UNDERSTANDING MULTISPECTRAL SATELLITE IMAGE LAND COVER CLASSIFICATION WITH THE DEEP LEARNING

Wei-Zhen Lin[1], Chih-Yuan Huang[1]
[1]Center for Space and Remote Sensing Research,
National Central University, Taoyuan 320, Taiwan,
Email: alice772365@g.ncu.edu.tw, cyhuang@csrsr.ncu.edu.tw

ABSTRACT: As satellite images provide periodical observations of a large area, Remote Sensing (RS) data is important for analyzing Land Use and Land Cover (LULC). In recent years, with the advancement of Deep Learning, better LULC classification and prediction are achieved via artificial neural networks (ANNs), such as convolutional neural networks (CNNs). As land cover classifications usually apply spatial and spectral properties, the objective of this research is to design a deep learning network to retrieve spatial and spectral features from remote sensing images and perform the land cover classification. To be specific, a CNN for extracting textural information is combined with a network extracting cross-band spectral relationships for the final classification. The thirteen bands of EuroSAT dataset are applied in this research, where more than 90% accuracy can be achieved. An ongoing work is to identify important factors for a better understanding to the remote sensing classification, and compare with current neural network achieves state-of-the-art results on EuroSAT dataset with 98.65% accuracy.

KEY WORDS: Satellite image, Land cover classification, Deep Learning

1. INTRODUCTION

Artificial Intelligence (AI) collects knowledge and information to build computational models that are capable of solving problems, learning, recognizing, classifying, self-improvement, and reasoning. With the improvement of AI methods in recent years, the Deep Learning is the latest progress and has been applied in various fields. The Deep Learning constructs Deep Neural Networks (DNNs) such as Convolutional Neural Networks (CNNs) with multilayer nonlinear structures designed to identify relationships between inputs and outputs.

Currently, the Land Use and Land Cover (LULC) classification in the Remote Sensing field, CNNs are usually the best method to identify and extract the spatial and spectral features within satellite images. In this paper will also utilize the structure of CNN to retrieve the spatial and spectral information provided by satellite image.

As in the domain of data science, DNNs have been successfully applied in various fields and have provided astonishing results. To describe complex behaviors of real world phenomena, more complex and non-linear models have been designed. The models contain a series of mathematical calculations that cannot be easily traced by a human being. Without suitable explanatory mechanisms, DNNs are considered as "black boxes", whose internal inference processes are neither known to the observer nor interpretable by humans (Guidotti et al., 2019). In order to address this issue, Explainable Artificial Intelligence (XAI) methods were proposed to unlock the black boxes and even visualize the information learned by a DNN model. Hence, the contribution distribution within the process and any error can be identified.

1.1 Explainable Artificial Intelligence (XAI)

There are already various AI applications around us creating a convenient and efficient world. However, AI still has the barrier of explainability, the XAI focusing on "explaining" AI is the new trend to solve this kind of issue. Even the accuracy of the results performed well, the feature information learned by AI model may not be correct because of without any understanding about the work principle of AI decision-maker. E.g., the AI model classified target with high accuracy while with wrong classification basis (Lapuschkin et al. 2016), the decision made by AI will turn into untrustworthy, and when training beyond the scope of original training data, the results might demonstrate incorrectly.

By backpropagating relevance through a DNN, XAI can identify the importance levels of inputs. When applying XAI on a CNN, a visualization indicating the general locations on the input image that contributes more information for the decision-making, e.g., saliency maps or heatmaps. With the saliency maps, XAI can provide qualitative analyses on the inferencing process based on these intuitive visualizations. Since XAI covers the techniques used to convert a non-interpretable model into an explainable model. XAI will become increasingly important to all groups of stakeholders, including the users, the affected people, and the developers of AI systems. XAI can help users further evaluate their models more than standard performance metrics, and reveal biases in the trained dataset, classes, multiple labels.

1.2 Objectives

In order to perform satellite image classification, the first objective in this paper is to design a CNN-based model to extract spatial features from multispectral images and perform land cover classification with the extracted spatial and spectral information. On the other hand, the second objective is to utilize XAI methods on the constructed model for a better understanding to the important information in multispectral satellite image classification.

2. METHODOLOGY AND STUDY DATA

The structure of CNN with the convolutional operation were utilized in this research. Taking the 64x64 shape of patch with thirteen spectrums as the inputs and going through with 32 convolutional kernels. By sliding 32 different convolutional kernels on the input image, generated 32 different feature maps in first layer. Following four hidden layers with 64 convolutional kernels produced 64 feature maps within each hidden layer.

In each hidden layer after forward convolutional computation, processing with the maximum pooling operation by dividing the input image into several sub-areas, and only keep the maximum value as one output for each sub-area. The maximum pooling operation is a non-linear downsampling that could reduce the amount of computational pixel and cause the size of the processing image smaller while still could preserve spatial information in next layer. Ultimately connected to the flatten layer for the purpose of generating the value of the classification results of ten labeled classes. Then putting the validating data into the constructed model to generate the prediction results. And illustrating the classification accuracy with the confusion matrix via comparing the results between the prediction and the original labeled.

Through constructing the satellite image classification model, applying the XAI method so as to evaluate the model performance. By calculating the backpropagation algorithm, the output given by the XAI method could derive the explanations including important feature or pixel represent the basis of classification. In this research utilized the Guided Grad-CAM method (Selvaraju et al., 2017), which is the element-wise product of the Guided Backpropagation method with the Grad-CAM method. Backward computing the heatmap layer by layer from the final feature map to the first input layer.

2.1 Datasets

In this paper, employing remote sensing multispectral image dataset for land use classification, which is the “EuroSAT dataset” that contain thirteen spectrums from RGB to Short Wave Infrared of sentinel-2 satellite images. The amount of satellite image is with 27000 labeled samples and have been classified into 10 different LULC classes, including “Industrial Buildings”, “Residential Buildings”, “Annual Crop”, “Permanent Crop”, “River”, “Sea & Lake”, “Herbaceous Vegetation”, “Highway”, “Pasture” and “Forest”. Each class contains 2000 to 3000 images equally. Bands with a lower spatial resolution have been upsampled to 10 meters per pixel using cubic-spline interpolation (De Boor et al., 1978). The “EuroSAT dataset” have been split into training, testing and validating dataset individually in this research.

2.2 Guided Grad-CAM

Nowadays, researchers have already applied different XAI methods on their own research. While some XAI methods deliver better metric performance, the shortcoming of low spatial resolution still could be the potential issue for further fine-grained feature extraction. Here utilized one of the XAI method, namely “Guided Grad-CAM” (Selvaraju et al., 2017) that has better spatial resolution of the heatmap to extract the distribution of features within input image (Iwana et al., 2019).

The philosophy of the Guided Grad-CAM method combines two XAI methods, including the Grad-CAM method and the Guided Backpropagation method in order to retrieve better spatial resolution of the heatmap. The Grad-CAM method (Selvaraju et al., 2017) is the extension of the CAM-based method, computes the gradients of the output compare to feature map activations, then the gradients are multiplied by the layer activations. Appling ReLU activation function, returning only non-negative attributions and using bilinear interpolation in order to upsample back to the original shape of input image (Kakogeorgiou and Karantzalos, 2021). The other one, the Guided Backpropagation method (Springenberg et al., 2015) that combine the backpropagation and the deconvolution mechanism (Zeiler and Fergus, 2014). The Guided Backpropagation computes the gradients of the output compare to the input, which only keep two non-negative gradients derived from two mechanisms in order to preserve and strengthen the positive gradient distribution. The Grad-CAM method could extract the coarse localization, as for the Guided Backpropagation could contain with the fine-grained details. Multiplied two outputs within the Grad-CAM and the Guided Backpropagation methods that could retrieve the heatmap of the Guided Grad-CAM to emphasize the important information or pixel location in input layer.

3. RESULTS

Constructing the LULC classification model with EuroSAT dataset and the accuracy of the classification results achieved about 92%. Applying the Guided Grad-CAM method, retrieved the importance classification pixel with precise location. In Figure 1 displays the river example with RGB bands in EuroSAT dataset, and with the heatmap output of the Guided Grad-CAM in Figure 2, visualized the importance information with yellow parts. The result shows that the model classified with correct pixel location and robust the model classification accuracy.

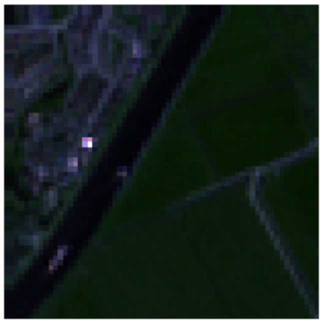

Figure 1 River example with RGB bands in EuroSAT dataset

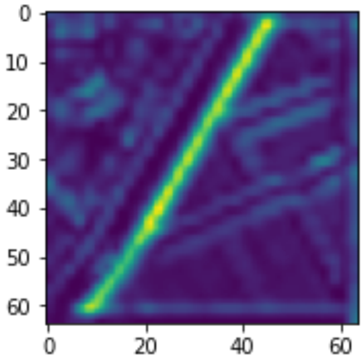

Figure 2 Heatmap of river example after applying Guided Grad-CAM method

4. CONCLUSIONS AND FUTURE WROK

The aim of the LULC classification is to automatically provide labels describing the represented physical land type or how a land area is used. When handling RS image segmentation with Deep Learning based methods, the goal is to achieve the spectral and spatial features representation.

XAI technique provides the interpretation of the features within satellite images. Explainable AI set a novel workflow to make AI from black boxes to glass boxes. The “explainable” process provided the results that are understandable for developers and users, clear to identify the internal inference process. The following work is to extract the spatial and spectral texture in CNN, search inner relationship within the contribution visualization of the XAI model and focus on spectral information extraction from XAI methods for better interpretation.

REFERENCE

Gómez, P., & Meoni, G. (2021). MSMatch: Semisupervised Multispectral Scene Classification With Few Labels. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 11643-11654.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-42.

Lapuschkin, S., Binder, A., Montavon, G., Muller, K. R., & Samek, W. (2016). Analyzing classifiers: Fisher vectors and deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2912-2920).

De Boor, C., & De Boor, C. (1978). A practical guide to splines (Vol. 27, p. 325). New York: springer-verlag.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).

Iwana, B. K., Kuroki, R., & Uchida, S. (2019, October). Explaining convolutional neural networks using softmax gradient layer-wise relevance propagation. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) (pp. 4176-4185). IEEE.

Kakogeorgiou, I., & Karantzalos, K. (2021). Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. International Journal of Applied Earth Observation and Geoinformation, 103, 102520.

Springenberg, J. T., Dosovitskiy, A., Brox, T., Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806.

Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham.