

RESEARCH ON LANDCOVER CLASSIFICATION SOLUTIONS FROM OPTICAL REMOTE SENSING IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

Dinh Sinh Mai^{1,*}, Le Hung Trinh¹, Khanh Hoai Dao¹

¹Le Quy Don Technical University, Hanoi, Vietnam

DOI: 10.56651/lqdtu.jst.v6.n01.664.sce

Abstract

With the explosion of data and information, data analysis techniques are becoming increasingly important. Remote sensing image data has many advantages in analyzing features of the earth's surface due to its wide coverage and frequent updating. However, with the strong development of remote sensing technology, the source of remote sensing image data has increased rapidly. This requires powerful tools to solve problems on this data source. In recent years, artificial intelligence is developing rapidly and is one of the extremely powerful data analysis tools in many fields, including remote sensing image data. In this article, we present a solution to apply the convolutional neural network (CNN) to classify landcovers from remote sensing image data. The landcover classification experiment shows that the application of the convolutional neural network model can give classification results with an accuracy of over 95%. Meanwhile, the RF model only gives results with less than 90% accuracy. This result shows great potential for the application of deep learning models in remote sensing image analysis.

Keywords: Convolutional neural networks; deep learning; remote sensing images; object classification.

1. Introduction

Remote sensing image data is usually acquired by many sensors, which can be from optical sensors (multispectral and hyperspectral images), lidar sensors, and radar sensors. Remote sensing image data is geolocated or pixels are naturally located in geospatial [1]. Each pixel corresponds to spatial coordinates, facilitating the association of pixel information with other data sources, such as geographic information systems, and geotagged images from media or sensors, etc. This allows solving data consolidation with non-traditional data methods. On the other hand, it will pave the way for new applications such as service-based map digitization. Remote sensing data also contains geodetic data such as coordinate system, terrain elevation information [2].

Remote sensing image data also faces the “big data” challenge. In an age of information explosion, we are dealing with growing volumes of data, often on a global

* Email: maidsinh@lqdtu.edu.vn

scale [3]. For example, with Sentinel-1 imagery, which began to be collected in 2014, approximately 25 Pb (Peta Bytes) of data were acquired by 2017 $\sim 25 \times 1,024$ TB $\sim 26,214,400$ GB, and this number has grown significantly since then. This requires algorithms to be fast enough to solve problems on such massive datasets.

The Sentinel-2 mission consists of two satellites developed to support vegetation, landcover, and environmental monitoring. The Sentinel-2A satellite was launched by ESA on June 23, 2015, and operates in a sun-synchronous orbit with a 10-day repeat cycle. A second identical satellite (Sentinel-2B) was launched on March 7, 2017. Together they cover all Earth's land surfaces, large islands, and inland and coastal waters every five days. Remote sensing image data is continuously acquired according to the satellite's rotation period. This capability enables individual image analysis to time series image processing. Therefore, the network architecture must be developed to be able to optimally exploit the temporal, spatial and spectral information of these data.

Deep learning is the fastest-growing trend in big data analytics and is considered as one of the ten disruptive technologies of 2013 [4]. Deep learning is characterized by artificial neural networks, which usually involve more than two hidden layers (hence they are called deep learning). A convolutional neural network (CNN) is a specific type of deep neural network [5]. A convolutional neural network has an input layer, an output layer, and many different hidden layers. Hidden layers include types such as convolutional layer, dimensionality reduction layer, data correction layer, normalization layer, and fully connected layer, etc. In which, the convolution layer is used to create the connection between adjacent layers in a small area, limited to a local area [6]. This greatly reduces the computation of the transfer functions between layers while maintaining the connections between neurons to extract features of the data in later layers of the network [7].

In recent times, there have also been a number of studies applying deep learning networks in landcover classification from UAV images [8]. Especially recently, when UAV technology is developing strongly, the research and development of UAV image processing methods is necessary and has great potential [9, 10]. Recent studies have shown that the attribute representations learned by CNNs are capable of giving high accuracy in large image recognition [11], in object detection [12]. The article researches a solution to apply convolutional neural network in remote sensing image classification to take advantage of CNN in large data processing. Overcoming the disadvantages that traditional classification techniques are difficult or impossible to handle.

The article is organized into 4 parts: Part 1 is general introduction; Part 2 presents background knowledge, and the theoretical basis of the convolutional neural network and remote sensing image classification problem; Part 3 is experimental research; Section 4 presents conclusions and recommendations in the future.

2. Theoretical foundations

2.1. Convolutional Neural Network

Convolutional Neural Networks (CNN) is one of the most popular and most influential deep learning models in the computer vision community [13]. CNN is used in many fields such as image classification, image recognition, video analysis, or natural language processing.

Image classification is one of the most important techniques in the field of image processing. There are many studies to solve this problem such as feature extraction (popular as SIFT, HOG), clustering-based data mining techniques, etc. However, these methods show that the output is not really effective on large images such as remote sensing images. The classification model will generally have to predict the class of objects from the matrix of pixels, such as plants, water or roads, etc.



Figure 1. Image and image data matrix.

To represent an image with a resolution of 256×256 pixels (Figure 1), we need a matrix that will be 256×256 in size, and depending on whether the image is color or multispectral, this matrix will have the number of bands corresponding to the number of spectral bands. For example, with an RGB color image, size 256×256 , we will have a $256 \times 256 \times 3$ matrix to represent this image.

The CNN network model includes a set of basic layers including convolution layer + nonlinear layer, pooling layer, fully connected layer. These classes are linked together in a certain order. Normally, the image data will be propagated through the first

convolution layer + nonlinear layer, then the calculated values will propagate through the pooling layer. The triple convolution layer + nonlinear layer + pooling layer can be repeated many times in the network. And then it is propagated through the fully connected layer and finally using the softmax function to calculate the probability of the output image layer.

Convolution layer

The convolution layer is the most important and also the first layer of the CNN model. This class has the main function of good detection of spatial features. There are four main objects: input matrix, filters, and receptive field, feature map. The convolution layer takes as input an n -dimensional matrix (n is the number of attributes) and a set of filters to learn. This filter will slide through each position on the image to calculate the convolution between the filter and the corresponding part of the image. This corresponding part of the image is called the receptive field. The area where a neuron can be seen to make a decision, and the matrix generated by this process is called the feature map.

Pooling layer

After the activation function, we normally use the pooling layer. Some common types of pooling layers are max-pooling, average pooling, whose main function is to reduce the dimension of the previous layer. With a 2×2 pooling window, you need to slide this 2×2 filter over similar sized areas and then calculate the max, or average, for that area. The implication behind the pooling layer is that the relative position of the features in the image space is no longer necessary, but instead the relative position of the features is sufficient for object classification. Furthermore, reducing the pooling layer has the potential to reduce the dimensionality greatly. From there can limit overfit, and reduce good training time.

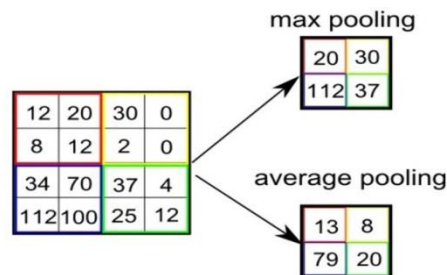


Figure 2. Pooling process.

Fully connected layer

The last layer of the CNN model in the image classification problem is the fully connected layer, also known as the classification layer to produce the output results.

This layer has the function of converting the feature matrix in the previous layer into a vector containing the probabilities of the objects that need to be predicted. For example, landcover classification problem with 10 classes corresponding to 10 numbers from 0 - 9, the fully connected layer will convert the feature matrix of the previous layer into a 10-dimensional vector representing the probabilities of the 10 corresponding classes.

And finally, the process of training the CNN model for the image classification problem is similar to the training of other models. We need an error function to calculate the error between the model's prediction results and the labeled data, as well as use the backpropagation algorithm for the weight update process.

2.2. Convolutional neural networks in remote sensing image analysis

In the field of remote sensing, CNN networks have been interested in applications in many problems because they have the ability to work well with large and multi-dimensional data. In this section, the article will present the method of using CNN network for remote sensing image classification problem. Instead of passing in the model the entire image (the object is mixed with the background). We start with the idea that an object is the same no matter where it is in the image, and divide the image into a grid.

Step 1: Split the image into small pieces

Split the original image into grid cells and store the result in pieces. By doing this, we have converted the original image into small pieces of the same size.

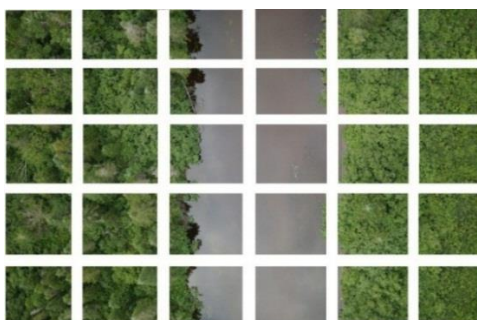


Figure 3. Splitting an image into small pieces of the same size.

Step 2: Pass each image piece into the small neural network

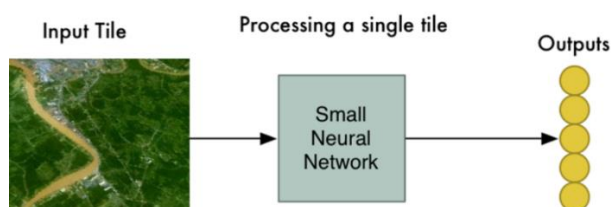


Figure 4. Passing each image piece into a small neural network.

In the past, traditional network models would transfer the entire image to the neural network. However, this approach can lead to network overload. With CNN, we will transmit the image piece by piece and iterate with all the image fragments. CNN uses the same neural network weight for all the image fragments to obtain the output data.

Because of the same weights, the input data is different, so the output is different. The purpose of this section is to find special features of the image. In a large sized image, it is difficult to identify an object. However, when separating the image into many different pieces to find, no matter where the object is on the image, it still has the same feature as a small area of that object.

Step 3: Store the result from each image piece into the array

In order not to lose the sequential arrangement of the original image fragments. We can store the result of processing each piece into a grid like the original image, as shown below:

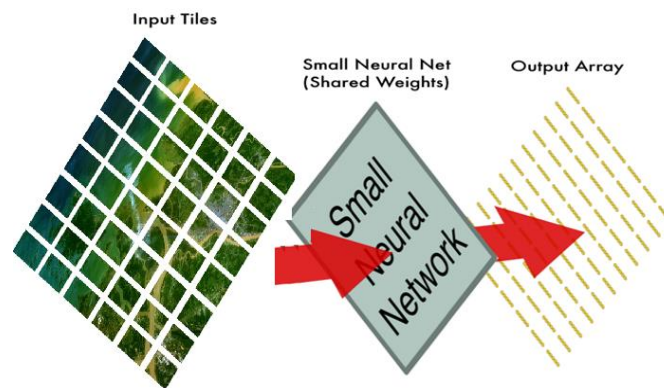


Figure 5. Storing the result from each image piece into an array.

We start with a large image, and end up with a lot of smaller image pieces. These image fragments store notable parts from the original image. From step 1 to step 3 is called convolutional process - take the output value with each image piece, through fixed weight.

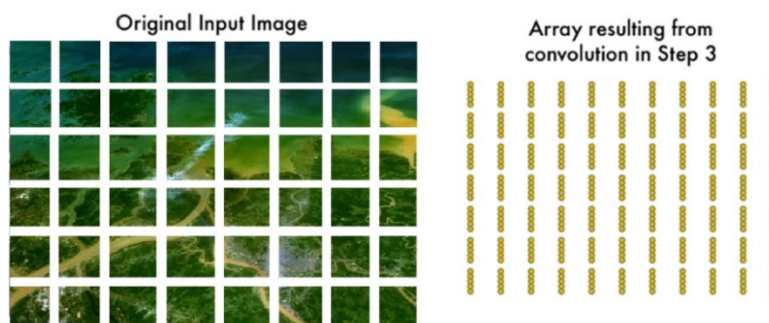


Figure 6. Storing the result from each image piece into an array.

Step 4: Reduce the sample

In step 3, we already have an array of image features. However, this array string is still very large, making calculations by computers difficult. Furthermore, in addition to features, arrays also contain noise. To reduce the dimension of the array, we use the max pooling algorithm - the largest filter separation. For each empty 2×2 cell of the array of values, we store only the largest value - the most obvious and interesting feature of the image.

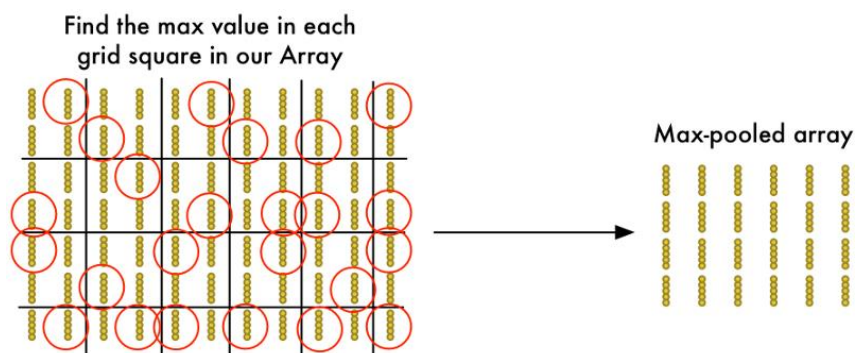


Figure 7. Sample reduction.

Step 5: Transfer the feature to another neural network and make predictions

The above steps transformed the original large image into a collection of the most prominent features. Now, we connect all those features to form a fully connected neural network. The value of the new network will be used as input to the layered neural network to find the final result. Networking is like finding the parts of an object, and we need to put it back together into a complete object. Since this method does not depend on the position of the person in the image, but on the relative position of the features to each other, the image will be recognizable no matter where the image is located. The whole process is as follows:

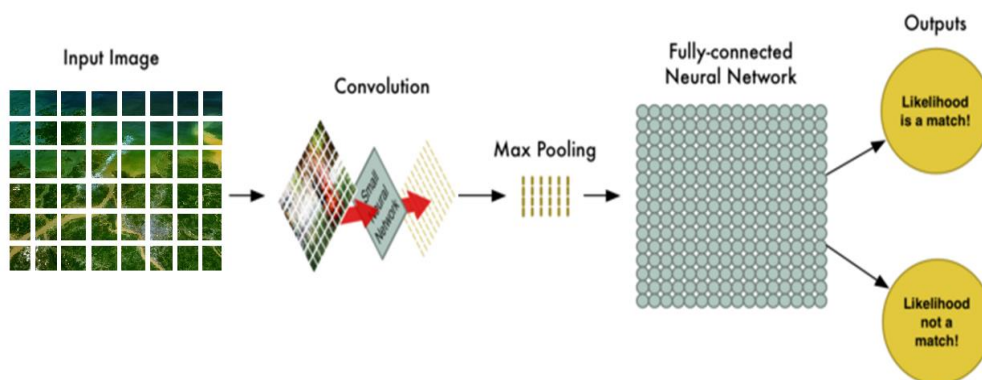


Figure 8. Feature transfer to other neural networks and prediction.

The image recognition process is a series of steps: convolution, sample reduction, and complete networking. When solving the actual problem, these steps can be connected many times. There can be two, three or even ten convolutional layers. Or you can insert sample reduction anywhere in the network. The idea is to start from a large image, step by step we extract the features, reducing the dimensions to get the final result. The more convolution steps there are, the more complex features can be detected by the model.

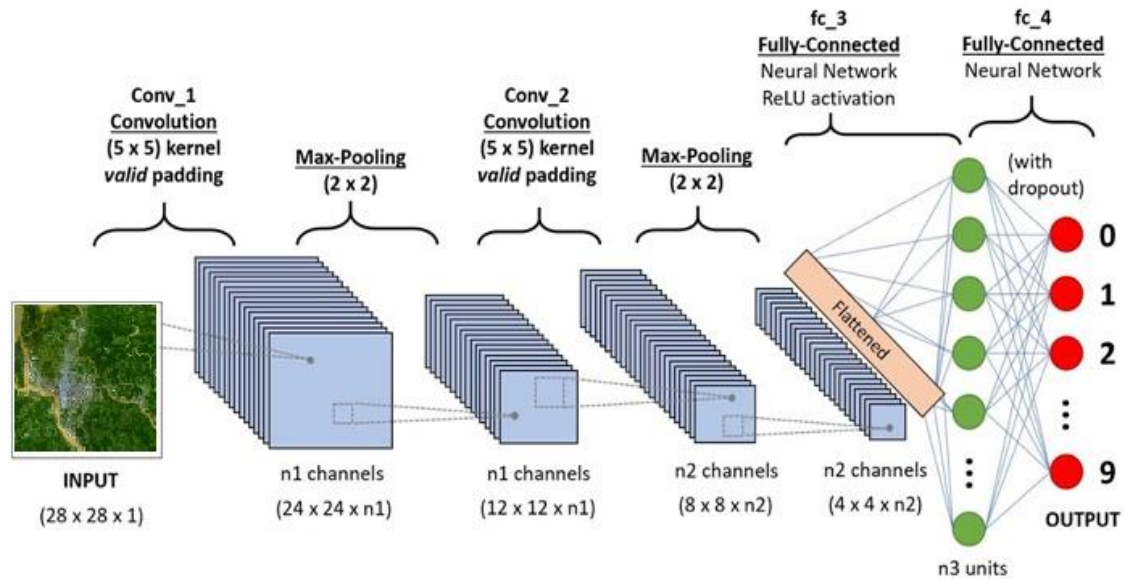


Figure 9. General model of remote sensing image classification using CNN.

Figure 9 is a landcover classification model from remote sensing image with two convolutions, two times using max-pooling technique to minimize data. The same fully connected layer is then used to give the output classification results (Example 10 landcovers).

3. Experimental study

3.1. Experimental data

Experimental data were obtained from the sample data library of Durham University, UK [14]. These images, obtained from manned and unmanned flights, have a resolution of less than 10 cm. Input data are images in .jpg format; The size of these images is 1000×1000 . When applying the CNN model, the image is divided into small image pieces of size 50×50 pixels. Figure 11 is the original and labeled image, there are 5 labeled landcover including water, dry sediment, green vegetation, senescent vegetation and paved roads. In which, pixels that do not know which class they belong to are assigned to unclassified, these are pixels that are not used during training and testing. In which, the data is labeled as 5 million pixels, using 1 million pixels for training and 4 million pixels for testing.

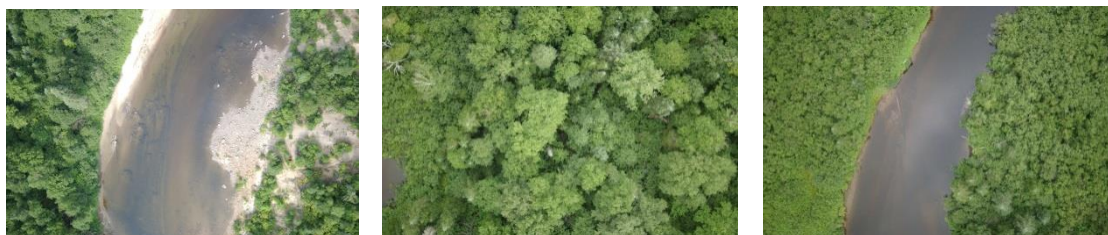


Figure 10. Some images in the test dataset.

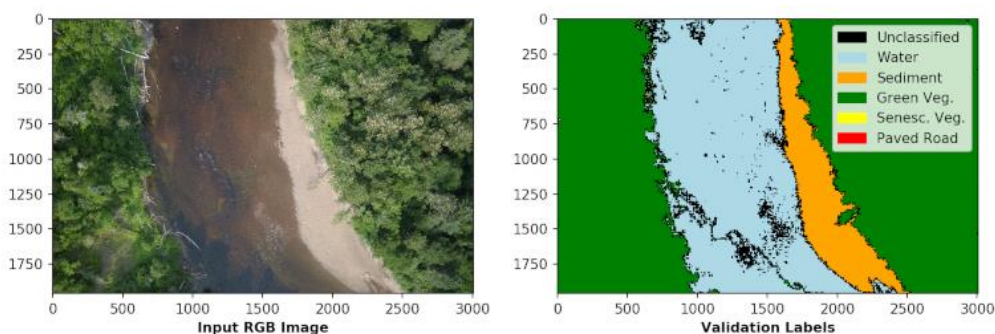


Figure 11. Original image and labeled data.

The experiment was run on an Intel Core i2-2.66GHz computer, with Windows 10 64bit operating system, 16Gb RAM, NVIDIA 2Gb graphics card; The language used for programming is Python 3.11; The library used for experimental implementation is tensorflow. The trained model can be used to classify the landcovers from remote sensing image data.

3.2. Experimental results

In this section, the article uses the Random Forest (RF) algorithm to compare the accuracy of the classification results with the CNN network model. In which, at the fully connected layer of the CNN network, we use a multilayer network (MPL) to give the final classification result. The images are partially classified into 5 classes: water, sediment, green vegetation, senescent vegetation, and paved roads.

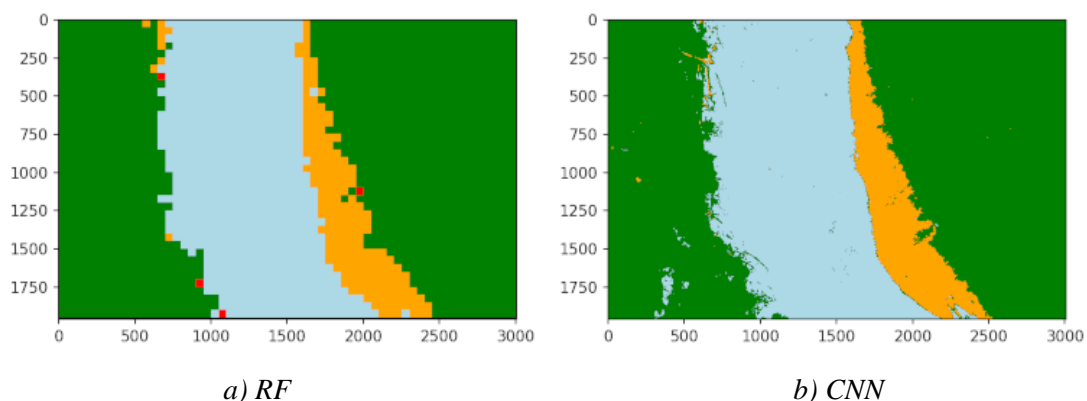


Figure 12. Landcover classification results.

Table 1. Classification results on the training and testing dataset

No.	Class/ Methods	Training dataset (%)		Testing dataset (%)	
		RF	CNN	RF	CNN
1	water	92.46	99.23	91.93	98.76
2	sediment	89.76	93.61	90.78	95.49
3	green vegetation	89.64	97.45	86.32	96.61
4	senescent vegetation	87.67	96.72	89.21	92.87
5	paved roads	88.88	95.45	87.79	93.13
	Total	89.68	96.49	89.21	95.37

The results of the accuracy assessment on the training dataset and the testing dataset show that using CNN network to classify the landcovers can give more than 95% accuracy on the whole dataset. Meanwhile, the RF algorithm only gives an accuracy of less than 90%. The accuracy on the training data set is higher than that on the test data set on both RF and CNN methods.

With this result, it can be seen that the method using the CNN network gives the classification results with much higher accuracy than the classification method using other algorithms. This also confirms the great potential of applying deep learning networks in remote sensing image classification in particular and image analysis in general.

4. Conclusion and recommendations

Due to the strong development of remote sensing technology, traditional image analysis methods are not suitable for big data. Alternative classification approaches based on traditional machine learning models and deep machine learning are gradually appearing in the earth sciences as a good solution to the classification problem. The article has researched and understood the CNN network and its application to the problem of landcover classification on remote sensing images. The classification results achieved an accuracy of over 95% compared to the RF model only gives results with less than 90% accuracy.

Based on the experimental results, we believe that the CNN network architecture is suitable and gives good performance in classifying remote sensing images compared to traditional methods. In the future, we will continue to study the application of CNN network for remote sensing images with different resolutions.

References

- [1] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, B.A. Johnson. "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 152, pp. 166-177, 2019. DOI: 10.1016/j.isprsjprs.2019.04.015
- [2] L. Zhang, L. Zhang and B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art," in *IEEE Geoscience and Remote Sensing Magazine*, Vol. 4, No. 2, pp. 22-40, 2016. DOI: 10.1109/MGRS.2016.2540798
- [3] M. Emmanuel, Y. Tarabalka, G. Charpiat, A. Pierre. "Convolutional Neural Networks for Large-Scale Remote Sensing Image Classification," *IEEE Transactions on Geoscience and Remote Sensing*, 55, pp. 645-657, 2017. DOI: 10.1109/TGRS.2016.2612821
- [4] MIT Technology Review. 10 breakthrough technologies 2013. Available: <https://www.technologyreview.com/lists/technologies/2013/>
- [5] I. Krizhevsky, Sutskever, and G. Hinton. "Imagenet classification with deep convolutional neural networks," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, pp. 1097-1115, 2012. DOI: 10.1145/3065386
- [6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv*, 2014. Available: <https://arxiv.org/pdf/1409.1556.pdf>
- [7] K. He, X. Zhang, S. Ren, and J. Sun. "Deep residual learning for image recognition," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016. DOI: 10.1109/CVPR.2016.90
- [8] M.S. Minu, R. Aroul Canessane. "Deep learning-based aerial image classification model using inception with residual network and multilayer perceptron," *Microprocessors and Microsystems*, Vol. 95, 104652, 2022, ISSN 0141-9331. DOI: 10.1016/j.micpro.2022.104652
- [9] J. Chen and W. Ding (2018), "UAV Image Classification and Recognition Based on SVM and Non-Reference Adaptive Quality Evaluation," *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp. 407-410. DOI: 10.1109/IMCEC.2018.8469647
- [10] Ö. Akar, "Mapping land use with using Rotation Forest algorithm from UAV images," *European Journal of Remote Sensing*, Vol. 50, No. 1, pp. 269-279, 2017. DOI: 10.1080/22797254.2017.1319252
- [11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 38, No. 1, pp. 142-158, 2016. DOI: 10.1109/TPAMI.2015.2437384
- [12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016. DOI: 10.1109/CVPR.2016.91
- [13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 3431-3440, 2015. DOI: 10.1109/CVPR.2015.7298965
- [14] <https://collections.durham.ac.uk/downloads/r2f1881k904>

NGHIÊN CỨU GIẢI PHÁP PHÂN LOẠI LỚP PHỦ TỪ ẢNH VIỄN THÁM QUANG HỌC SỬ DỤNG MẠNG NƠ-RON TÍCH CHẬP

Mai Đình Sinh^a, Trịnh Lê Hùng^a, Đào Khánh Hoài^a

^aTrường Đại học Kỹ thuật Lê Quý Đôn

Tóm tắt: Với sự bùng nổ về dữ liệu và thông tin, các kỹ thuật phân tích dữ liệu ngày càng trở nên quan trọng. Dữ liệu ảnh viễn thám có nhiều ưu điểm trong việc phân tích đặc điểm bề mặt trái đất do phạm vi phủ trùm rộng và thường xuyên được cập nhật. Tuy nhiên, với sự phát triển mạnh mẽ của công nghệ viễn thám dẫn đến nguồn dữ liệu ảnh viễn thám gia tăng nhanh chóng. Điều này đòi hỏi phải có những công cụ mạnh mẽ để giải quyết các bài toán trên nguồn dữ liệu này. Hiện nay, trí tuệ nhân tạo đang có bước phát triển nhanh chóng và là một trong những công cụ phân tích dữ liệu cực kỳ mạnh mẽ trong nhiều bài toán, trong đó có dữ liệu ảnh viễn thám. Trong bài báo này, chúng tôi trình bày giải pháp và đưa ra mô hình ứng dụng mạng nơ-ron tích chập (CNN) cho bài toán phân loại ảnh viễn thám. Thực nghiệm phân loại lớp phủ cho thấy việc ứng dụng mô hình mạng nơ-ron tích chập có thể cho kết quả phân loại với độ chính xác trên 95%. Trong khi đó, mô hình RF chỉ cho kết quả với độ chính xác thấp hơn 90%. Kết quả này cho thấy tiềm năng ứng dụng các mô hình học sâu trong phân tích ảnh viễn thám rất lớn.

Từ khóa: Mạng nơ-ron tích chập; học sâu; ảnh viễn thám; phân loại đối tượng.

Received: 16/03/2023; Revised: 31/05/2023; Accepted: 23/06/2023; Published: 30/06/2023

