COMBINATION OF 2D-CNN AND RANDOM FOREST MODELS FOR CROP CLASSIFICATION WITH UAV IMAGERY

Geun-Ho Kwak (1), Chan-Won Park (2), Kyung-Do Lee (2),
Sang-Il Na (2), Ho-Yong Ahn (2), No-Wook Park (1)

[1] Inha University, 100 Inha-ro, Michuhol-gu, Incheon 22212, Korea
[2] National Institute of Agriculture Sciences, 166 Nongsaengmyeong-ro, Iseo-myeon, Wanju 55365, Korea
Email: root0109@inha.edu; cwpark@korea.kr; kdlee11@korea.kr; sangil917@korea.kr; hyahn85@korea.kr; nwpark@inha.ac.kr

KEY WORDS: Deep learning, classification, training samples

ABSTRACT: Recently, a convolutional neural network (CNN) has been regarded as an effective deep learning model that can extract spatial contextual information without user's intervention for classification. However, to extract useful spatial features may be difficult from the CNN model when limited training data are used for supervised learning. In this case, if the simple application of softmax activation functions, the final classification may not lead to satisfactory classification performance due to less informative spatial features. As an alternative, conventional machine learning algorithms can improve the classification performance because more sophisticated algorithms than the softmax operator are applied to the final classification. In this paper, a hybrid model is presented that combines two dimensional CNN (2D-CNN) and random forest (RF). Spatial contextual information extracted from 2D-CNN is used as input features of RF-based classification. To evaluate the potential of the hybrid model for crop classification, a case study of crop classification with unmanned aerial vehicle images was carried out. The classification performance of the hybrid model proposed in this study was superior to those of 2D-CNN and RF classifiers, implying the effectiveness of the proposed model when small training data are used for supervised classification.

1. INTRODUCTION

Crop maps, which have been routinely crop yield prediction and crop growth monitoring, are produced using remote sensing data. Recently, the importance of food security has been emphasized in relation to the rise of international grain price due to environmental and climate changes. Therefore, the producing a reliable crop maps are one of the most important issues in agriculture.

The selection of an appropriate classification methodology for producing reliable crop maps is very important. Machine learning (ML) models including support vector machine (SVM) and random forest (RF) have been widely applied for crop classification. When the conventional ML models are applied to supervised classification, appropriate feature extraction and selection should be performed in advance (Löw *et al*., 2013; Kwak and Park, 2019). However, extraction of optimal features may be time-consuming and much effort should be made prior to classification (Sidike *et al*., 2019).

In recent years, deep learning (DL) algorithms have been widely applied to video recognition and signal processing, as well as classification in remote sensing (Rußwurm and Körner, 2018; Wei *et al*., 2019). A convolutional neural network (CNN), which is one of DL algorithms, can consider spatial contextual information between neighboring pixels and be effectively applied to the classification of areas with similar spatial characteristics such as crop fields (Zhong *et al*., 2019). The CNN model is capable of automatically extracting high-level spatial features without user's intervention (Sidike *et al*., 2019; Zhu *et al*., 2017). Despite this advantage, the CNN model still has some limitations. It is often difficult to estimate the optimal model parameters for the extraction of high-level spatial features when small training data are used (Hu *et al*., 2015). Kim *et al*. (2018) compared the classification accuracy of SVM with that of two dimensional CNN (2D-CNN) when different numbers of training data and hyper-parameters were applied. The classification performance of 2D-CNN decreases significantly as the number of training data was smaller. Another problem is that a simple classifier such as softmax activation functions is applied to fully connected layers like conventional neural networks. When optimal features that are useful to distinguish classes could not be extracted from the CNN model with limited training data, satisfactory classification performance may not be obtained by the softmax activation functions (Zhou *et al*., 2017).

To overcome these drawbacks, some researches have been conducted to combine CNN with conventional ML algorithms (Li *et al*., 2019). The conventional CNN architecture is first designed to automatically extract the high-level features. Then, the conventional ML classifier that is more sophisticated than the softmax activation functions is applied to classify the features extracted from CNN. Despite the great potential of this hybrid model, crop classification is still challenging because similar spatial and spectral characteristics between crops make it difficult to extract sufficient features from DL algorithms.

In this study, we investigate the potential of the hybrid model that combines 2D-CNN and RF for crop classification. The 2D-CNN model is employed as a high-level spatial feature extractor. RF, which is relatively insensitive to hyper-parameters compared to another ML algorithm such as SVM, is applied to a classifier. The applicability of this hybrid model for crop classification is illustrated via a case study of crop classification with unmanned aerial vehicle (UAV) imagery acquired in Anbandegi, a major highland Kimchi cabbage cultivation area in Korea.

2. STUDY AREAS AND MATERIALS

Anbandegi, one of three major highland Kimchi cabbage cultivation areas in Korea, was selected as a case study area (Figure 1(a)). Major crops in the study area include highland Kimchi cabbage, cabbage, and potato, and there are some fallow areas. For crop mapping, the UAV imagery acquired on July 27, 2017 was selected because highland Kimchi cabbage shows the highest plant vitality in the late July. Three visible bands from the UAV imagery from a fixed wing drone equipped with a Cannon IXUS/ELPH camera were used as inputs for classification.

Crop parcels in ground truth data based on field surveys (Figure 1(b)) were divided into two independent parcel groups, one for training parcels and the other for reference parcels. To test the classification performance with respect to the change of the number of training data, training data were randomly extracted from the training parcels. Five different proportions with respect to reference data (0.1 %, 0.5 %, 1 %, 3 %, and 5 %) were tested.

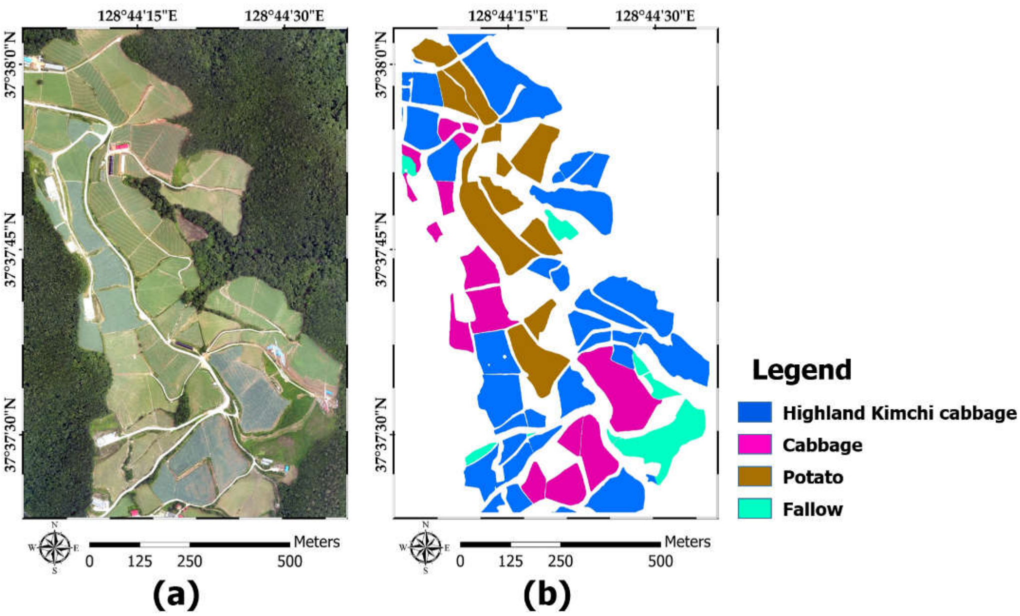

Figure 1. (a) The unmanned aerial vehicle imagery acquired in the study area and (b) ground truth data

3. METHODOLOGY

As a representative DL model, the 2D-CNN model first extracts spatial features and then performs classification using extracted features. It uses spectral information composed of patches as input data to extract spatial features. Suppose that the input data of 2D-CNN consist of $K \times K \times D$ dimensions ($K \times K$ and D denote input patch size and spectral depth, respectively). In this study, K and D were set to 11 and 3, respectively through a preliminary test. First, the 2D-CNN uses convolution layers to extract useful features of neighboring pixels within input patches. The convolution layers extract a lot of features, which are called feature maps or activation maps, through predefined kernels and activation functions. In this study, the kernel size was set to 3 by 3 and rectified linear unit (ReLU) was used as an activation function. Max-pooling is also applied to artificially reduce the spatial dimension of feature maps. A series of these operations are applied to extract useful spatial features. To perform classification, feature maps should be first converted to one dimensional form through flattening, and softmax activation functions are then applied to the flattened fully connected layers. Additional features can be extracted from the fully connected layers through activation functions prior to applying softmax activation functions. The optimal 2D-CNN model architecture was constructed through several preliminary tests and the details of the architecture are given in Table 1.

Table 1. 2D-CNN model structure used in this study

Layer (activation function)	Output dimension	# of parameters
Input layer	(11, 11, 3)	0
2D Convolution_1 (ReLU)	(11, 11, 32)	896
2D Convolution_2 (ReLU)	(11, 11, 32)	9,248
Max pooling_1	(5, 5, 32)	0
2D Convolution_3 (ReLU)	(5, 5, 64)	18,496
Flatten_1	(1600)	0
Dropout_1	(1600)	0
Dense_1 (ReLU)	(64)	12,808
Dense_2 (softmax)	(4)	36
Total trainable parameters: 41,484		

If the number of training data is small, the 2D-CNN model may be overfitted to the small training data and the classification performance may be unsatisfactory. RF is generally known to mitigate overfitting and is not significantly affected by outliers (Li *et al*., 2013). By considering this advantage, RF was selected as a classifier of the final layer (referred to as Dense_2 in Table 1) to improve the classification performance of the conventional 2D-CNN model. To evaluate the classification performance of the proposed hybrid model, we compared overall accuracy statistics of RF, 2D-CNN, and the proposed hybrid model.

4. RESULTS AND DISCUSSION

Figure 2 presents the variation of overall accuracy of classification results with respect to different numbers of training data. Regardless of the number of training data, the proposed hybrid model showed the best classification accuracy. The overall accuracy of pixel-based RF was about 60 % regardless of the variation in the number of training data. As the number of training data increased, the classification accuracy values of 2D-CNN and the proposed hybrid model also increased. The highest overall accuracy was about 80 %. When relatively large numbers of training data were used (e.g., 1 %, 3 %, and 5 %), the difference in overall accuracy between pixel-based and patch-based classification was about 20 %p. When spatial features were extracted from large training data in 2D-CNN (e.g., 5 %), the difference in classification accuracy between conventional 2D-CNN and the hybrid model was not great. In contrast, when fewer training data were used for classification, the overall accuracy of the two patch-based classification models was significantly different. In particular, the difference in overall accuracy between conventional 2D-CNN and the hybrid model was about 10 %p when 0.1% training data were used. The overall accuracy of 2D-CNN was about 10 %p lower than that of the RF model. Based on these results, it could be concluded that enough training data should be used to extract useful spatial features from the 2D-CNN model. The hybrid model presented in this study could effectively classify less informative spatial features.

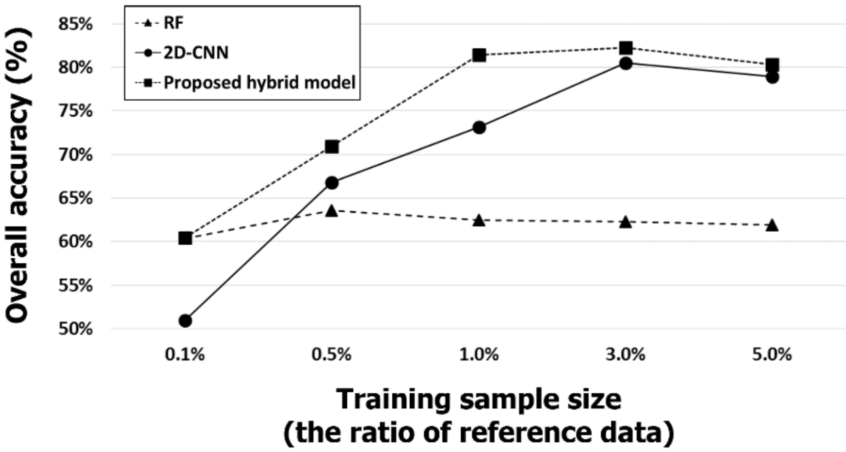

Figure 2. Variation of overall accuracy of three models with respect to the variation of the number of training data.

The classification results are shown in Figure 3. Many isolated pixels were observed in the RF classification result, which is common in pixel-based classification results, regardless of variation in the number of training data. In addition, highland Kimchi cabbage was misclassified as potato or cabbage. In the 2D-CNN classification result, isolated pixels were greatly reduced, compared to the RF classifier, but misclassified pixels tend to be spatially clustered. On the other hand, the proposed hybrid model partially mitigated misclassification patterns that appeared when the number of training data was small.

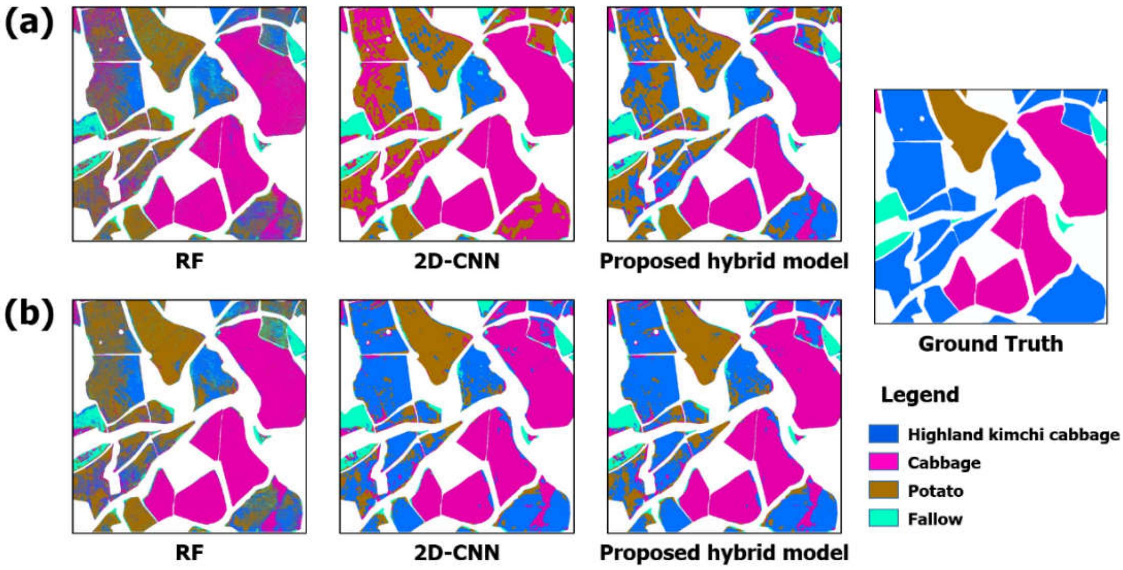

Figure 3. Classification results in sub-areas with significant differences between the three models: the number of training data at (a) 0.1% (the fewest training data) and (b) 0.3% (the highest classification accuracy).

5. CONCLUSIONS

In this study, we proposed the hybrid model that combines 2D-CNN and RF as a feature extractor and a final classifier, respectively, to mitigate overfitting problems that commonly occur in 2D-CNN when small training data are used. From the case study of crop classification in the highland Kimchi cabbage cultivation area, it was found that the proposed hybrid model outperformed conventional 2D-CNN and RF. Particularly, the improvement in classification accuracy of the hybrid model was significant when small training data were used for classification. Extensive experiments in other areas where different crops are cultivated will be carried out to strengthen the potential of the proposed hybrid model.

ACKNOWLEDGEMENT

This study was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01350004)" Rural Development Administration, Republic of Korea.

REFERENCES

Hu, F., Xia, G.S., Hu, J., and Zhang, L., 2015. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing, 7(11), pp. 14680-14707.

Kim, Y., Kwak, G.-H., Lee, K.-D., Na, S.-I., Park, C.-W., and Park, N.-W., 2018. Performance evaluation of machine learning and deep learning algorithms in crop classification: impact of hyper-parameters and training sample size. Korean Journal of Remote Sensing, 34 (5), pp. 811-827 (In Korean with English abstract).

Kwak, G.-H., and Park, N.-W., 2019. Impact of texture information on crop classification with machine learning and UAV images. Applied Sciences, 9(4), pp. 643.

Li, M., Im, J., and Beier, C., 2013. Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntington Wildlife Forest. GIScience and Remote Sensing, 50(4), pp. 361-384.

Li, T., Leng, J., Kong, L., Guo, S., Bai, G., and Wang, K., 2019. DCNR: deep cube CNN with random forest for hyperspectral image classification. Multimedia Tools and Applications, 78 (3), pp. 3411-3433.

Löw, F., Michel, U., Dech, S., and Conrad, C., 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. ISPRS Journal of Photogrammetry and Remote Sensing, 85, pp. 102-119.

Rußwurm, M., and Körner, M., 2018. Multi-temporal land cover classification with sequential recurrent encoders. ISPRS International Journal of Geo-Information, 7(4), pp. 129.

Sidike, P., Sagan, V., Maimaitijiang, M., Maimaitiyiming, M., Shakoor, N., Burken, J., Mockler, T., Fritschi, F.B., 2019. dPEN: deep Progressively Expanded Network for mapping heterogeneous agricultural landscape using WorldView-3 satellite imagery. Remote Sensing of Environment, 221, pp. 756-772.

Wei, S., Zhang, H., Wang, C., Wang, Y., and Xu, L., 2019. Multi-temporal SAR data large-scale crop mapping based on U-Net model. Remote Sensing, 11(1), pp. 68.

Zhong, L., Hu, L., and Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote sensing of Environment, 221, pp. 430-443.

Zhou, L., Li, Q., Huo, G., and Zhou, Y., 2017. Image classification using biomimetic pattern recognition with convolutional neural networks features. Computational Intelligence and Neuroscience, 2017, pp. 1-12.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., and Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 5(4), pp. 8-36.