Sensors and Materials, Vol. 31, No. 11 (2019) 3849–3858
MYU Tokyo

S & M 2054

Comparative Analysis of Generalized Intersection over Union and Error Matrix for Vegetation Cover Classification Assessment

Hyun Choi,[1] Hyun-Jik Lee,[2] Ho-Jin You,[1] Sang-Yong Rhee,[3] and Wang-Su Jeon[3*]

[1]Department of Civil Engineering, Kyungnam University,
7, Gyeongnamdaehak-ro, Masanhappo-gu, Changwon-si, Gyeongsangnam-do 51767, Republic of Korea
[2]Department of Civil Engineering, Sangji University,
83, Sangjidae-gil, Wonju-si, Gangwon-do 26339, Republic of Korea
[3]Department of Computer Engineering, Kyungnam University,
7, Gyeongnamdaehak-ro, Masanhappo-gu, Changwon-si, Gyeongsangnam-do 51767, Republic of Korea

(Received August 29, 2019; accepted October 16, 2019)

Keywords: normalized difference vegetation index (NDVI), remote sensing, error matrix, Intersection over Union (IoU)

The result of vegetation cover classification greatly depends on the classification methods. Accuracy analysis is mostly performed using the error matrix in remote sensing. In recent remote sensing, image classification has been carried out on the basis of deep learning. In the field of image processing in computer science, Intersection over Union (IoU) is mainly used for accuracy analysis. In this study, the error matrix, which is frequently used in remote sensing, and IoU, which is mainly used for deep learning images, were compared and reviewed to analyze their accuracy levels for the results of vegetation index calculation. The results of vegetation index calculation were applied to the comparison of the accuracy levels of IoU and the error matrix. According to the results of accuracy analysis using the error matrix, which is based on random points, the accuracy of the normalized difference vegetation index (NDVI) was shown to be 82.4% and that of deep learning was shown to be 93.7%, with a difference of about 11.3%.

1. Introduction

Remote sensing is a technology for observing distant objects. It measures physical properties by detecting electromagnetic waves that radiated from the target without direct contact with the target. Remote sensing data are important for various types of decision making in various fields, in which they are used through processes such as maintenance, analysis, construction, and editing of spatial information. In South Korea, it is very important to build accurate databases of forests, which occupy most of the territory, in order to prevent disasters and accidents. The accuracy of image classification is examined using the error matrix technique. However, when the accuracy is analyzed with the error matrix, which is based on experience points, the reliability of the accuracy declines. Therefore, the accuracy is analyzed

[*]Corresponding author: e-mail: jws2218@naver.com
https://doi.org/10.18494/SAM.2019.2584

ISSN 0914-4935 © MYU K.K.
https://myukk.org/

by the Intersection over Union (IoU) method, but the difference in accuracy level between the vegetation cover type identified using the normalized difference vegetation index (NDVI) and that identified using the deep learning technique cannot be known. IoU is the most popular evaluation metric used in object detection benchmarks. Also known as the Jaccard index, it is the most commonly used metric for determining the similarity between two arbitrary shapes. Therefore, to solve this problem, in this study, the vegetation cover types were calculated by the deep learning method, and the accuracy of the vegetation index calculated using the existing error matrix and that by the IoU method were compared.

2. Experimental Methods

2.1 Convolution neural network (CNN)

Before moving to specific solutions for NDVI estimation, in this section, we provide some basic notions and terminologies about CNNs. Over the last few years, CNNs have been successfully applied to many classical image processing problems, such as denoising,(1) super-resolution,(2) pansharpening,(3,4) segmentation,(5,6) object detection,(7,8) change detection,(9) and classification.(10–13) The main strengths of CNNs are (i) an extreme versatility that allows them to approximate any sort of linear or nonlinear transformation, including scaling or hard thresholding; (ii) no need to design handcrafted filters, replaced by machine learning; and (iii) high-speed processing, due to parallel computing. On the downside, for correct training, CNNs require the availability of a large amount of data with the ground truth (examples). In our specific case, data are not a problem, given the unlimited quantity of cloud-free Sentinel-2 time series that can be downloaded from web repositories. However, using large datasets has a cost in terms of complexity and may lead to unreasonably long training times. Usually, a CNN is a chain (parallels, loops, or other combinations are also possible) of different layers, such as convolution, nonlinearities, pooling, and deconvolution. For image processing tasks in which the desired output is an image at the same resolution of the input, as in this work, only convolutional layers that interleaved with nonlinear activations are typically employed.

The generic l-th convolutional layer, with N-band input $x^{(l)}$, yields an M-band stack $z^{(l)}$ computed as

$$z^{(l)} = w^{(l)} * x^{(l)} + b^{(l)}, \tag{1}$$

whose m-th component can be written in terms of ordinary 2D convolutions:

$$z^{(l)}(m,\cdot,\cdot) = \sum_{n=1}^{N} w^{(l)}(m,n,\cdot,\cdot) * x^{(l)}(n,\cdot,\cdot) + b^{(l)}(m). \tag{2}$$

The tensor w is a set of M convolutional $N \times (K \times K)$ kernels, with a $K \times K$ spatial support (receptive field), while b is an M-vector bias. These parameters, compactly, $\Phi_l \triangleq \left(w^{(l)}, b^{(l)}\right)$, are learned during the training phase. If the convolution is followed by a pointwise activation function $g_l(\cdot)$, then the overall layer output is given by

$$y^{(l)} = g_l\left(z^{(l)}\right) = g_l\left(w^{(l)} * x^{(l)} + b^{(l)}\right) \triangleq f_l\left(x^{(l)}, \Phi_l\right). \tag{3}$$

Owing to the good convergence properties it ensures,(11) the rectified linear unit (ReLU), defined as $\mathrm{g}(\cdot) \triangleq \max(0, \cdot)$, is a typical activation function of choice for input or hidden layers. Assuming a simple L-layer cascade architecture, the overall processing will be

$$f(x, \Phi) = f_L\left(f_{L-1}\left(\dots f_1(x, \Phi_1), \dots, \Phi_{L-1}\right), \Phi_L\right), \tag{4}$$

where $\Phi \triangleq (\Phi_1, \dots, \Phi_L)$ is the whole set of parameters to learn. In this chain, each layer l provides a set of so-called feature maps, $y^{(l)}$, to become more and more representative of abstract and global phenomena in subsequent ones (large l). In this work, all proposed solutions are based on a simple three-layer architecture and differ only in the input layer, as different combinations of input bands are considered.

Once the architecture has been chosen, its parameters are learned by means of some optimization strategy. An example is the stochastic gradient descent (SGD) algorithm, specifying the cost to be minimized over a properly selected training dataset. Details on training will be given below for our specific solution.

2.2 Atrous convolution

Semantic image segmentation using existing CNNs such as AlexNet, VGGNet, and Googlenet is simpler than that using histograms or graphs or using machine learning methods, such as decision trees and clustering by identifying feature regions, and shows higher performance. However, it has the shortcoming that when continuous layers such as fully convolution network (FCN) and DeconvNet are calculated, the sizes of features are reduced and the amount of computation and memory use increase. To solve this problem, atrous convolution was devised for the deepLab model. This method uses a filter made by increasing the size of the existing convolution filter and filling the spaces between weights with 0 so that holes are formed. First, when signals are assumed as one-dimensional signals as shown in Eq. (1), the output of $x[k]$ with a length k in the input signal $\omega[k]$ is equal to $y[i]$.

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k]\,\omega[k], \tag{5}$$

where r is stride and is in the range of atrous convolution. In the case of basic convolution operation, 1 is used as the stride. It can be seen that, in the case of atrous convolution, the amount of calculation decreases even when the acceptance range is widened, as shown in Fig. 1.

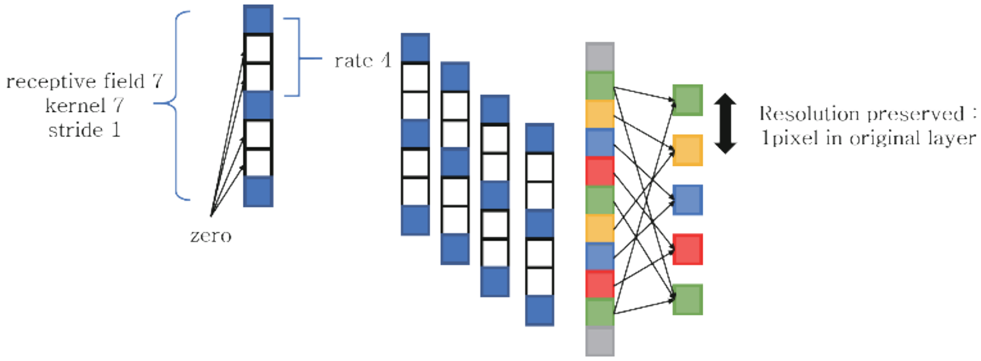

Fig. 1. (Color online) Atrous convolution structure.

2.3 Detailed structure of the model

To increase the amount of data used for learning, a method of additionally considering the original image and the R, G, and B channels are shown in Fig. 2. We used the data augmentation and the R, G, B split channel in this paper. ResNet has a problem that when it is used with the existing VGGNet and GoogleNet, more than 30 layers are not learned well. To solve this problem, a bottleneck structure is used so that the depth of the network can be increased to 50–150 layers. In the case of the bottleneck structure, since 1 × 1 and 3 × 3 convolutions are used, the effects of being able to reduce and expand the dimension while reducing the calculation amount can be obtained. In addition, the residual connection that adds the previous values is used to solve the problem that the slope disappears when the network is deepened. The ResNet-51 layer structure is used for the backbone network. In this case, the 51 layers of ResNet are calculated using convolution operation, and the remaining 50 layers are calculated using atrous convolution. After using 3 × 3, which extracts features from the front layer as the atrous convolution, atrous spatial pyramid pooling (ASPP), which is a technique used in image pyramids, is applied to the last layer at different rates {2,4,6,8,10,12,14} to adjust the size of the image, and the results obtained are concatenated to produce the result. In this case, the structure of the ASPP layer is as shown in Fig. 3.

2.4 Generalized Intersection over Union (GIoU)

The IoU for determining the similarity between two arbitrary shapes (volumes) $\boldsymbol{A}, \boldsymbol{B} \subseteq \boldsymbol{S} \in \boldsymbol{R}^n$ is attained as

$$\boldsymbol{IoU} = \frac{|\boldsymbol{A} \cap \boldsymbol{B}|}{|\boldsymbol{A} \cup \boldsymbol{B}|}. \tag{6}$$

Here, $\boldsymbol{A}$ is the output result, $\boldsymbol{B}$ is the Ground Truth, $\boldsymbol{S}$ is overlapping space. Two appealing features, which make this similarity measure popular for evaluating many 2D/3D computer vision tasks, are as follows:

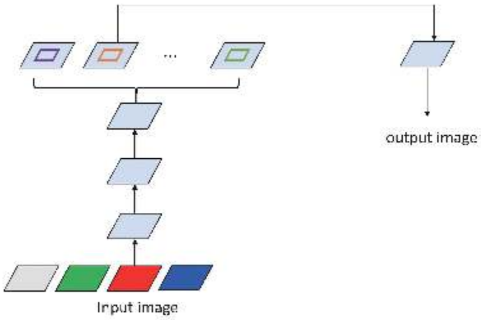

Fig. 2. (Color online) Preprocessing considering R, G, and B channels.

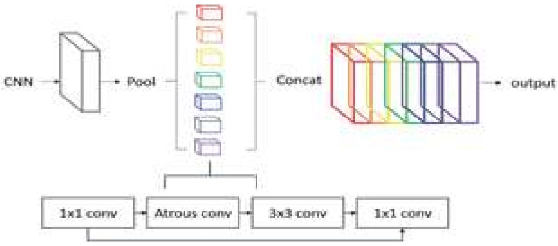

Fig. 3. (Color online) Structure of ASPP layer.

Algorithm 1: GIoU

```
Input: Two arbitrary convex shapes: $\boldsymbol{A}, \boldsymbol{B} \subseteq \mathbb{S} \in \mathbb{R}^{\boldsymbol{n}}$
Output: $\boldsymbol{GIoU}$
1.  For $\boldsymbol{A}$ and $\boldsymbol{B}$, find the smallest enclosing convex object $\boldsymbol{C}$, where $\boldsymbol{C} \subseteq \mathbb{S} \in \mathbb{R}^{\boldsymbol{n}}$
2.  $\boldsymbol{IoU} = \frac{|\boldsymbol{A} \cap \boldsymbol{B}|}{|\boldsymbol{A} \cup \boldsymbol{B}|}$
3.  $\boldsymbol{GIoU} = \boldsymbol{IoU} - \frac{|\boldsymbol{C} / (\boldsymbol{A} \cup \boldsymbol{B})|}{|\boldsymbol{C}|}$
```

- IoU as a distance, $\mathcal{L}_{IoU} = 1 - \boldsymbol{IoU}$, is a metric by mathematical definition.(14) It means that $\mathcal{L}_{IoU}$ fulfills all properties of a metric such as non-negativity, the identity of indiscernibles, symmetry, and triangle inequality.
- IoU is invariant to the scale of the problem. This means that the similarity between two arbitrary shapes $\boldsymbol{A}$ and $\boldsymbol{B}$ is independent of the scale of their space $\boldsymbol{S}$ (the proof is provided in Supplementary Material).

However, IoU has a major weakness:

- If $|\boldsymbol{A} \cap \boldsymbol{B}| = 0$, $\boldsymbol{IoU}(\boldsymbol{A}, \boldsymbol{B}) = 0$. In this case, IoU does not reflect if two shapes are in the vicinity of each other or very far from each other.

To address this issue, we propose a general extension to IoU, namely, GIoU. For two arbitrary convex shapes (volumes) $\boldsymbol{A},\boldsymbol{B} \subseteq \boldsymbol{S} \in \boldsymbol{R}^n$, we first find the smallest convex shapes $\boldsymbol{C} \subseteq \boldsymbol{S} \in \boldsymbol{R}^n$ enclosing both $\boldsymbol{A}$ and $\boldsymbol{B}$. For comparing two specific types of geometric shapes, $\boldsymbol{C}$ can be from the same type. For example, in the case of two arbitrary ellipsoids, $\boldsymbol{C}$ could be the smallest ellipsoids enclosing $\boldsymbol{A}$ and $\boldsymbol{B}$. Then, we calculate the ratio between the volumes (areas) occupied by $\boldsymbol{C}$ excluding $\boldsymbol{A}$ and $\boldsymbol{B}$**,** and divide it by the total volume (area) occupied by $\boldsymbol{C}$. This represents a normalized measure that focuses on the empty volume (area) between $\boldsymbol{A}$ and $\boldsymbol{B}$. Finally, GIoU is attained by subtracting this ratio from the IoU value. The calculation of GIoU is summarized in measurement.

GIoU as a new metric has the following properties:

(1) Similarly to IoU, GIoU as a distance, e.g., $\mathcal{L}_{IoU} = 1 - \boldsymbol{IoU}$, holds all the properties of a metric such as non-negativity, the identity of indiscernibles, symmetry, and triangle inequality.

(2) Similarly to IoU, GIoU is invariant to the scale of the problem.

(3) GIoU is always a lower bound for IoU, i.e., $\forall \boldsymbol{A}, \boldsymbol{B} \subseteq \mathbb{S}$, $\boldsymbol{GIoU}(\boldsymbol{A}, \boldsymbol{B}) \le \boldsymbol{IoU}(\boldsymbol{A}, \boldsymbol{B})$, and this lower bound becomes tighter when A and B have stronger shape similarity and proximity, i.e, $\lim_{\mathbf{A}\to\mathbf{B}} \boldsymbol{GIoU}\left(\boldsymbol{A},\boldsymbol{B}\right) = \boldsymbol{IoU}\left(\boldsymbol{A},\boldsymbol{B}\right)$.

(4) $\forall \boldsymbol{A}, \boldsymbol{B} \subseteq \mathbb{S}$, $0 \le \boldsymbol{IoU}(\boldsymbol{A}, \boldsymbol{B}) \le 1$, but GIoU has a symmetric range, i.e, $\forall \boldsymbol{A}, \boldsymbol{B} \subseteq \mathbb{S}$, $-1 \le \boldsymbol{GIoU}(\boldsymbol{A}, \boldsymbol{B}) \le 1$.

I) Similarly to IoU, the value 1 occurs only when two objects overlay perfectly, i.e, if $\left|\boldsymbol{A} \cup \boldsymbol{B}\right| = \left|\boldsymbol{A} \cap \boldsymbol{B}\right|$, then $\boldsymbol{GIoU} = \boldsymbol{IoU} = 1$

II) The GIoU value asymptotically converges to −1 when the ratio between occupying regions of two shapes, $\left|\boldsymbol{A} \cup \boldsymbol{B}\right|$, and the volume (area) of the enclosing shape $\left|\boldsymbol{C}\right|$ tends to be zero, i.e, $\lim_{\frac{|\boldsymbol{A}\cup\boldsymbol{B}|}{|\boldsymbol{C}|}\to 0} \boldsymbol{GIoU}\left(\boldsymbol{A},\boldsymbol{B}\right) = \mathbf{-1}$

In summary, this generalization keeps the major properties of IoU while rectifying its weakness. Therefore, GIoU can be a proper substitute for IoU in all performance measures used in 2D/3D computer vision tasks. In this paper, we only focus on 2D object detection where we can easily derive an analytical solution for GIoU to apply it as both metric and loss. The extension to non-axis aligned 3D cases is left as future work.

3. Experiment

3.1 Study area

The classification images used in this study were taken from areas near Cheongju and Okcheon, Chungbuk, South Korea on November 24, 2016 with Kompsat – 3A (Fig. 4). The images were rearranged into RMSE 0.6 m by the pan sharpening pretreatment process that make images clearer using PCI Geomatica.

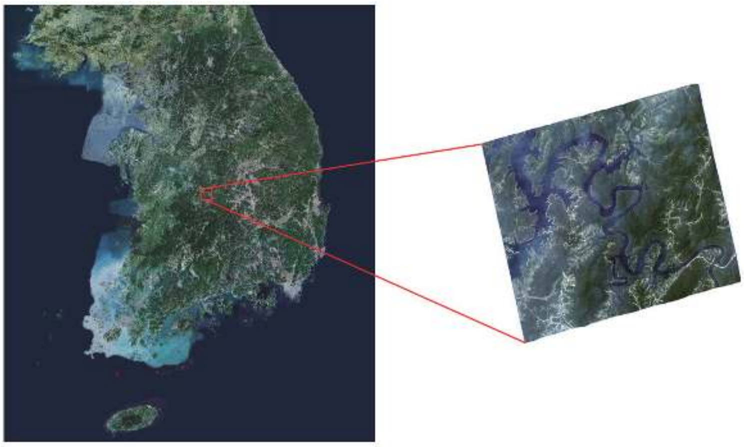

Fig. 4. (Color online) Satellite image used in this study.

3.2 Study method

The deep learning technique has mainly been used in remote sensing recently. We were used to CNN classify vegetation cover types. The CNN technique is a convolution-based neural network that extracts and learns various features from massive data. Using this technique, spatial object items with similar spectral characteristics were extracted. The colors were labeled according to a geographic feature (red: deciduous trees, yellow: coniferous trees, and black: watershed) by pixel by the semantic segmentation method, developed on the basis of deepLab, and the baseline model was used to improve the detection speed and performance. To extract objects from satellite images, the objects were classified by the semantic image classification technique in the CNN technique using the deepLab model based on atrous convolution, which was selected because of its advantage of easy image classification, owing to its wide receptive field. Since ResNet has a problem that more than 30 layers are not learned when VGGNET and GoogleNet are used, a bottleneck structure that can increase the depth of the network from to 50 to 150 layers was used to solve the problem. The bottleneck structure can reduce and expand the dimension while reducing the amount of analysis by using 1 × 1 and 3 × 3 convolutions. The computer environment was configured with a CPU i7-6700k, 32 GB of memory, and GPU TITAN × 2, and implemented by a fine tuning process using Image-net 2012 pretraining data. The rate of learning used was 0.001, Adam was used as a gradient descent method, and the number of repetitions was 100000. The data used in classification were satellite images with a size of about 8000 × 8000 grids. Since it would use a lot of GPU memory, each image was divided into 321 × 321 grid size pieces for efficient calculation. Thereafter, about 700 images were used for learning and 294 images were used for tests. In addition, a residual connection that adds previous results was used so that the learning was carried out stably. For accuracy analysis, as shown in Fig. 5, the error matrix frequently used in remote sensing was compared and analyzed with IoU, which is mainly used in deep learning image quality analysis.

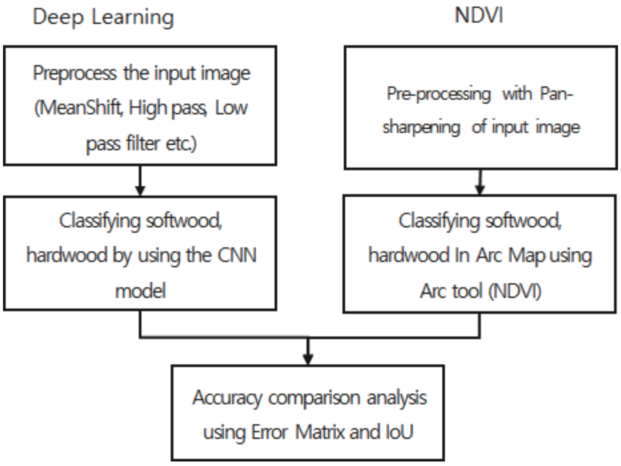
Deep Learning
NDVI
Preprocess the input image (MeanShift, High pass, Low pass filter etc.)
Pre-processing with Pan-sharpening of input image
Classifying softwood, hardwood by using the CNN model
Classifying softwood, hardwood In Arc Map using Arc tool (NDVI)
Accuracy comparison analysis using Error Matrix and IoU

Fig. 5. Accuracy comparison analysis in this study.

3.3 Accuracy analysis

To extract objects from satellite images, conifers and deciduous trees were identified using the semantic image classification technique in the CNN technique using a deepLab model. DeepLab models are based on atrous convolution. These models have the advantage of easy image classification since their receptive field is wide and they can be used to carry out calculations if unpooling and convolution are combined. To evaluate the performance of the model, the accuracy of classification was calculated using IoU, and the results are shown. To analyze performance, the results of classification were compared with the conifers and deciduous trees in the land cover map provided by the Environmental Geographic Information Service (EGIS) of South Korea. According to the results of accuracy analysis using the error matrix, which was based on random points, the accuracy of NDVI was 82.4% and that of deep learning was 93.7%, with a difference of 11.3%. According to the results of accuracy analysis using IoU, which is based on grids, the accuracy of NDVI was 49.4% and that of deep learning was 71.6%, with a difference of about 22.2%. Figure 6 shows the results of the classification of trees into coniferous and deciduous for the test data. Cases of misclassification occur because the pixels of conifers and deciduous trees are similar. The resultant vegetation cover types obtained using NDVI and the resultant vegetation cover types obtained by the deep learning method were compared in terms of area (Table 1) and accuracy (Table 2).

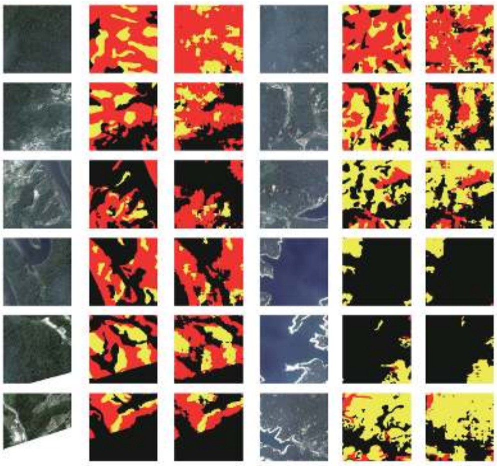

Fig. 6. (Color online) Results of classification of trees into coniferous and deciduous trees for the test data.

Table 1
IoU and error matrix area ratio. (Unit: ha)

	Total area	Deciduous trees	Coniferous trees	Watershed
Ground truth	11107.75	2187.29	4136.84	4783.62
Deep learning	11107.75	1900.32	4406.74	4800.69
NDVI	11107.75	1819.83	3474.95	5812.97

Table 2
IoU accuracy compared with the error matrix. (Unit: %)

	Technique	Deciduous trees	Coniferous trees	Watershed
NDVI	Error matrix	83.2	84.0	80.0
	IoU	48.7	49.2	50.4
Deep learning	Error matrix	94.0	94.1	93.0
	IoU	72.0	72.0	71.0

4. Conclusions

In this study, the accuracy of the classification result of vegetation cover is analyzed using IoU and the error matrix method; the accuracy of NDVI was 82.4% and that of deep learning was 93.7%, with a difference of 11.3%. According to the results of accuracy analysis using IoU, which is based on grids, the accuracy of NDVI was 49.4% and that of deep learning was 71.6%, with a difference of about 22.2%. Therefore, according to the results of the study on deep learning, the accuracy according to IoU was shown to be about 70% in the classification of the same image and about 60% when learned images were applied to other images. In the analysis of the vegetation cover types, the accuracy of the results of analysis using the error matrix was shown to be higher according to the accuracy setting method, and it could be seen that if the deep learning technique was used, the accuracy was high. In the case of the deep learning technique, studies to improve performance through learning the regional characteristics of conifers and deciduous trees are necessary.

Acknowledgments

This research was supported by a grant (19SIUE-B148326-02) from the KAS 500-1/2 Image Acquisition and Utilization Technology Development Program, funded by the Ministry of Land, Infrastructure and Transport of the Korean government.

References

1 K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang: IEEE Trans. Image Process. **26** (2017) 3142. https://doi.org/10.1109/tip.2017.2662206

2 C. Dong, C. Loy, K. He, and X. Tang: IEEE Trans. Pattern Anal. **38** (2016) 295. https://doi.org/10.1109/tpami.2015.2439281

3 G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa: Remote Sens. **8** (2016) 594. https://doi.org/10.3390/rs8070594

4 G. Scarpa, S. Vitale, and D. Cozzolino: IEEE Trans. Geosci. Remote Sens. **56** (2018) 5443. https://doi.org/10.1109/tgrs.2018.2817393

5 L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. Yuille: IEEE Trans. Pattern Anal. Mach. Intell. **40** (2018) 838. https://doi.org/10.1109/tpami.2017.2699184

6 J. Long, E. Shelhamer, and T. Darrell: IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (2015) 3431. https://doi.org/10.1109/cvpr.2015.7298965

7 N. Zhang, J. Donahue, R. Girshick, and T. Darrell: Proc. European Conf. Computer Vision **54** (2014) 834. https://doi.org/10.1007/978-3-319-10590-1_54

8 E. Maltezos, N. Doulamis, A. Doulamis, and C. Ioannidis: J. Appl. Remote Sens. **11** (2017) 1931. https://doi.org/10.1117/1.JRS.11.042620

9 N. Zhang, J. Donahue, R. Girshick, and T. Darrell: Proc. European Conf. Computer Vision **54** (2014) 834. https://doi.org/10.1007/978-3-319-10590-1_54

10 F. Jahan and M. Awrangjeb: ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. (2017) 711. https://doi.org/10.5194/isprs-archives-xlii-2-w7-711-2017

11 A. Krizhevsky, I. Sutskever, and G. E. Hinton: Proc. Adv. Neural Inf. Process. Syst. **60** (2012) 1106. https://doi.org/10.1145/3065386

12 L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao: IEEE Trans. Geosci. Remote Sens. 55(2017) 5585. https://doi.org/10.1109/tgrs.2017.2710079

13 K. Fotiadou, G. Tsagkatakis, and P. Tsakalides: Electron. Imaging (2017) 185. https://doi.org/10.2352/issn.2470-1173.2017.17.coimg-445

14 S. Kosub: Pattern Recognit. Lett. **120** (2019) 36. https://doi.org/10.1016/j.patrec.2018.12.007

About the Authors

Hyun Choi received his B.S. degree from Pukyung National University, Korea, in 1998 and his M.S. and Ph.D. degrees from Pusan National University, Korea, in 2000 and 2004, respectively. Since 2006, he is currently a professor of the Department of Civil Engineering at Kyungnam University, Changwon, South Korea. His research interests are in the utilization of various types of spatial information. (hchoi@kyungnam.ac.kr)

Hyun Jik Lee received his B.S. degree from Chungbuk National University, Republic of Korea, in 1984. He received his M.S. and Ph.D. degrees from Yonsei University, Republic of Korea, in 1986 and 1992, respectively. Since 1996, he has been a professor at Sangji University, Republic of Korea. He has worked on numerous research studies and projects related to measurements and GIS. His research interests are in the utilization of various types of spatial information. (hjiklee@sangji.ac.kr)

Ho-Jin You received his B.S. degree from Kyungnam University, Republic of Korea, in 2018. Since 2018, he has been a graduate student at Kyungnam University, Republic of Korea. His research interests are in the utilization of various types of spatial information. (susu4103@hanma.kr)

Sang-Yong Rhee received his B.S. and M.S. degrees in industrial engineering from Korea University, Seoul, South Korea, in 1982 and 1984, respectively, and his Ph.D. degree in industrial engineering from Pohang University, Pohang, South Korea. He is currently a professor of computer engineering at Kyungnam University, Changwon, South Korea. His research interests include computer vision, augmented reality, neuro-fuzzy, and human-robot interface. (syrhee@kyungnam.ac.kr)

Wang-Su Jeon (corresponding author) received his B.S. and M. S. degrees in computer engineering and IT convergence engineering from Kyungnam University, Masan, South Korea, in 2016 and 2018, respectively, and is currently pursuing his Ph.D. degree in IT convergence engineering at Kyungnam University, Changwon, South Korea. His present interests include computer vision, pattern recognition, and machine learning. (jws2218@naver.com)