A Deep Convolutional Neural Network to Detect the Existence of Geospatial Elements in High-Resolution Aerial Imagery †

Calimanut-Ionut Cira *, Ramón Alcarria, Miguel-Ángel Manso-Callejo and Francisco Serradilla

Universidad Politécnica de Madrid, 28031 Madrid, Spain

* Correspondence: cira.calimanut-ionut@alumnos.upm.es

† Presented at the II Congress in Geomatics Engineering, Madrid, Spain, 26–27 June 2019.

Abstract: This paper tackles the problem of object recognition in high-resolution aerial imagery and addresses the application of Deep Learning techniques to solve a challenge related to detecting the existence of geospatial elements (road network) in the available cartographic support. This challenge is addressed by building a convolutional neural network (CNN) trained to detect roads in high resolution aerial orthophotos divided in tiles (256 × 256 pixels) using manually labelled data.

Keywords: convolutional neural network; remote sensing; road detection

1. Introduction

Machine-learning techniques have become increasingly important in the paradigm shift toward data-intensive sciences [1]. The use of appropriate statistical analysis and artificial intelligence algorithms allows us to better analyse geospatial data. In addition, aerial imagery proved to be equally important for many applications including infrastructure monitoring. These applications traditionally required manual identification of objects, but recent deep learning methods enable the automation of such tasks and achieve great success in imagery classification by means of convolutional neural networks [2].

Nowadays, scientists in geomatics are exploiting the power of deep learning to tackle existing challenges and instigated a new wave of promising research [3,4]. Studies have indicated that feature representations learned by CNNs are highly effective in large-scale image recognition [5]. Deep neural networks are experiencing a rapid development fuelled by big tech companies' investments in research and by the support shown by tech communities (open source projects like TensorFlow/Keras) [6].

Remote sensing image analysis has been a hot research topic in recent years. Researchers have successfully analysed land use patterns in urban neighbourhoods using large-scale satellite imagery data and state-of-the-art computer vision techniques based on deep convolutional neural networks [7]. In [8], the authors train a simple one-dimensional CNN that contains five layers (an input layer, a convolutional layer, a max-pooling layer, a fully connected layer, and an output layer) to directly classify hyperspectral images. The authors of [9] focused on road extraction from aerial images and purposed a semantic segmentation neural network tested on a public road dataset which demonstrated great effectiveness. Recently, the detection of roads from aerial imagery was approached from a spatio-temporal perspective, by training a deep fully convolutional neural network for image segmentation and adding a temporal processing block [10].

The goal of this project is to build a CNN capable of learning the relationships between training inputs and training categories using supervised learning. We evaluate the network's efficiency levels through the accuracy metric given by its performance on the test data.

2. Material and Methodology

The specific machine-learning problem in this case is binary classification (an instance of supervised learning). The research is experimental and has a quantitative approach, the variables being pixels of orthoimages divided in tiles (256 × 256 pixels).

2.1. Data

Labelled data was needed for training and testing the CNN. This dataset was generated by visually comparing the aerial images (PNOA—*Plan Nacional de Ortofotografía Aérea*) to existing vector data (MTN25—*Mapa Topográfico Nacional* 1:25.000) via a web viewer built for this task. This way we were able to assign the correspondent label to each tile (category 1—no road, category 2—road) and allowed the network to learn about the existence of roads in any given tile. For consistency reasons, the same zoom-level has been used during the labelling operations. We took into consideration representative areas in Spain (samples of vegetation coverage—Figure 1) and built a dataset containing around 9000 labelled tiles (png format) with a size on disk of approximately 1.22 GB. On this dataset we follow the convention in the literature of using half of the images for training and half for testing.

(**a**)　　　　　　　　　　　　　　　　　　　　　　　　　　　(**b**)
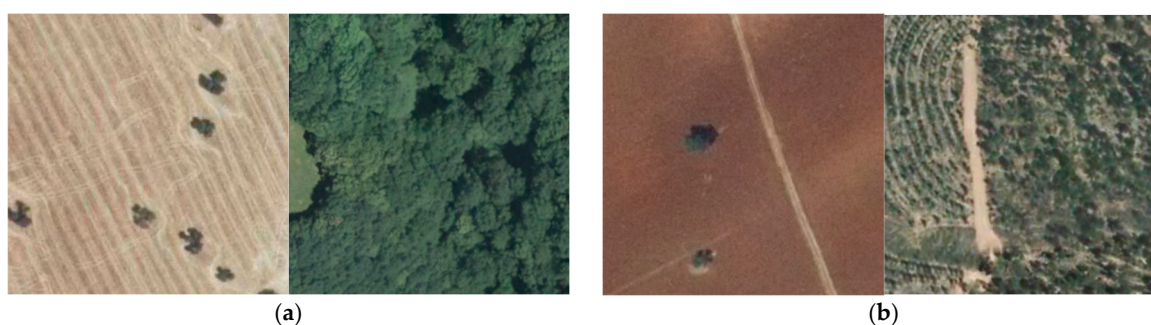
Figure 1. Examples of tiles extracted from the labelled dataset (**a**) Category 1—"No road" and (**b**) Category 2—"Road exists".

2.2. Network's Architecture

In the scientific literature we can find a variety of CNNs built for different tasks (AlexNet, GoogLeNet, VGG16, VGG19, ResNet) often with depths of more than 50 layers and trained with large datasets. However, we built a smaller ConvNet better suited for our task.

In this case, the tiles are passed through four convolutional layers, where we use filters with a 3 × 3 receptive field. Spatial pooling is carried out by four max-pooling layers, which follow every convolutional layer. Max-pooling is performed over a 2 × 2-pixel window. Next, a flatten operation on the tensor reshapes the tensor to have a 1-d share that is equal to the number of elements contained in the tensor.

Finally, the stack of convolutional layers is followed by two Fully-Connected (FC) layers: the first has 512 channels while the second performs the 2-way classification and contains 1 channel. Five of the eleven hidden layers are equipped with ReLU [11] non-linearity activation functions, while the last FC uses sigmoid activation function.

2.3. Network's Training

Convolution operates over 3D tensors and works by sliding these filters of size 3 × 3 over the 3D tensor, stopping at every possible location. It consists of learning to map input data to known targets (annotations) given a set of labelled examples.

The only pre-processing done was turning the image files to pre-processed tensors. Given that our model has few samples we had to control the overfitting. To do so, we augmented the sample via a number of transformations (rotation, zoom range, horizontal flip, height and width shift)—Figure

2. This allows the model to generalize better given it never sees same image twice. Furthermore, we added a dropout layer (with a dropout rate of 0.5) before the fully connected layers.
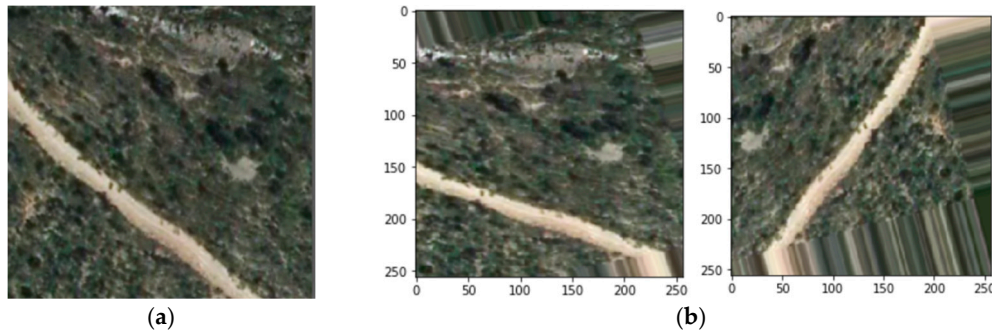

Figure 2. Data augmentation process: (**a**) Original tile; (**b**) Transformed tiles (rotated, flipped, zoomed).

For compiling the model we used binary crossentropy as loss, given that we've chosen to end the network with a sigmoid unit. RMSprop was chosen as optimizer, with a learning rate set at 0.0001. The model was trained for 200 epochs using a Linux cloud computing system with TensorFlow installed and an integrated GPU (Nvidia Tesla K80).

3. Results and Discussion

The network achieved a classification accuracy of 90% on the training and test set (Figure 3a), the loss decaying to less than 0.3 (Figure 3b). We consider these error rates to be small considering the size of the dataset and the number of convolutional layers. Our results show that a large, deep convolutional neural network is not always necessary for achieving high accuracy results on a challenging dataset using only supervised learning.
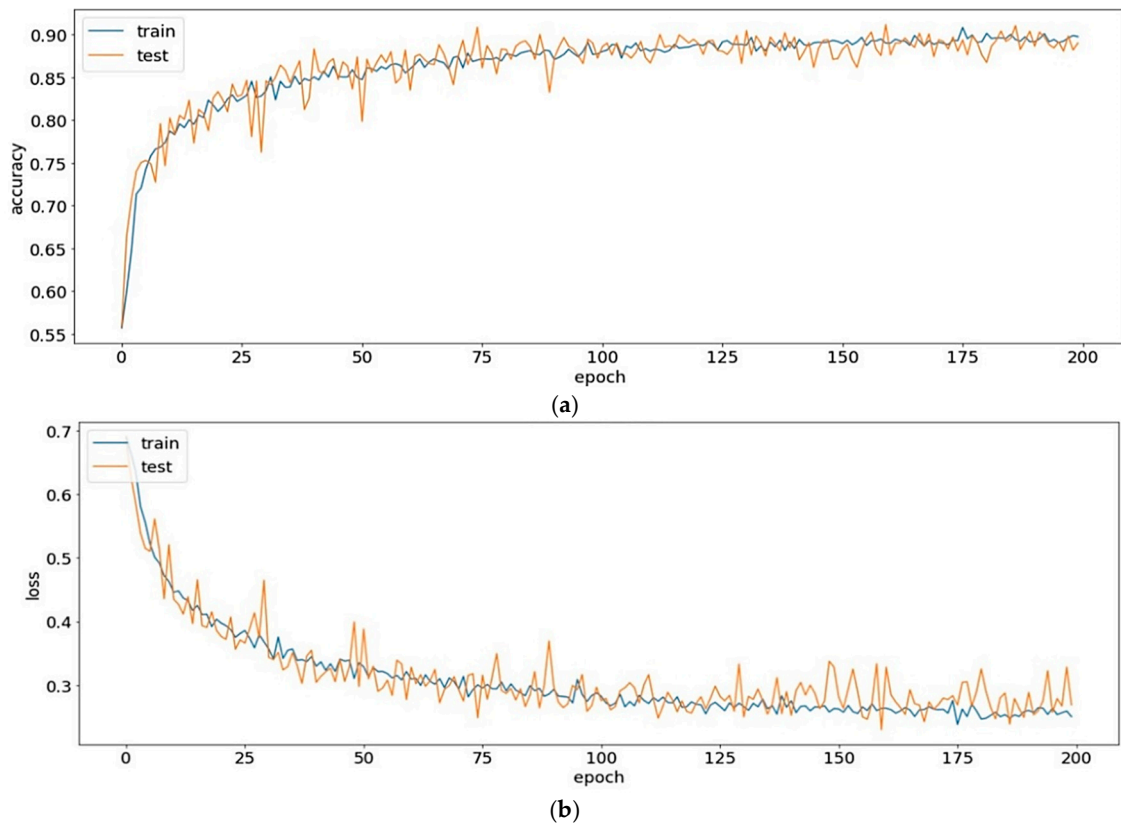
Figure 3. Evaluation of the model's performance: (**a**) Model's Accuracy; (**b**) Model's loss.

Thus far, our results have not improved as we have made our network larger (added 1 and 2 convolutional layers + 1 fully connected layer) or trained it for longer (500 epochs). The accuracy of the model should increase with the number of training samples and would be interesting to see how the model reacts when the dataset increases.

Deep learning proved to be an extremely powerful tool for remote-sensing data analysis, convolutional networks being known for their performance in addressing visual-classification problems. Next, we plan to review two more techniques for applying deep learning to small datasets: feature extraction with a pretrained network and fine-tuning the pretrained model and to use image segmentation techniques to extract such the infrastructures in the form of vector data layers. We also plan to increase the dataset and improve these values. We hope to obtain high efficiency levels that could reduce reliance on human participation in detecting changes in geospatial elements, help state administration in reducing mapping costs and the citizens by having up-to-date cartography in less time.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36.
2. Pritt, M.; Chern, G. Satellite Image Classification with Deep Learning. In Proceedings of the IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 10–12 October 2017; pp. 1–7, doi:10.1109/AIPR.2017.8457969.
3. Camps-Valls, G.; Tuia, D.; Bruzzone, L.; Benediktsson, A. Advances in hyperspectral image classification. *IEEE Signal Process* **2014**, *31*, 45–54.
4. Tuia, D.; Persello, C.; Bruzzone, L. Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 41–57, doi:10.1109/MGRS.2016.2548504.
5. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *Proc. ICLR* **2015**. Available online: https://arxiv.org/pdf/1409.1556.pdf (accessed on 24 April 2019).
6. Bughin, J.; Hazan, E.; Ramaswamy, S.; Chui, M.; Allas, T.; Dahlström, P.; Henke, P.; Trench, M. *Artificial Intelligence the Next Digital Frontier*?; Discussion Paper; McKinsey & Company: New York, NY, USA, 2017.
7. Albert, A.; Kaur, J.; González, M. Using Convolutional Networks and Satellite Imagery to Identify Patterns in Urban Environments at a Large Scale. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; doi:10.1145/3097983.3098070.
8. Hu, W.; Huang, Y.; Wei, L.; Zhang, F.; Li, H. Deep convolutional neural networks for hyperspectral image classification. *J. Sens.* **2015**, 2015, 258619, doi:10.1155/2015/258619.
9. Zhangy, Z.; Liuy, Q.; Wang, Y. Road Extraction by Deep Residual U-Net 2017. *IEEE Geosci. Remote Sens. Lett.* **2017**, doi:10.1109/LGRS.2018.2802944.
10. Luque, B.; Morros, J.R.; Ruiz-Hidalgo, J. Spatio-Temporal Road Detection from Aerial Imagery using CNNs. In Proceedings of the International Conference on Computer Vision Theory and Applications, Porto, Portugal, 27 February 2017; doi:10.5220/0006128904930500.
11. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**.