APPLICATION OF DEEP LEARNING ALGORITHMS CONSIDERING SPATIO-TEMPORAL FEATURES FOR CROP CLASSIFICATION

Min-Gyu Park, No-Wook Park

Inha University, 100 Inha-ro, Michuhol-gu, Incheon 22212, Korea
Email: mingyu9321@inha.edu; nwpark@inha.ac.kr

KEY WORDS: Classification, Deep learning, Image sequence, Feature map

ABSTRACT: The purpose of this study is to compare deep learning models that consider characteristics of crops in the classification of multi-temporal and high spatial resolution images. We applied 2D-convolutional neural network (2D-CNN) and long short-term memory (LSTM) for crop classification to consider spatial and temporal features, respectively. In addition, 3D-CNN and convolutional LSTM (Conv-LSTM), which can account for both temporal and spatial features, were also applied and compared. From a case study of crop classification with multi-temporal unmanned aerial vehicle images, Conv-LSTM showed the best classification accuracy thanks to its ability to account for both spatial and temporal features. Since the growth cycles of crops should be properly considered for crop classification, LSTM-based models including LSTM and Conv-LSTM are more effective than CNN-based models.

1. INTRODUCTION

Remote sensing data have been widely used in various fields such as agriculture, environment, and weather. Particularly, in agriculture, a crop classification map has been regarded as one of important information sources for crop yield modeling and prediction. Since each crop has its own growth cycle, multi-temporal images have been used for crop classification to fully account for temporal characteristics of crops (Ienco et al., 2017; Sharma et al., 2018). Satellite images with different spatial resolutions are usually used as inputs for crop classification, depending on the scales of crop fields (Bohler et al., 2018). From a viewpoint of data availability, it is often difficult to acquire multi-temporal optical images in major crop growth periods due to weather conditions. Recently, there is a growing interest in unmanned aerial vehicle (UAV) images, since UAV images are less affected by weather conditions and can be easily acquired at the desired times (Jeon and Kim, 2018; Kwak and Park, 2019). Since UAV images are acquired at ultra-high spatial resolutions, the number of detectable objects usually increases. When ultra-high spatial resolution images are used for classification, however, the spectral variability of classes of interest tends to increase. If conventional pixel-based classification algorithms that do not use spatial contextual information from neighboring pixels are applied to classification of UAV images, very noisy classification results are usually obtained (Castelluccio et al., 2015). In addition, most classification algorithms are not designed for a proper consideration of temporal features (Wu and Prasad, 2017; Zhong et al., 2019). Therefore, when multi-temporal UAV images are used for crop classification, it is necessary to consider a classification model that can account for both temporal and spatial features to obtain reliable classification results. Recently, as deep learning models, convolutional neural network (CNN) and recurrent neural network (RNN) have been widely applied to classification (Ruβwurm and Körner., 2017; Lyu et al., 2016). However, the comparison of classification performance of various deep learning models that consider temporal and/or spatial features has not been fully conducted for crop classification.

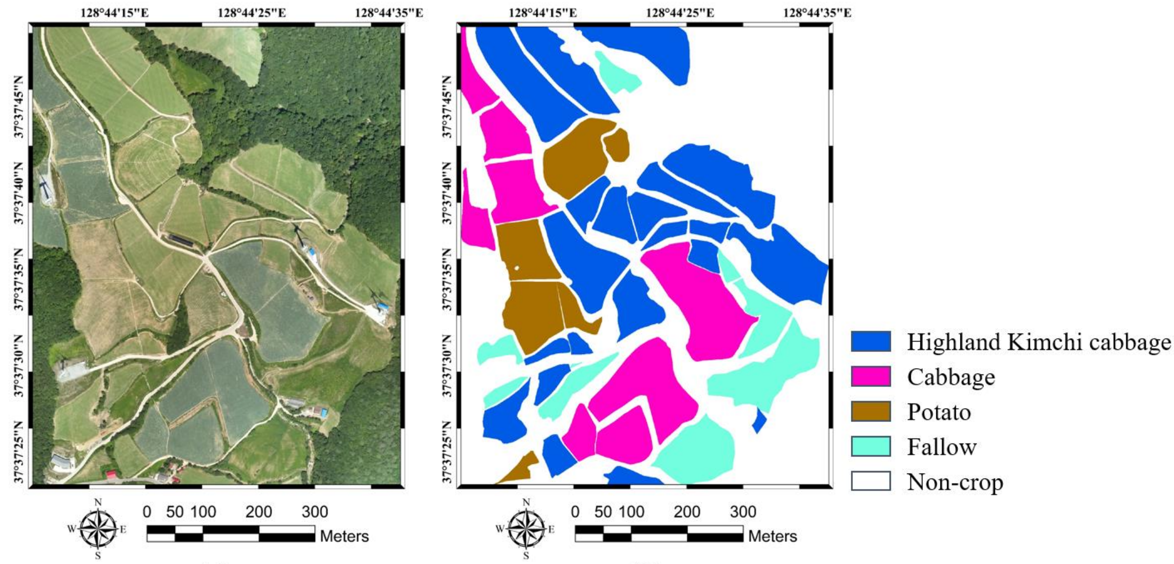
In this paper, the classification performance of deep learning models is compared for crop classification with multi-temporal UAV images. We compared four deep learning models including 2-D CNN, long short-term memory (LSTM), 3-D CNN, and convolutional LSTM (Conv-LSTM). A case study of crop classification in Anbandegi, Korea was carried out to quantitatively evaluate the classification performance.

2. STUDY AREA AND DATA

The case study site is Anbandegi in Gangneung, Korea, which is the major highland Kimchi cabbage cultivation area. For supervised classification, we considered four classes including highland Kimchi cabbage, cabbage, potatoes, and fallow (Table 1).

Table 1. Crop classes and their respective area and proportion in the study area

Class	Area (m^2)	Proportion (%)
Highland Kimchi cabbage	125,168.51	44.69
Cabbage	73,373.99	26.20
Potato	37,379.76	13.34
Fallow	44,172.55	15.77
Total	280,094.81	100



(a) (b)

Figure 1. (a) UAV imagery (Aug. 1st, 2018) and (b) ground truth.

To consider the grow cycles of crops in the study area, nine UAV images acquired from May to October 2018 were used as inputs for classification (Figure 1). The UAV images were taken using Canon IXUS / ELPH with red, green, and blue bands and had a spatial resolution of 0.5m. In the study area, cabbage and potato, and highland Kimchi cabbage are planted in June and July, respectively. Potato is harvested in August, while highland Kimchi cabbage and cabbage are harvested in September. Ground truth crop types provided by National Academy of Agricultural Sciences were partitioned into two spatially independent parcel groups to avoid spatial correlation between training and reference data (Sharma et al., 2018; Audebert et al., 2019). 0.5% of training parcels were randomly selected as training pixels for supervised classification.

3. CLASSIFICATION MODEL

In this study, 2D-CNN using spatial features and LSTM using temporal features were applied to consider to the characteristics of crop in multi-temporal images. In addition, 3D-CNN, which extends the spectral dimension to spectral-temporal dimension, and Conv-LSTM, which conveys spatial features of input data to LSTM were applied.

3.1 CNN

CNN has been known to be more effective for extracting spatial feature than traditional classification algorithms because CNN uses spatial information around pixels. In general, the CNN structure includes two types of typical layers, a convolutional layer and a pooling layer (LeCun et al. 2015). The convolutional layer generates various feature maps by applying a convolution operator to each pixel in order to extract spatial features. The pooling layer is used to reduce or abstractly emphasize the dimensions of feature maps generated by convolutional filters (Kim et al., 2018). Unlike traditional classification algorithms that use pixel-based input data, CNN uses patch-based input data. As shown in Figure 2, the CNN that considers the spectral dimension in two-dimensional input data is generally called 2D-CNN, and the CNN that extends from only spectral dimension to the temporal and spectral dimension is called 3D-CNN. In other words, 2D-CNN does not distinguish spectral and temporal dimensions, but 3D-CNN can consider temporal features by dividing spectral and temporal dimensions (Ji et al., 2018).

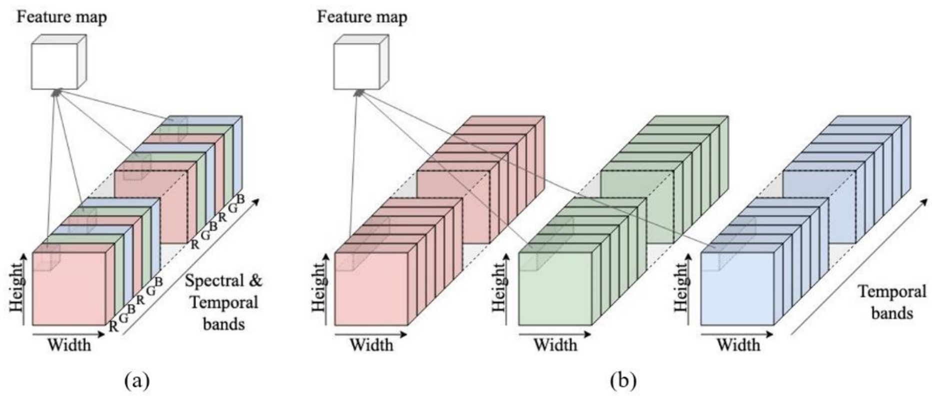


Figure 2. Process of extracting feature maps in (a) 2D-CNN and (b) 3D-CNN

3.2 LSTM

Unlike CNN or conventional machine learning, RNN use the output in time t-1 with the next input, to feed itself at time t. Therefore, it is suitable to analyze time series data by clearly managing temporal dependencies. However, RNN has long-term dependencies when it has deeper model or long input sequences, causing vanishing gradient problems in back-propagation and thus makes it difficult to learn for long-term dependencies. To overcome this problem, LSTM with improved learning efficiency uses more complicated functions that learn to control the flow of information, allowing the recurrent layer to capture long-term dependencies more easily (Wu and Prasad, 2017).

LSTM consists of two stats including cell states that convey the entire sequence, and hidden states that update new information by each gate (Shi et al., 2015). As shown in Figure 3 (a), the output (C_{t-1}, H_{t-1}) at time t-1 is used together with the input of time t controlled by gates. So, LSTM can take advantage of the temporal features in time series data. Since LSTM cannot use spatial information, however, it receives pixel-based data unlike patch-based CNN. To solve this problem, Conv-LSTM was applied by replacing a pointwise operator (x) in the network with convolutional operator(*) as shown in Figure 3 (b) (Liu et al., 2015). By using advantages of conventional LSTM that can consider characteristics of time series, thus, it is possible to simultaneously learn spatio-temporal features.

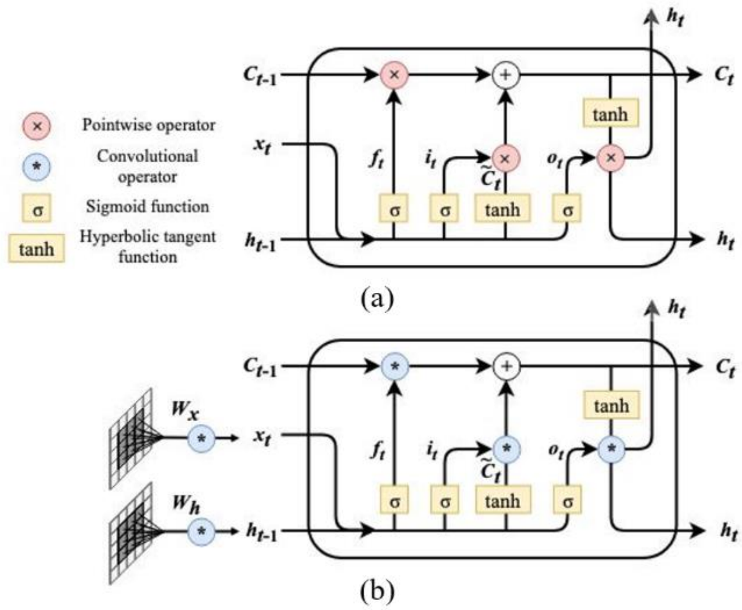


Figure 3. Structure of the (a) LSTM and (b) Conv-LSTM unit.

4. RESULTS AND DISCCUSSION

The classification results of four different deep learning models are presented in Figure 4. In all classification results, misclassification is dominant in boundaries of parcels. For quantitative comparisons, statistical measures from confusion matrix including overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) were computed and compared (Table 2). As shown in Figure 4, fallow was misclassified as highland Kimchi cabbage and potatoes in both 2D-CNN and 3D-CNN. The PA and UA of fallow in 2D-CNN and 3D-CNN are relatively lower than LSTM-based models (Table 2). Unlike crops in the study area, fallow has no specific seeding or harvesting period, so they may be misclassified as other crops depending on its spectral characteristics. When comparing two CNN-based models, 3D-CNN, which additionally uses information from adjacent temporal bands, showed slightly improved classification accuracy as a result of the reduction of misclassification of high-land Kimchi cabbage.

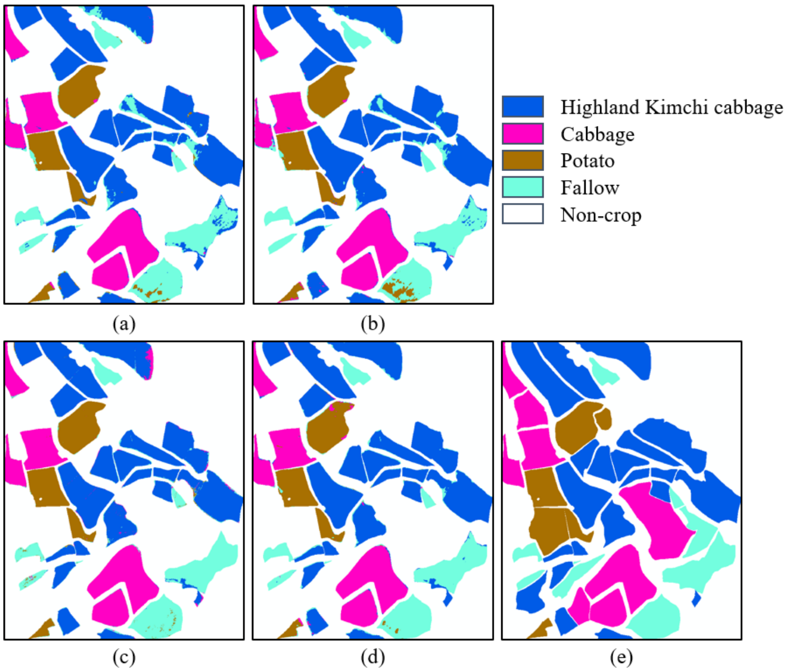


Figure 4. Classification results: (a) 2D-CNN, (b) 3D-CNN, (c) LSTM, (d) Conv-LSTM, and (e) ground truth.

LSTM-based models showed improved classification accuracy statistics in all classes because they can consider unique spectral patterns of different crops per image acquisition dates. Conv-LSTM showed the best OA, but the difference in OA between LSTM and Conv-LSTM is not great. Unlike CNN-based models, when spectral patterns at a certain time are different from growth cycles, LSTM can ignore those patterns that may degrade classification performance, leading to superior classification performance. If crops have similar growth cycles such that cabbage is sown at a time similar to that of potato and harvested with highland Kimchi cabbage at a similar time, other crops may be misclassified as cabbage on some parcels, as shown in Figure 4 (c) and (d).

Table 2. Summary of classification accuracy statistics

		2D-CNN	3D-CNN	LSTM	Conv-LSTM
PA	Highland Kimchi cabbage	93.15	94.79	95.09	**96.64**
	Cabbage	97.46	96.74	**98.99**	98.21
	Potato	95.79	95.86	**98.96**	95.33
	Fallow	89.74	89.06	96.52	**96.85**
UA	Highland Kimchi cabbage	96.62	97.96	**99.32**	98.85
	Cabbage	**98.86**	98.55	95.68	98.10
	Potato	93.38	91.20	96.16	**98.30**
	Fallow	80.84	82.02	**90.40**	88.79
OA		93.80	94.38	96.59	**96.84**

The classification patterns in a certain sub-area are shown in Figure 5. In the LSTM classification result, as a pixel-based classifier, misclassification occurs in a pixel unit, while other models showed clustered misclassified pixels, because spatial features have been used for classification. Particularly, since CNN extracts features from all layers, misclassification is prominent as shown in Figure 5 (a) and (b) when the peculiar features appearing in some periods are strongly reflected. On the other hand, Conv-LSTM could reduce the misclassification around the parcel boundary and of isolated pixels, but misclassification of highland Kimchi cabbage as follow still occur. When comparing LSTM-based models, Conv-LSTM could overcome the limitation of pixel-based LSTM and showed the best OA and PA of highland Kimchi cabbage that is a major crop type. Based on these results, it can be concluded that Conv-LSTM that combines spatial features of CNN and temporal correlation of LSTM is a promising deep learning model in the study area when multi-temporal UAV images are used for crop classification.

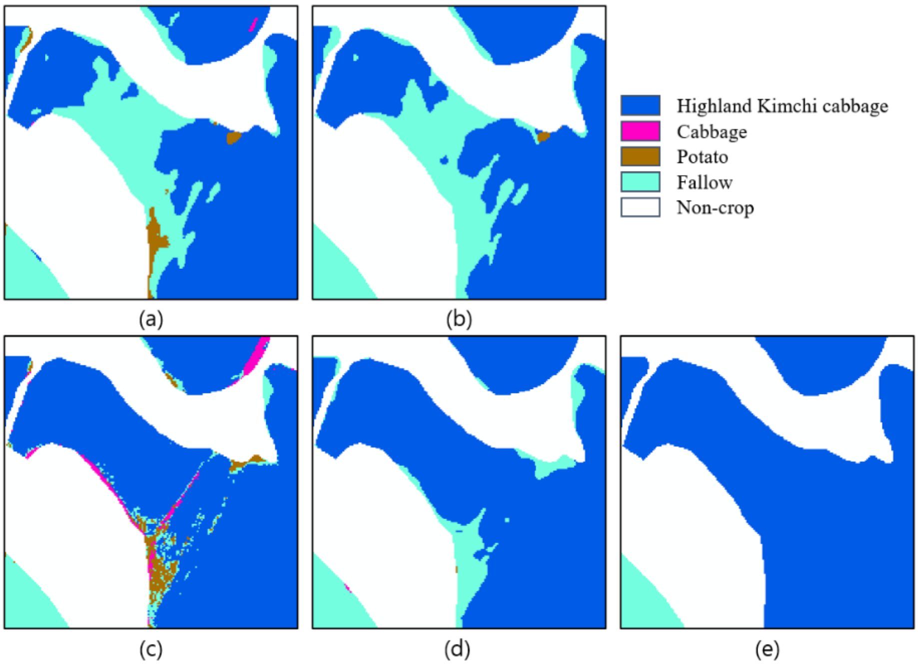


Figure 5. Comparison of misclassification patterns in a sub-area: (a) 2D-CNN, (b) 3D-CNN, (c) LSTM, (d) Conv-LSTM, and (e) ground truth.

5. CONCLUSION

In this study, we compared four different deep learning models for crop classification that consider spatial and temporal features separately or together. As the models that consider temporal and spatial features, respectively, LSTM and 2D-CNN were applied and compared hybrid models that consider both features together including 3D-CNN and Conv-LSTM. From a case study in crop classification with multi-temporal UAV images, we found that Conv-LSTM showed the highest OA, followed by LSTM, 3D-CNN, and 2D-CNN, but the difference in classification accuracy between Conv-LSTM and LSTM was not great. Conv-LSTM could improve the classification accuracy by combining the respective advantages of 2D-CNN and LSTM. In addition, the classification performance of LSTM-based models that can convey sequential information was superior to that of CNN-based models for crop classification. To strengthen the results of this study, more experiments in other regions and the development of an advanced model that can combine spatial features from different sizes and shapes of crop parcels with temporal features will be carried out in the future.

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. The authors thank National Institute of Agricultural Sciences for providing UAV images and ground truth data used in this study.

REFERENCES

Audebert, N., B. Le Saux, and S. Lefèvre, 2019. Deep learning for classification of hyperspectral data: A comparative review. IEEE Geoscience and Remote Sensing Magazine, 7 (2), pp. 159-173.

Böhler, J., M. Schaepman, and M. Kneubühler, 2018. Crop classification in a heterogeneous arable landscape using uncalibrated UAV data. Remote Sensing, 10 (8), 1282.

Castelluccio, M., G. Poggi, C. Sansone, and L. Verdoliva, 2015. Land use classification in remote sensing images by convolutional neural networks. arXiv preprint arXiv: 1508:00092.

Ienco, D., R. Gaetano, C. Dupaquier, and P. Maurel, 2017. Land cover classification via multitemporal spatial data by deep recurrent neural networks. IEEE Geoscience and Remote Sensing Letters, 14 (10), pp. 1685-1689.

Jeon, W. and Y. Kim, 2018. An assessment of a random forest classifier for a crop classification using airborne hyperspectral imagery. Korean Journal of Remote Sensing, 34 (1), pp. 141-150.

Ji, S., C. Zhang, A. Xu, Y. Shi, and Y. Duan, 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sensing, 10 (1), 75.

Kim, Y., G.-H. Kwak, K.-D. Lee, S.-I. Na, C.-W. Park, and N.-W. Park, 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), 811-827, (In Korean with English abstract).

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

LeCun, Y., Y. Bengio, and G. Hinton, 2015. Deep learning. Nature, 521 (7553), 436.

Liu, Q., F. Zhou, R. Hang, and X. Yuan, 2017. Bidirectional-convolutional LSTM based spectral-spatial feature learning for hyperspectral image classification. Remote Sensing, 9 (12), 1330.

Lyu, H., H. Lu, and L. Mou, 2016. Learning a transferable change rule from a recurrent neural network for land cover change detection. Remote Sensing, 8 (6), 506.

Rußwurm, M. and M. Körner, 2017. Temporal vegetation modelling using long short-term memory networks for crop identification from medium-resolution multi-spectral satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 11-19.

Sharma, A., X. Liu, and X. Yang, 2018. Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. Neural Networks, 105, pp. 346-355.

Shi, X., Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems, pp. 802-810.

Wu, H. and S. Prasad, 2017. Convolutional recurrent neural networks for hyperspectral data classification. Remote Sensing, 9 (3), 298.

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