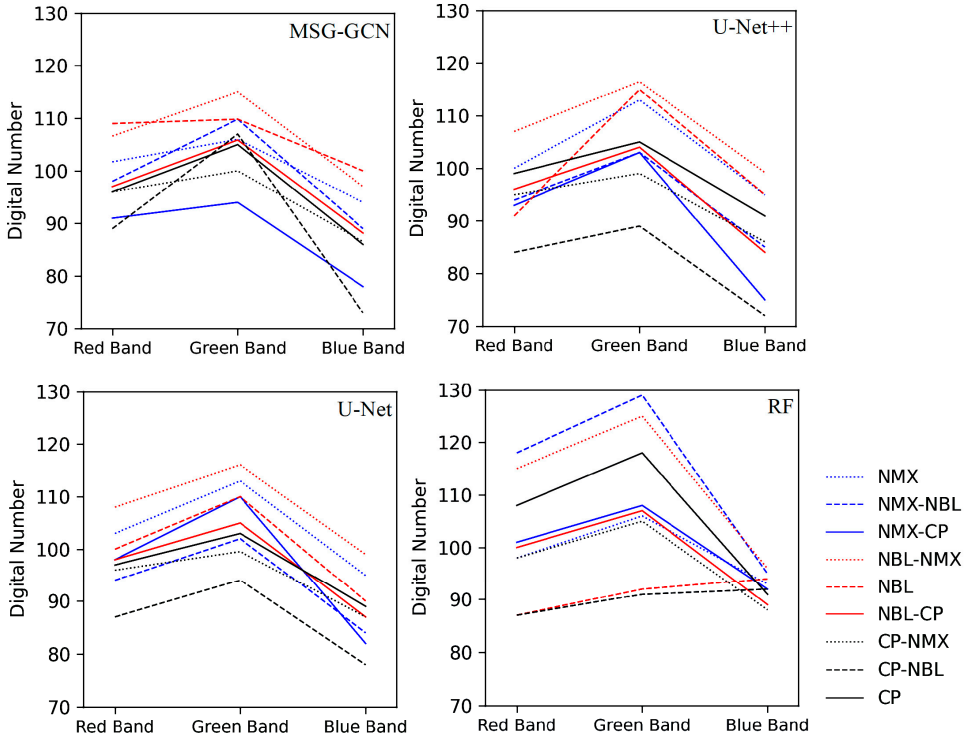

Figure 4. Mean digital numbers (DNs) of forest types correctly classified and misclassified by different models.

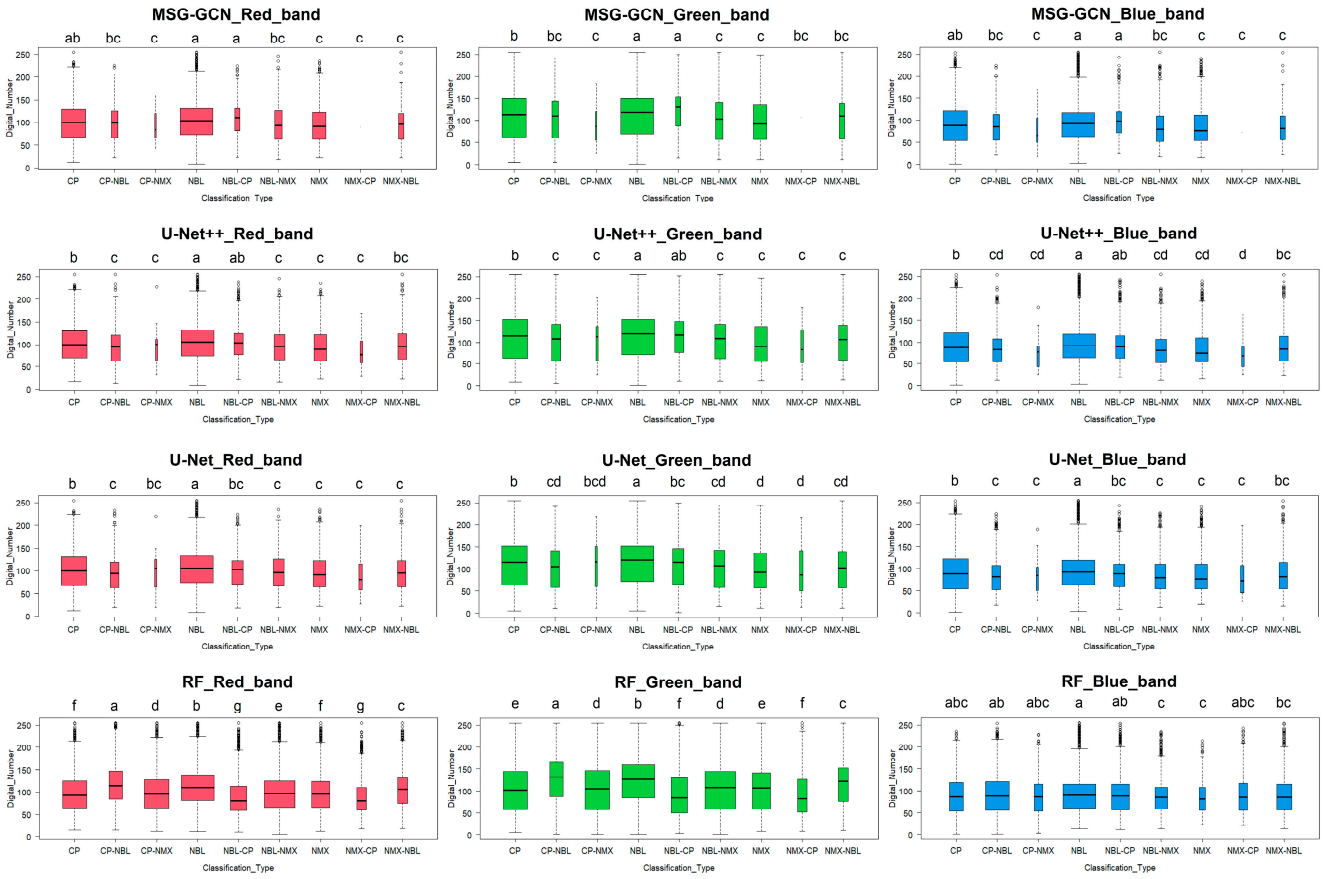

Figure 5. Least significant difference (LSD) analysis of the digital numbers (DNs) of the various bands for different forest types and different models.

The differences between the DNs of forest types correctly classified and misclassified by the four models were analysed using the LSD test. In terms of the DNs of correctly classified forest types, NMX(c) and NBL(a) differed significantly. In terms of the DNs of misclassified forest types, taking the red band of the MSG-GCN model as an example (Figure 5), a significant difference between NBL(a) and NBL-NMX(bc) was shown, while there was no significant difference between NBL-NMX(bc) and NMX(c). There was no significant difference between CP(ab) and CP-NBL(bc), but there was a difference between CP-NBL(bc) and NBL(a). There was no significant difference between NMX(c) and NMX-NBL(c), but there was a difference between NMX-NBL(c) and NBL(a).

In terms of the prediction maps (Figure 6) for the test data set, severe salt-and-pepper noise was apparent in the RF map, but this was less severe in the U-Net and U-Net++ maps. However, there were patches where NBL and NMX areas were misclassified as CP or confused with each other. Moreover, the forest type edges were unclear, especially for NMX in U-Net and U-Net++, whereas the MSG-GCN map showed fewer misclassified patches and the edge areas between different forest types were sharper (Figure 6).

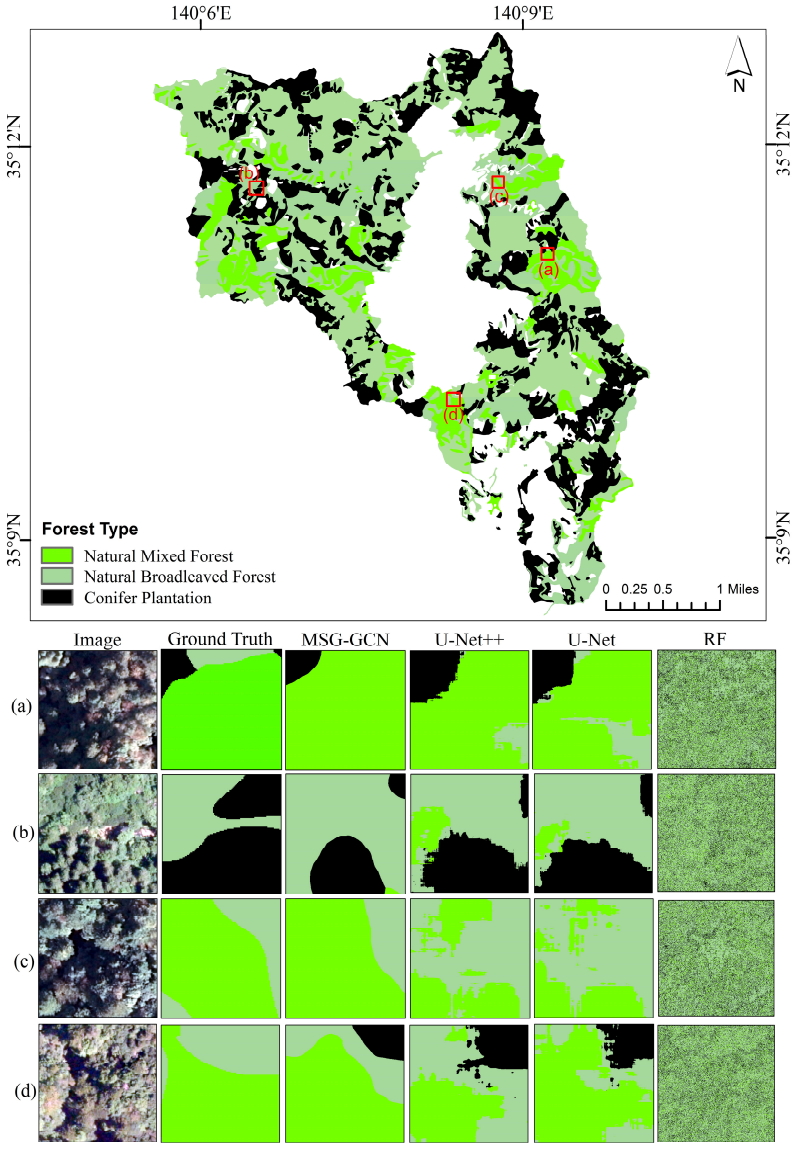

Figure 6. Ground truths of the three forest types. Example of the limitations of spectral classification (i.e., RF) and the U-Net, U-Net++, and MSG-GCN spectral–spatial classification methods. (a) a patch tile contains NMX, NBL and CP forests, (b) a patch tile contains NBL and CP forests, (c) a patch tile contains NMX and NBL forests, (d) a patch tile contains NMX and NBL forests.

3.3. Mapping and Spatial Distributions of Forest Types by the Different Models

For MSG-GCN, misclassified (Figures 7 and 8) areas were mostly distributed in boundary areas. For U-Net and U-Net++, the misclassified areas were not confined to the boundaries; instead, they were randomly distributed in internal regions. Buffer zone analysis (Figures 8 and 9) showed that MSG-GCN errors were mostly in the external 30-m boundary buffer areas. The error rate decreased rapidly from the internal 40-m buffer zones and plateaued at 3.26% for the 50-m buffer zones. For U-Net and U-Net++, the errors were concentrated in the external 30-m boundary buffer zones; the internal buffer zones had lower error rates. However, because of the randomly misclassified fragments, the internal error rates were larger than for MSG-GCN.

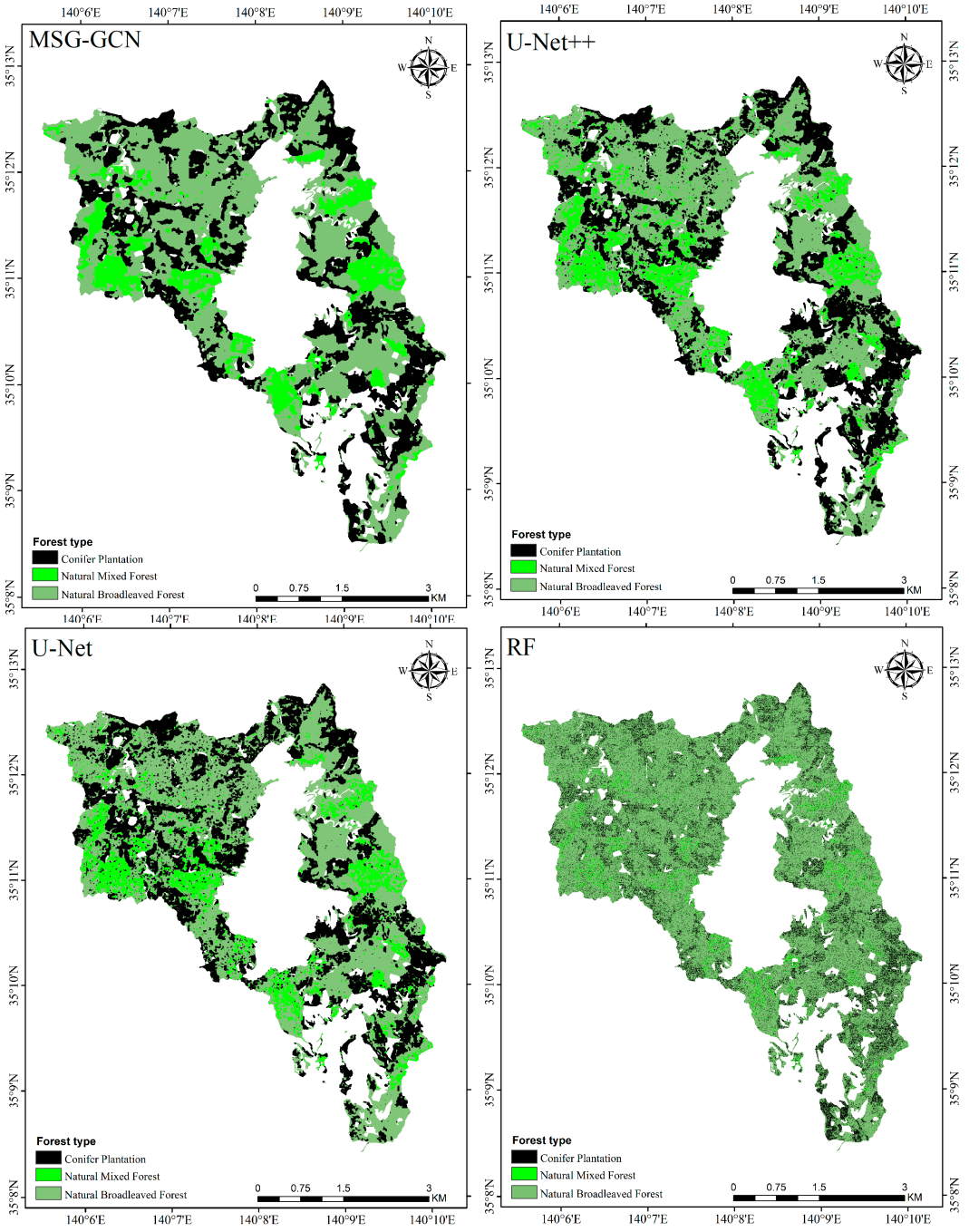

Figure 7. Visualisation of the entire study area.

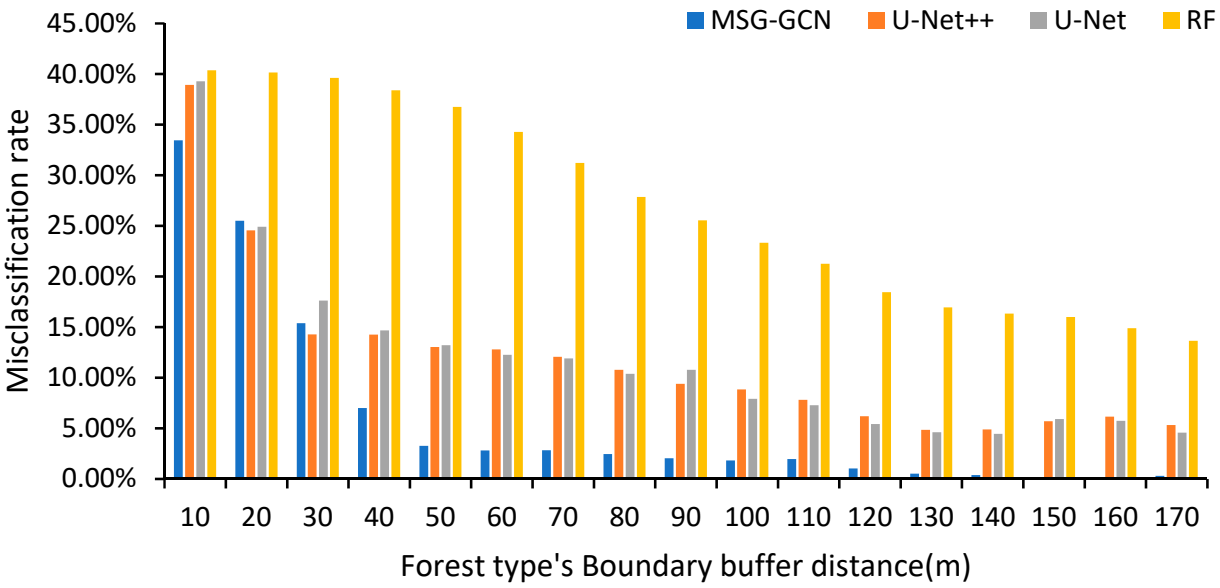

Figure 8. Variation in the misclassification rates of different models over buffer scales of 10–170 m.

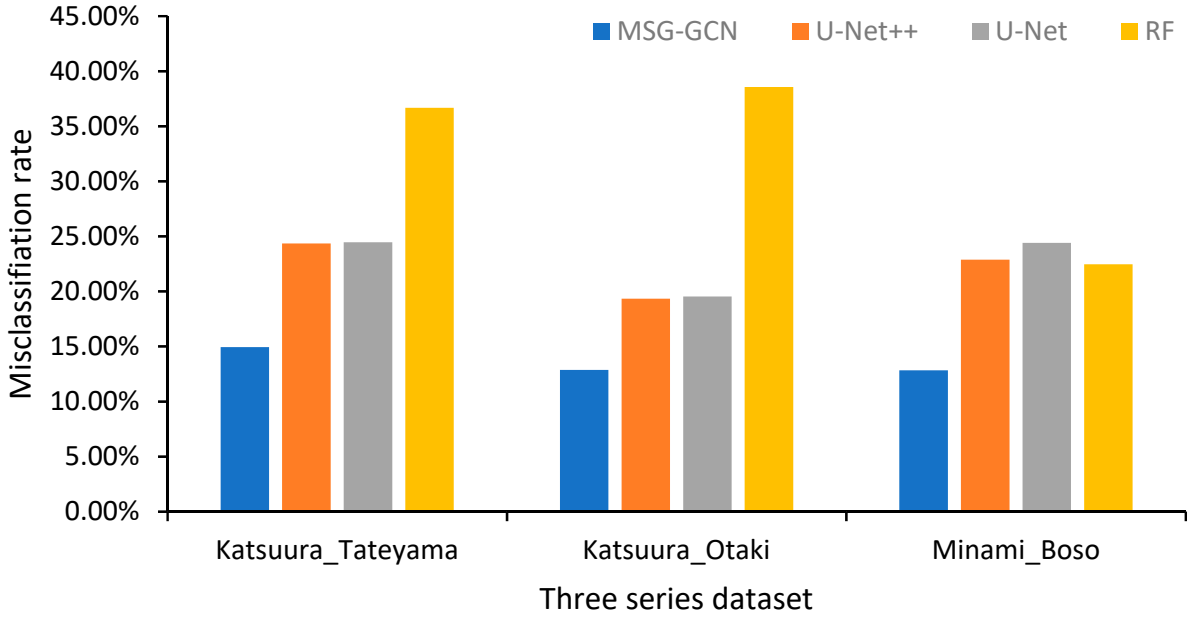

Figure 9. Misclassification rates for the different districts of all models.

The misclassification rates of the three district images showed that year and season did not greatly influence classification accuracy (Figure 9). The Katsuura Tateyama and Minami Boso areas showed slightly higher error rates than the Katsuura Otaki area with U-Net, U-Net++, and RF models; for the MSG-GCN model, the error rate for Katsuura Tateyama was higher than those for Katsuura Otaki and Minami Boso. The RF, U-Net, and U-Net++ made more errors than MSG-GCN (Figures 9 and 10).

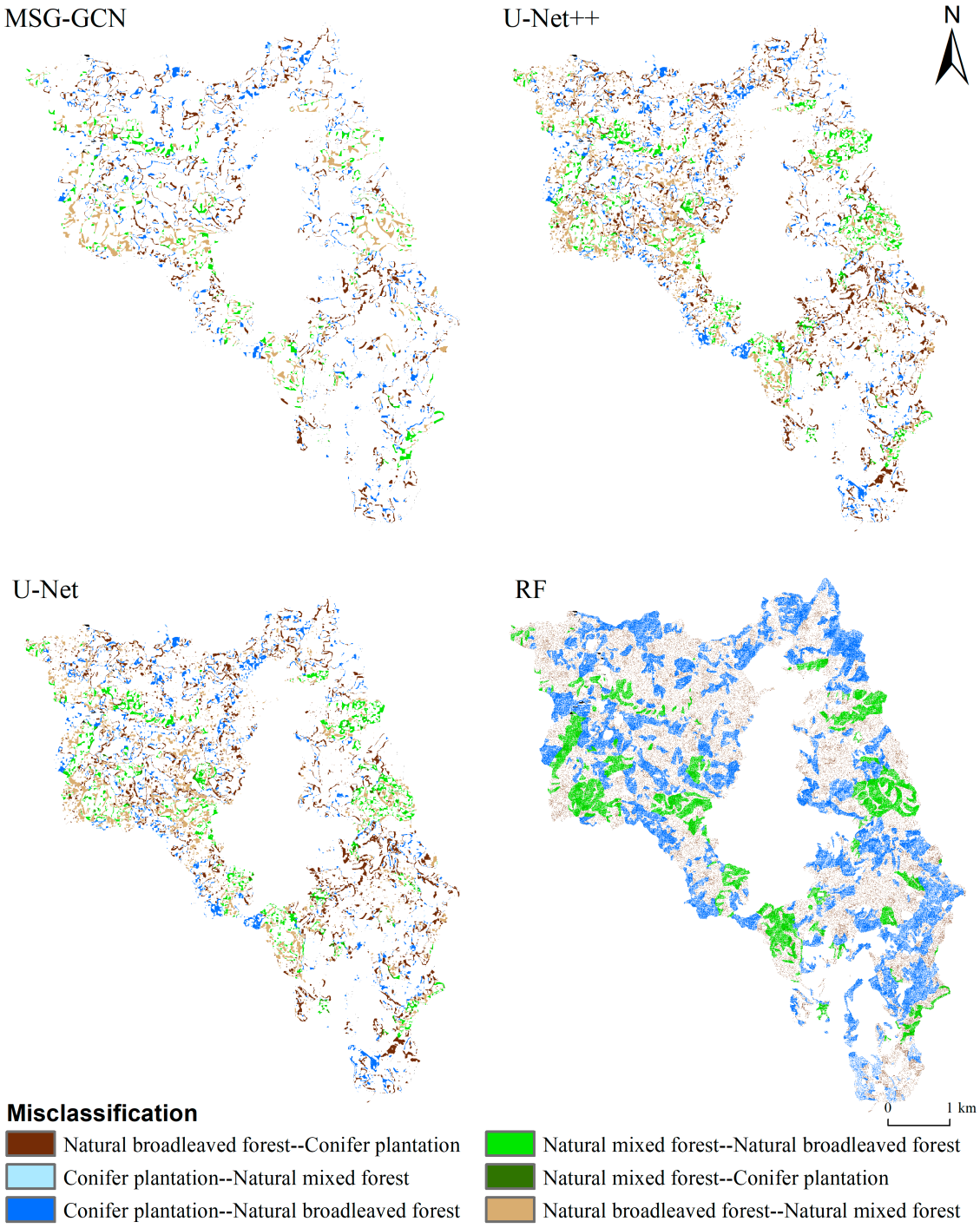

Figure 10. The forest types misclassified by the various models.

4. Discussion

4.1. Classification Accuracy

The results (Tables 1 and 2) showed that MSG-GCN outperformed the other models. MSG-GCN better extracted the features of minority classes. The IoU accuracy for NMX was 0.4374, which was higher than those of U-Net++ (0.3367) and U-Net (0.3440), and shows the utility of augmentation techniques, including the cropping, rotating, and flipping (of input images), as well as the Dice loss function [103]. The results are in line with previous studies; the classification accuracy for NMX was lower than those for CP or NBL [10]. Cheng et al. [32] provided a relatively reliable planted forest (OA = 84.93 %,

F1 = 0.85) classification at a national scale using multisource remote sensing data, while the poor data quality and complex terrain and vegetation conditions brought errors in planted and natural forest classification. In CP areas, seedlings are usually planted in lines, such that high classification accuracy is possible. However, if the lines are not evident because some trees have died or been harvested or broadleaved trees have naturally regenerated, there is a tendency toward error [2]. The borders of NMX areas are not sharp and the manual and automatic classification results varied. Although visual classification can be considered accurate, errors may exist due to inaccurate manual delineation and objective variance [3]. NMX areas with heterogeneous and irregular boundaries, diverse species, and multicanopy layers were the most difficult to map. The edges were not smooth and the canopy textural features were heterogeneous. This severely imbalanced the forest types in terms of the relative proportions of the NMX, NBL, and CP areas, which complicated classification because the dataset was biased toward the majority class; this was associated with errors among the minority classes [104]. For the SOTA experiments, FCN [89] can effectively capture the local feature information of the target, while it cannot obtain the global information due to the small receptive field. A shallow GCN [78] with a single structure cannot effectively spread label information in a large area. During the transmission of node information, it is easy to cause the node feature vector to be oversmoothed and result in similar characteristics of each node, which is not conducive to the final segmentation performance. ViT [90] regards each patch as a task by extracting and refining relational information through a multihead self-attention mechanism, while the model requires for heavy computation capacity and high-quality training data. In addition, the oversmoothing problem limited the number of stacking layers and prevented the model from encoding the position information, which ultimately degraded the segmentation performance. The computation complexity performance of MSG-GCN is not the best among the SOTA experiments, our main aim is to improve classification accuracy while ensuring moderate FLOPs. In the future, we will try to build concise efficient graph semantic models to improve the accuracy of forest type classification.

4.2. Area and Digital Number for Each Forest Type Predicted by the Different Models

In terms of forest type misclassification, our model (Table 2) made fewer errors than the U-Net, U-Net++, and RF models. A drawback of U-Net and U-Net++ is the use of identical kernel weights, which can cause loss of semantic information during information transfer between layers [105]. Unlike previous studies that used fixed graphs, we imported a multiscale graph, which can perform flexible convolution on any irregular image area and describe the target object in the image at different scales, increase the receptive field, and improve the feature representation ability. The MSG-GCN can exploit multiscale information, remove noise, and preserve edges [85]. We found that MSG-GCN segmented clear edges. The GCN uses edge information to aggregate node information and then generate a new node representation that automatically learns both node features and associations between nodes [106]. This allows the model to learn the characteristic information of boundary pixels better, establishes spatial correlations, effectively resolves differences within or between classes, and improves the classification accuracy of objects in aerial forest images. Although the segment edges of MSG-GCN were clearer than those of U-Net and U-Net++, some misclassification (Figure 8) still occurred in edge areas, in line with previous research [107,108]. In an NMX area, broadleaved species covered 25–75% of the land occupied by coniferous canopies [66]. Some tree species are found in more than one forest type. In UTCBF areas, the main tree species are evergreen *Quercus* spp., but *C. siedoldii,* which is found in NBL areas, can also be seen in the NMX forest type; this may explain the misclassification of NBL as NMX. Spectral similarities, crown overlap, and similar noise among different forest types compromise classification accuracy [109], leading to misclassification. The models find it difficult to delineate forest types based on the surface, colour, and patterns. Thus, we analysed the spectral values of correctly classified and misclassified areas to understand why misclassification occurred. The DNs

(Figure 4) of different forest types varied, but the DNs of misclassified forest types were similar to those of correctly classified areas [107]. The DN of NBL was relatively high compared with that of CP, in line with previous results [28,110]; the forward-scattering direction of a coniferous forest shows lower reflectance than that of a deciduous forest because of the distinct optimal angles and leaf directions. An example LSD analysis of the digital red band of MSG-GCN showed that, within the NMX group, there was no significant difference between NMX(c) and NMX-NBL(c). Although a significant group difference between NMX-NBL(c) and NBL(a) was apparent, misclassification remained an issue. There was no significant difference between CP(ab) and CP-NBL(bc) within the CP group, but MSG-GCN showed a significant difference between CP-NBL(bc) and NBL(a). However, some CP-NBL misclassification still occurred, perhaps because the spectral information was not adequate to allow the model to classify the forest types. Other factors may also trigger misclassification. There were only three information bands and the spectral reflectance varied among the bands for different forest types. Within-species variation in reflectance may be caused by site conditions, species composition, the vertical structure, and shadowing effects [10,111].

4.3. Mapping and Spatial Distributions of Forest Types for the Different Models

As shown in the visualisation map (Figure 7), RF showed severe salt-and-pepper noise. Object-based classification by U-Net and U-Net++ reduced this noise, but many patches remained where NBF was misclassified as CP and the edges were not sharp and clear (unlike MSG-GCN) (Figures 7 and 10). With MSG-GCN, misclassified forests clustered mainly along the transition zones of the forest boundary areas (Figure 8), similar to previous findings [30,73], perhaps because of homogeneity in the spectral responses and shade due to the highly enclosed overlapping crowns [2], as well as the greater sensitivity of edges compared with internal areas. This is a longstanding problem in semantic segmentation. Unlike the regular morphology of well-defined urban land, the morphology and texture of forest canopies are heterogeneous and complex and vary according to species composition [66]. Thus, ground truths may not be recorded accurately given the inevitable subjective boundary errors and variation in remote sensing images caused by weather or the sensor type used. Weight or edge masks could be assigned to boundary area pixels [112] to effectively mine edge and neighbourhood information. Apart from edge misclassification, U-Net and U-Net++ misclassified internal forest areas (i.e., NBF as CP; Figure 8), perhaps because of the similar spectral values and crown sizes of certain species. However, MSG-GCN rarely made such misclassifications. Node transfer and aggregation yielded multiscale graphical information that highlighted both intra- and interclass differences, thereby improving the recognition of different categories [106].

5. Conclusions

In this paper, we developed a novel MSG-GCN model that uses a combination of multiscale convolutional kernels, a MSGCN module, LA, and output features from different decoding blocks to extract both high- and low-level features. To our knowledge, this is the first application of multiscale graph convolutions to forest type classification with aerial photos. Our results show that MSG-GCN is useful for the segmentation task. The main contributions of this study are as follows: First, multiscale convolutional kernels were used to learn features from different receptive field scales for forest type classification using aerial photos. Unlike the traditional fixed square area convolution, this method successfully learned the correlations between adjacent pixels in an irregular area with a multiscale graph convolutional kernel filter. Second, LA was used to refine the features and to highlight the representation ability of salient features. In the stage of high-level feature representation, excessive interference from low-level features with the representation of high-level features (and vice versa) is avoided. Finally, the MSGCN module should resolve the incompatibility between convolution and graph convolution in the data structure, which has an excessive influence on encoding and decoding features. Moreover, CNN (encoding and decoding

modules) and GCN (multiscale graph convolution) are used to perform feature learning on small-scale regular areas and large-scale irregular areas, respectively, which aids decision making regarding boundary pixels.

We also found that NMXs and CPs were more susceptible to misclassification than NBLs. Classification of forest types using only the similar and overlapping spectral DNs is not sufficient. The visualisation map of the entire area revealed that edge pixels were more likely to be misclassified as neighbourhood pixels by all networks and that the CNN approaches were associated with random misclassification patches in internal zones. NMF was more challenging to classify than other forest types, given the imbalances in datasets, the heterogeneous canopy texture, and the fact that edge regions are evident only in very high-resolution aerial RGB images.

In future research, we will aim to combine a multisensor remote sensing dataset and a multimodal machine learning model [113–116] to enhance the multidisciplinary nature of remote sensing images and deep-learning technology, to overcome the remaining problems with the MSG-GCNs method (such as how to make full use of multimodal [117] data to aid segmentation of edge pixels). Simultaneously, multisensor data will be applied to build concise efficient graph semantic models to improve the accuracy of forest type classification.

Author Contributions: Conceptualization, methodology, formal analysis, and writing (original draft preparation), H.P.; resources, supervision, and writing (review and editing), T.O. and S.T.; writing (methodology review and editing), Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number [JP18K05742], Hainan Provincial Natural Science Foundation, No.322MS019. This research is partially supported by Initiative on Recommendation Program for Young Researchers and Woman Researchers [2022-I-004], Information Technology Center, The University of Tokyo.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank the technical staff from the UTCBF for their contribution to the forest type interpretation data collection, and we would like to thank Keisuke Toyama and Takuya Hiroshima of the University of Tokyo for their kind help through the provision of professional information and comments about this paper. The authors wish to thank the comments from Hong Danfeng of the Aerospace Information Research Institute, Chinese Academy of Sciences and Song Xiqiang of College of Forestry, Hainan University for improvements in this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thompson, I.D.; Baker, J.A.; Ter-Mikaelian, M. A review of the long-term effects of post-harvest silviculture on vertebrate wildlife, and predictive models, with an emphasis on boreal forests in Ontario, Canada. *For. Ecol. Manag.* **2003**, *177*, 441–469. [CrossRef]
2. Wagner, F.H.; Sanchez, A.; Tarabalka, Y.; Lotte, R.G.; Ferreira, M.P.; Aidar, M.P.M.; Gloor, E.; Phillips, O.L.; Aragão, L.E.O.C. Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. *Remote Sens. Ecol. Conserv.* **2019**, *5*, 360–375. [CrossRef]
3. Kislov, D.E.; Korznikov, K.A.; Altman, J.; Vozmishcheva, A.S.; Krestov, P.V. Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images. *Remote Sens. Ecol. Conserv.* **2021**, *7*, 355–368. [CrossRef]
4. Muhammad, K.; Ahmad, J.; Baik, S.W. Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing* **2018**, *288*, 30–42. [CrossRef]
5. Zhao, F.; Sun, R.; Zhong, L.; Meng, R.; Huang, C.; Zeng, X.; Wang, M.; Li, Y.; Wang, Z. Monthly mapping of forest harvesting using dense time series Sentinel-1 SAR imagery and deep learning. *Remote Sens. Environ.* **2022**, *269*, 112822. [CrossRef]
6. Pandit, S.; Tsuyuki, S.; Dube, T. Landscape-scale aboveground biomass estimation in buffer zone community forests of Central Nepal: Coupling in situ measurements with Landsat 8 Satellite Data. *Remote Sens.* **2018**, *10*, 1848. [CrossRef]
7. Jayathunga, S.; Owari, T.; Tsuyuki, S. The use of fixed–wing UAV photogrammetry with LiDAR DTM to estimate merchantable volume and carbon stock in living biomass over a mixed conifer–broadleaf forest. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 767–777. [CrossRef]
8. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N. Prabhat Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [CrossRef]

9. Yang, R.; Wang, L.; Tian, Q.; Xu, N.; Yang, Y. Estimation of the conifer-broadleaf ratio in mixed forests based on time-series data. *Remote Sens.* **2021**, *13*, 4426. [CrossRef]
10. Ohsawa, T.; Saito, Y.; Sawada, H.; Ide, Y. Impact of altitude and topography on the genetic diversity of Quercus serrata populations in the Chichibu Mountains, central Japan. *Flora Morphol. Distrib. Funct. Ecol. Plants* **2008**, *203*, 187–196. [CrossRef]
11. Pfeifer, M.; Lefebvre, V.; Peres, C.A.; Banks-Leite, C.; Wearn, O.R.; Marsh, C.J.; Butchart, S.H.M.; Arroyo-Rodríguez, V.; Barlow, J.; Cerezo, A.; et al. Creation of forest edges has a global impact on forest vertebrates. *Nature* **2017**, *551*, 187–191. [CrossRef]
12. Bonan, G.B.; Pollard, D.; Thompson, S.L. Effects of boreal forest vegetation on global climate. *Nature* **1992**, *359*, 716–718. [CrossRef]
13. Raft, A.; Oliier, H. Forest restoration, biodiversity and ecosystem functioning. *BMC Ecol.* **2011**, *11*, 29. [CrossRef]
14. Rozendaal, D.M.A.; Requena Suarez, D.; De Sy, V.; Avitabile, V.; Carter, S.; Adou Yao, C.Y.; Alvarez-Davila, E.; Anderson-Teixeira, K.; Araujo-Murakami, A.; Arroyo, L.; et al. Aboveground forest biomass varies across continents, ecological zones and successional stages: Refined IPCC default values for tropical and subtropical forests. *Environ. Res. Lett.* **2022**, *17*, 014047. [CrossRef]
15. Thurner, M.; Beer, C.; Santoro, M.; Carvalhais, N.; Wutzler, T.; Schepaschenko, D.; Shvidenko, A.; Kompter, E.; Ahrens, B.; Levick, S.R.; et al. Carbon stock and density of northern boreal and temperate forests. *Glob. Ecol. Biogeogr.* **2014**, *23*, 297–310. [CrossRef]
16. Coppin, P.R.; Bauer, M.E. Digital Change Detection in Forest Ecosystems with Remote Sensing Imagery. *Remote Sens. Rev.* **1996**, *13*, 207–234. [CrossRef]
17. Cowardin, L.M.; Myers, V.I. Remote Sensing for Identification and Classification of Wetland Vegetation. *J. Wildl. Manag.* **1974**, *38*, 308–314. [CrossRef]
18. Schiefer, F.; Kattenborn, T.; Frick, A.; Frey, J.; Schall, P.; Koch, B.; Schmidtlein, S. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2020**, *170*, 205–215. [CrossRef]
19. Kentsch, S.; Karatsiolis, S.; Kamilaris, A.; Tomhave, L.; Lopez Caceres, M.L. Identification of Tree Species in Japanese Forests based on Aerial Photography and Deep Learning. *arXiv* **2020**. [CrossRef]
20. Komárek, J. The perspective of unmanned aerial systems in forest management: Do we really need such details? *Appl. Veg. Sci.* **2020**, *23*, 718–721. [CrossRef]
21. Ray, R.G. *Aerial Photographs in Geologic Interpretation and Mapping*; Professional Paper; US Government Printing Office: Washington, DC, USA, 1960. [CrossRef]
22. Ozaki, K.; Ohsawa, M. Successional change of forest pattern along topographical gradients in warm-temperate mixed forests in Mt Kiyosumi, central Japan. *Ecol. Res.* **1995**, *10*, 223–234. [CrossRef]
23. Chianucci, F.; Disperati, L.; Guzzi, D.; Bianchini, D.; Nardino, V.; Lastri, C.; Rindinella, A.; Corona, P. Estimation of canopy attributes in beech forests using true colour digital images from a small fixed-wing UAV. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *47*, 60–68. [CrossRef]
24. Bagaram, M.B.; Giuliarelli, D.; Chirici, G.; Giannetti, F.; Barbati, A. UAV remote sensing for biodiversity monitoring: Are forest canopy gaps good covariates? *Remote Sens.* **2018**, *10*, 1397. [CrossRef]
25. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [CrossRef]
26. Heydari, S.S.; Mountrakis, G. Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites. *Remote Sens. Environ.* **2018**, *204*, 648–658. [CrossRef]
27. Dostálová, A.; Wagner, W.; Milenković, M.; Hollaus, M. Annual seasonality in Sentinel-1 signal for forest mapping and forest type classification. *Int. J. Remote Sens.* **2018**, *39*, 7738–7760. [CrossRef]
28. Liu, Y.; Gong, W.; Hu, X.; Gong, J. Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sens.* **2018**, *10*, 946. [CrossRef]
29. Griffiths, P.; Kuemmerle, T.; Baumann, M.; Radeloff, V.C.; Abrudan, I.V.; Lieskovsky, J.; Munteanu, C.; Ostapowicz, K.; Hostert, P. Forest disturbances, forest recovery, and changes in forest types across the carpathian ecoregion from 1985 to 2010 based on landsat image composites. *Remote Sens. Environ.* **2014**, *151*, 72–88. [CrossRef]
30. Lapini, A.; Pettinato, S.; Santi, E.; Paloscia, S.; Fontanelli, G.; Garzelli, A. Comparison of machine learning methods applied to SAR images for forest classification in mediterranean areas. *Remote Sens.* **2020**, *12*, 369. [CrossRef]
31. Pasquarella, V.J.; Holden, C.E.; Woodcock, C.E. Improved mapping of forest type using spectral-temporal Landsat features. *Remote Sens. Environ.* **2018**, *210*, 193–207. [CrossRef]
32. Cheng, K.; Su, Y.; Guan, H.; Tao, S.; Ren, Y.; Hu, T. Mapping China's planted forests using high resolution imagery and massive amounts of crowdsourced samples. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 356–371. [CrossRef]
33. Kuppusamy, P.; Ieee, M. Retinal Blood Vessel Segmentation using Random Forest with Gabor and Canny Edge Features. In Proceedings of the 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 5–26 March 2022; pp. 1–4. [CrossRef]
34. Yoo, C.; Han, D.; Im, J.; Bechtel, B. Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images. *ISPRS J. Photogramm. Remote Sens.* **2019**, *157*, 155–170. [CrossRef]

35. Kattenborn, T.; Leitloff, J.; Schiefer, F.; Hinz, S. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 24–49. [CrossRef]
36. Zhou, Z.; Siddiquee, M.R.; Tajbakhsh, N.; Liang, J. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans. Med. Imaging* **2020**, *39*, 1856–1867. [CrossRef] [PubMed]
37. Ferreira, M.P.; de Almeida, D.R.A.; de Almeida Papa, D.; Minervino, J.B.S.; Veras, H.F.P.; Formighieri, A.; Santos, C.A.N.; Ferreira, M.A.D.; Figueiredo, E.O.; Ferreira, E.J.L. Individual tree detection and species classification of Amazonian palms using UAV images and deep learning. *For. Ecol. Manag.* **2020**, *475*, 118397. [CrossRef]
38. Pyo, J.C.; Han, K.J.; Cho, Y.; Kim, D.; Jin, D. Generalization of U-Net Semantic Segmentation for Forest Change Detection in South Korea Using Airborne Imagery. *Forests* **2022**, *13*, 2170. [CrossRef]
39. Fu, C.; Song, X.; Xie, Y.; Wang, C.; Luo, J.; Fang, Y.; Cao, B.; Qiu, Z. Research on the Spatiotemporal Evolution of Mangrove Forests in the Hainan Island from 1991 to 2021 Based on SVM and Res-UNet Algorithms. *Remote Sens.* **2022**, *14*, 5554. [CrossRef]
40. Li, L.; Mu, X.; Chianucci, F.; Qi, J.; Jiang, J.; Zhou, J.; Chen, L.; Huang, H.; Yan, G.; Liu, S. Ultrahigh-resolution boreal forest canopy mapping: Combining UAV imagery and photogrammetric point clouds in a deep-learning-based approach. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *107*, 102686. [CrossRef]
41. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; Volume 9351, pp. 234–241. [CrossRef]
42. Liu, Y.; Zhong, Y.; Qin, Q. Scene classification based on multiscale convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 7109–7121. [CrossRef]
43. Zhou, W.; Jin, J.; Lei, J.; Yu, L. CIMFNet: Cross-Layer Interaction and Multiscale Fusion Network for Semantic Segmentation of High-Resolution Remote Sensing Images. *IEEE J. Sel. Top. Signal Process.* **2022**, *16*, 666–676. [CrossRef]
44. Zhao, W.; Du, S. Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *113*, 155–165. [CrossRef]
45. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]
46. Liu, Q.; Hang, R.; Song, H.; Li, Z. Learning multiscale deep features for high-resolution satellite image scene classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 117–126. [CrossRef]
47. Wang, Q.; Member, S.; Liu, S.; Chanussot, J. Scene Classification with Recurrent Attention of VHR Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1155–1167. [CrossRef]
48. Bi, Q.; Qin, K.; Li, Z.; Zhang, H.; Xu, K.; Xia, G.S. A Multiple-Instance Densely-Connected ConvNet for Aerial Scene Classification. *IEEE Trans. Image Process.* **2020**, *29*, 4911–4926. [CrossRef] [PubMed]
49. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018*; Springer: Cham, Witzerland, 2018; Volume 11045 LNCS, pp. 3–11. [CrossRef]
50. Deng, Y.; Hou, Y.; Yan, J.; Zeng, D. ELU-Net: An Efficient and Lightweight U-Net for Medical Image Segmentation. *IEEE Access* **2022**, *10*, 35932–35941. [CrossRef]
51. Cao, W.; Zheng, J.; Xiang, D.; Ding, S.; Sun, H.; Yang, X.; Liu, Z.; Dai, Y. Edge and neighborhood guidance network for 2D medical image segmentation. *Biomed. Signal Process. Control* **2021**, *69*, 102856. [CrossRef]
52. Yan, Y.; Ren, J.; Liu, Q.; Zhao, H.; Sun, H.; Zabalza, J. PCA-domain Fused Singular Spectral Analysis for fast and Noise-Robust Spectral-Spatial Feature Mining in Hyperspectral Classification. *IEEE Geosci. Remote Sens. Lett.* **2021**. [CrossRef]
53. Bazi, Y.; Bashmal, L.; Al Rahhal, M.M.; Dayil, R.A.; Ajlan, N. Al Vision transformers for remote sensing image classification. *Remote Sens.* **2021**, *13*, 516. [CrossRef]
54. Liang, J.; Deng, Y.; Zeng, D. A Deep Neural Network Combined CNN and GCN for Remote Sensing Scene Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4325–4338. [CrossRef]
55. Xiong, Z.; Cai, J. Multi-scale Graph Convolutional Networks with Self-Attention. *arXiv* 2021. [CrossRef]
56. Khan, N.; Chaudhuri, U.; Banerjee, B.; Chaudhuri, S. Graph convolutional network for multi-label VHR remote sensing scene recognition. *Neurocomputing* **2019**, *357*, 36–46. [CrossRef]
57. Yuan, J.; Qiu, Y.; Wang, L.; Liu, Y. Non-Intrusive Load Decomposition Based on Graph Convolutional Network. In Proceedings of the 2022 IEEE 5th International Electrical and Energy Conference (CIEEC), Nangjing, China, 27–29 May 2022; pp. 1941–1944. [CrossRef]
58. Liu, Q.; Xiao, L.; Huang, N.; Tang, J.; Member, S. Composite Neighbor-Aware Convolutional Metric Networks for Hyperspectral Image Classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–15. [CrossRef]
59. Lu, Y.; Chen, Y.; Zhao, D.; Chen, J. Graph-FCN for Image Semantic Segmentation. *Comput. Vis. Pattern Recognit.* **2019**, *11554*, 97–105. [CrossRef]
60. Liu, Q.; Xiao, L.; Yang, J.; Wei, Z. Multilevel Superpixel Structured Graph U-Nets for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5516115. [CrossRef]

61. Liu, Q.; Xiao, L.; Yang, J.; Wei, Z. CNN-Enhanced Graph Convolutional Network with Pixel- and Superpixel-Level Feature Fusion for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 8657–8671. [CrossRef]
62. Ding, Y.; Zhang, Z.; Zhao, X.; Hong, D.; Cai, W.; Yu, C.; Yang, N.; Cai, W. Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification. *Neurocomputing* **2022**, *501*, 246–257. [CrossRef]
63. Wang, S.H.; Govindaraj, V.V.; Górriz, J.M.; Zhang, X.; Zhang, Y.D. COVID-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Inf. Fusion* **2021**, *67*, 208–229. [CrossRef]
64. Peng, F.; Lu, W.; Tan, W.; Qi, K.; Zhang, X.; Zhu, Q. Multi-Output Network Combining GNN and CNN for Remote Sensing Scene Classification. *Remote Sens.* **2022**, *14*, 1478. [CrossRef]
65. Knight, J. From timber to tourism: Recommoditizing the Japanese forest. *Dev. Chang.* **2000**, *31*, 341–359. [CrossRef]
66. Kosztra, B.; Büttner, G.; Hazeu, G.; Arnold, S. *Updated CLC Illustrated Nomenclature Guidelines*; European Environment Agency: Wien, Austria, 2017; Available online: https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018_Nomenclature_illustrated_guide_20190510.pdf (accessed on 8 December 2022).
67. de la Cuesta, I.R.; Blanco, J.A.; Imbert, J.B.; Peralta, J.; Rodríguez-Pérez, J. Changes in Long-Term Light Properties of a Mixed Conifer—Broadleaf Forest in Southwestern Europe Ignacio. *Forests* **2021**, *12*, 1485. [CrossRef]
68. Asner, G.P.; Martin, R.E. Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. *Remote Sens. Environ.* **2008**, *112*, 3958–3970. [CrossRef]
69. Zhang, C.; Ma, L.; Chen, J.; Rao, Y.; Zhou, Y.; Chen, X. Assessing the impact of endmember variability on linear Spectral Mixture Analysis (LSMA): A theoretical and simulation analysis. *Remote Sens. Environ.* **2019**, *235*, 111471. [CrossRef]
70. Wang, Q.; Ding, X.; Tong, X.; Atkinson, P.M. Spatio-temporal spectral unmixing of time-series images. *Remote Sens. Environ.* **2021**, *259*, 112407. [CrossRef]
71. Knyazikhin, Y.; Schull, M.A.; Stenberg, P.; Mõttus, M.; Rautiainen, M.; Yang, Y.; Marshak, A.; Carmona, P.L.; Kaufmann, R.K.; Lewis, P.; et al. Hyperspectral remote sensing of foliar nitrogen content. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E185–E192. [CrossRef]
72. Oreti, L.; Giuliarelli, D.; Tomao, A.; Barbati, A. Object oriented classification for mapping mixed and pure forest stands using very-high resolution imagery. *Remote Sens.* **2021**, *13*, 2508. [CrossRef]
73. Kattenborn, T.; Eichel, J.; Wiser, S.; Burrows, L.; Fassnacht, F.E.; Schmidtlein, S. Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. *Remote Sens. Ecol. Conserv.* **2020**, *6*, 472–486. [CrossRef]
74. Jayathunga, S.; Owari, T.; Tsuyuki, S. Analysis of forest structural complexity using airborne LiDAR data and aerial photography in a mixed conifer–broadleaf forest in northern Japan. *J. For. Res.* **2018**, *29*, 479–493. [CrossRef]
75. Zarco-Tejada, P.J.; Hornero, A.; Beck, P.S.A.; Kattenborn, T.; Kempeneers, P.; Hernández-Clemente, R. Chlorophyll content estimation in an open-canopy conifer forest with Sentinel-2A and hyperspectral imagery in the context of forest decline. *Remote Sens. Environ.* **2019**, *223*, 320–335. [CrossRef]
76. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large kernel matters—Improve semantic segmentation by global convolutional network. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017, Honolulu, HI, USA, 21–26 July 2016; pp. 1743–1751. [CrossRef]
77. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141. [CrossRef]
78. Ouyang, S.; Li, Y. Combining deep semantic segmentation network and graph convolutional neural network for semantic segmentation of remote sensing imagery. *Remote Sens.* **2021**, *13*, 119. [CrossRef]
79. Li, L.; Tang, S.; Deng, L.; Zhang, Y.; Tian, Q. Image caption with global-local attention. In Proceedings of the 31st AAAI Conference on Artificial Intelligence AAAI 2017, San Francisco, CA, USA, 4–9 February 2017; Volume 31, pp. 4133–4139. [CrossRef]
80. Zhang, C.; Chen, X.; Ji, S. Semantic image segmentation for sea ice parameters recognition using deep convolutional neural networks. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102885. [CrossRef]
81. The University of Tokyo Forests, Graduate School of Agricultural and Life Sciences. *Education and Research Plan (2021–2030) of the University of Tokyo Forests: Part 2 Standing Technical Committee Plans*; The University of Tokyo Forests: Tokyo, Japan, 2022; Volume 64, pp. 33–49. [CrossRef]
82. Fadnavis, S. Image Interpolation Techniques in Digital Image Processing: An Overview. *Int. J. Eng. Res. Appl.* **2014**, *4*, 70–73.
83. Ohsato, S.; Negisi, K. *Miscellaneous Information, the University of Tokyo Forests*; The Tokyo University Forests: Tokyo, Japan, 1994; Volume 32, pp. 9–35. (In Japanese) [CrossRef]
84. Gu, Z.; Cheng, J.; Fu, H.; Zhou, K.; Hao, H.; Zhao, Y.; Zhang, T.; Gao, S.; Liu, J. CE-Net: Context Encoder Network for 2D Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2019**, *38*, 2281–2292. [CrossRef] [PubMed]
85. Ma, Y.; Guo, Y.; Liu, H.; Lei, Y.; Wen, G. Global context reasoning for semantic segmentation of 3D point clouds. In Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 1–5 March 2020; pp. 2920–2929. [CrossRef]
86. Li, X.; Sun, X.; Meng, Y.; Liang, J.; Wu, F.; Li, J. Dice Loss for Data-imbalanced NLP Tasks. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 465–476. [CrossRef]

87. Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571. [CrossRef]
88. Chen, H.; Liu, X.; Jia, Z.; Liu, Z.; Shi, K.; Cai, K. A combination strategy of random forest and back propagation network for variable selection in spectral calibration. *Chemom. Intell. Lab. Syst.* **2018**, *182*, 101–108. [CrossRef]
89. Shao, Z.; Zhou, W.; Deng, X.; Zhang, M.; Cheng, Q. Multilabel Remote Sensing Image Retrieval Based on Fully Convolutional Network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 318–328. [CrossRef]
90. Deng, P.; Xu, K.; Huang, H. When CNNs Meet Vision Transformer: A Joint Framework for Remote Sensing Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8020305. [CrossRef]
91. Sangeetha, V.; Prasad, K.J.R. Deep Residual Learning for Image Recognition Kaiming. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2006**, *45*, 1951–1954. [CrossRef]
92. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. In Proceedings of the 7th International Conference on Learning Representations (ICLR 2019), New Orleans, LA, USA, 6–9 May 2019.
93. Culjak, I.; Abram, D.; Pribanic, T.; Dzapo, H.; Cifrek, M. A brief introduction to OpenCV. In Proceedings of the 2012 Proceedings of the 35th International Convention MIPRO, Opatija, Croatia, 21–25 May 2012; pp. 1725–1730.
94. Barupal, D.K.; Fiehn, O. Scikit-learn: Machine Learning in Python. *Environ. Health Perspect.* **2019**, *127*, 2825–2830. [CrossRef]
95. Acharjya, P.P.; Das, R.; Ghoshal, D. Study and Comparison of Different Edge Detectors for Image Segmentation. *Glob. J. Comput. Sci. Technol. Graph. Vis.* **2012**, *12*, 29–32.
96. Basu, M.; Member, S. Gaussian-Based Edge-Detection Methods—A Survey. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2002**, *32*, 252–260. [CrossRef]
97. Adrian, J.; Sagan, V.; Maimaitijiang, M. Sentinel SAR-optical fusion for crop type mapping using deep learning and Google Earth Engine. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 215–235. [CrossRef]
98. Carbonneau, P.E.; Dugdale, S.J.; Breckon, T.P.; Dietrich, J.T.; Fonstad, M.A.; Miyamoto, H.; Woodget, A.S. Adopting deep learning methods for airborne RGB fluvial scene classification. *Remote Sens. Environ.* **2020**, *251*, 112107. [CrossRef]
99. Molchanov, P.; Tyree, S.; Karras, T.; Aila, T.; Kautz, J. Pruning convolutional neural networks for resource efficient inference. In Proceedings of the 5th International Conference on Learning Representations ICLR 2017—ICLR 2017 Conference Track, Toulon, France, 24–26 April 2017; pp. 1–17. [CrossRef]
100. Markoulidakis, I.; Rallis, I.; Georgoulas, I.; Kopsiaftis, G.; Doulamis, A.; Doulamis, N. Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies* **2021**, *9*, 81. [CrossRef]
101. Aamir, M.; Li, Z.; Bazai, S.; Wagan, R.A.; Bhatti, U.A.; Nizamani, M.M.; Akram, S. Spatiotemporal Change of Air-Quality Patterns in Hubei Province—A Pre- to Post-COVID-19 Analysis Using Path Analysis and Regression. *Atmosphere* **2021**, *12*, 1338. [CrossRef]
102. Wilebore, B.; Coomes, D. Combining spatial data with survey data improves predictions of boundaries between settlements. *Appl. Geogr.* **2016**, *77*, 1–7. [CrossRef]
103. Perez, L.; Wang, J. The effectiveness of data augmentation in image classification using deep learning. *arXiv* **2017**. [CrossRef]
104. Karatas, G.; Demir, O.; Sahingoz, O.K. Increasing the Performance of Machine Learning-Based IDSs on an Imbalanced and Up-to-Date Dataset. *IEEE Access* **2020**, *8*, 32150–32162. [CrossRef]
105. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.W.; Wu, J. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1055–1059. [CrossRef]
106. Zhang, M.; Zhang, H.; Li, J.; Wang, L.; Fang, Y.; Sun, J. Supervised graph regularization based cross media retrieval with intra and inter-class correlation. *J. Vis. Commun. Image Represent.* **2019**, *58*, 1–11. [CrossRef]
107. Kosaka, N.; Akiyama, T.; Tsai, B.; Kojima, T. Forest type classification using data fusion of multispectral and panchromatic high-resolution satellite imageries. *Int. Geosci. Remote Sens. Symp.* **2005**, *4*, 2980–2983. [CrossRef]
108. Johnson, B.; Tateishi, R.; Xie, Z. Using geographically weighted variables for image classification. *Remote Sens. Lett.* **2011**, *3*, 491–499. [CrossRef]
109. Mellor, A.; Boukir, S.; Haywood, A.; Jones, S. Exploring issues of training data imbalance and mislabelling on random forest performance for large area land cover classification using the ensemble margin. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 155–168. [CrossRef]
110. Schlerf, M.; Atzberger, C. Vegetation structure retrieval in beech and spruce forests using spectrodirectional satellite data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 8–17. [CrossRef]
111. Grabska, E.; Hostert, P.; Pflugmacher, D.; Ostapowicz, K. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sens.* **2019**, *11*, 1197. [CrossRef]
112. McIlrath, L.D. A CCD/CMOS Focal-Plane Array Edge Detection Processor Implementing the Multi-Scale Veto Algorithm. *IEEE J. Solid-State Circuits* **1996**, *31*, 1239–1247. [CrossRef]
113. Wu, J.; Zhou, W.; Luo, T.; Yu, L.; Lei, J. Multiscale multilevel context and multimodal fusion for RGB-D salient object detection. *Signal Process.* **2021**, *178*, 63–65. [CrossRef]
114. Li, J.; Hong, D.; Gao, L.; Yao, J.; Zheng, K.; Zhang, B.; Chanussot, J. Deep Learning in Multimodal Remote Sensing Data Fusion: A Comprehensive Review. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *112*, 102926. [CrossRef]

115. Jin, H.; Mountrakis, G. Fusion of optical, radar and waveform LiDAR observations for land cover classification. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 171–190. [CrossRef]
116. Hong, D.; Hu, J.; Yao, J.; Chanussot, J.; Zhu, X.X. Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 68–80. [CrossRef]
117. Hong, D.; Yokoya, N.; Ge, N.; Chanussot, J.; Zhu, X. Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 193–205. [CrossRef]