International Journal of Innovative Studies in Sociology and Humanities
ISSN 2456-4931 | Open Access | Volume 8, Issue 1, 2023
DOI: https://doi.org/10.20431/2456-4931.080118

Using a Deep Learning Method for the Detection of Buildings from a High Spatial Resolution Satellite Image

Messaadi Ibtissem[1], Redjem Sandra[1], Raham Djamel[2]

[1]Faculty of Earth Sciences, Geography and Spatial Planning, University of Constantine1, Algeria.
[2]Research Center in Spatial Planning 'CRAT', Constantine, Algeria.

Received: January 05, 2023 **Accepted**: January 21, 2023 **Published**: January 25, 2023

Abstract

In the recent years, the detection of buildings from using high spatial resolution remotely sensed imageshas been a very active research topic. The use of Deep Learning based on convolutional neural networks (CNN), coupled with open-source available satellite or aerial images, which has greatly simplified the processes. However, the architectures of this type of Deep Learningrequires the production of large datasets with totally labeled ground truths. These operations appear to be expensive, because these annotations are limited or difficult to achieve, despite the availability of free-access images. Moreover, the model training can be long and requires high computational capacity. In this context, our contribution is to adjust and improve the performance of a pre-trained model, using an own dataset adapted to our needs and our study area.

Key words: Deep Learning; automatic classification; buildings; very high spatial resolution satellite image.

INTRODUCTION

The first airborne sensors, capable of obtaining Earth observation images, appeared in the 1970s, then, in the 1980s and 1990s, they witnessed the emergence of medium and high-resolution sensors. The latter were followed in the 2000s by space sensors with very high spatial resolution (VHSR). This progress is significant from the moment when the details observed in an image increased considerably at each stage.

Although this technological leap offers great potential of application in the field of remote sensing[1], it has led to real complications, because recent space observation missions allow the acquisition of more accurate and larger images. As these high resolution VHR remote sensing images contain a large amount of information, processing them requires long, laborious, thorough and costly work, with manual or semi-automatic methods, requiring heavy human intervention[2] [3].In order to efficiently process these remote sensing images, it is therefore necessary to have fully automatic methods. In recent years, considerable progress has been made in the field of automatic processing of satellite data, particularly in image classification techniques. Unlike manual approaches, automatic approaches respond to an urgent need to limit manual effort, accelerate the requiredtime for this task ofimage classification, and present good performance that optimizes costs.

Automatic image classification approaches involve associating each pixel of the image with a land use class. Traditionally, two approaches are distinguished: supervised and unsupervised, unfortunately, they all have limitations at different levels [4]. While deep learning based on convolutional neural networks (CNN) coupled with the provision of satellite images has greatly simplified the processes. Therefore, our study focuses on the use of Deep Learningmethods for the detection of buildings from remote sensing images of high spatial resolution. These information are needed in multiple applications, mainly related to the implementation of public policies [5].

METHODOLOGY

From its characteristics, the urban tissue is difficult to detect on satellite images, and conventional detection approaches do not provide satisfactory performance. However, Deep Learning has proven to be effective in this segmentation task [6][7]. Thus, the approach proposed in this article focuses on the extraction of buildings by Deep Learning from satellite images with high spatial resolution. It is divided into two main steps:

First, the creation of a ground truth that is composed of pairs of images (image and its label) and an annotation file. This step consists in the acquisition and preparation of the necessarydata for the training of a model which allows the extraction of buildings. Indeed, Deep Learning architectures require large data sets, with densely labelled ground truths, which accurately represent all the characteristics of the object, including its boundaries [8].

The next step focuses on the choice of the model, its training from already constituted ground truth data, and its use to extract the targeted objects. This model allows us to perform segmentation operations automatically. It is based on networks of convolutionalneurons that will search the images for the characteristic elements of the expected objects.

The Creation of a Ground Truth

The creation of an image dataset for the model training is the preliminary step to any use of a supervised Deep Learning model for extracting objects. In fact, the good performance of the current neural networks is based on the availability of large fully annotated databases [9]. Furthermore, the accuracy of the model depends on the quality, quantity and variety of this learning dataset. In addition, the manually created field truth has been adapted to our needs and our study area. The suggested procedure includes the following steps:

- The first step consists in uploadingopen-sourceshared images, the latter are of different architectures and different environments, in order not to have a specialized model on a specific type of landscape.
- The images were formatted with an identical dimension of 256x256 and a PNGextension. This work was done online on the website : www.convert-a-image.com
- The next step is label generation, labelling. This type of processing involves identifying and delimiting the elements of an image, by associating each pixel with a cartographic category (in our case the soughtcategory represents the buildings). The objects or areas of interest can be easily reconstructed by vectorizing the result. The vectorization was done manually.
- Aset of pairs (image and associated label) was created, in addition to the json annotation file, describing the presented objectsin the images. The obtainedlabel is in black and white, where the buildings are in white and the rest of the landscape is in black.

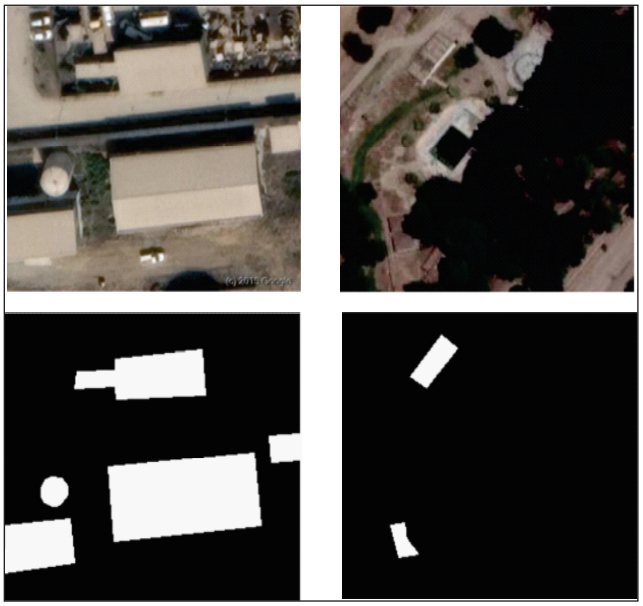

Figure 1. Examples of satellite images (on the top) and its associated labels (on the bottom) highlighting the sought element “Building”.

Preparation of the Model

This part consists mainly in teaching our model to identify the expected objects, based on ouralready builtground truth set. It will therefore define a set of rules to lead to a better prediction, and therefore to the detection of buildings.The various parameters of the model are defined in the Python scripts. For example:

The Location of the Input Data

We are talking about two image directories for the model training, the images sizeis 256x256. Each image in the image train directory has its opposite in the label train directory, while keeping the same file name.

The Image Augmentation Function

Creating a proper dataset is a task that can be costly, and image enhancement functionscontribute in making it more robust. They make it possible tosignificantly increase the number of images in input (without touching the dataset). The main idea behind the Data Augmentation is to reproduce the pre-existing data, by applying a random transformation (zoom, offset, rotation, cutting, stretching or mirror effect)

The Definition of the Soughtclasses: Called Labels

Among the parameters to be defined is the number of targeted categories, the "background" category is the default category. In our case, we have two categories of objects: our sought object category, the buildings, and the background category for the rest of the image.

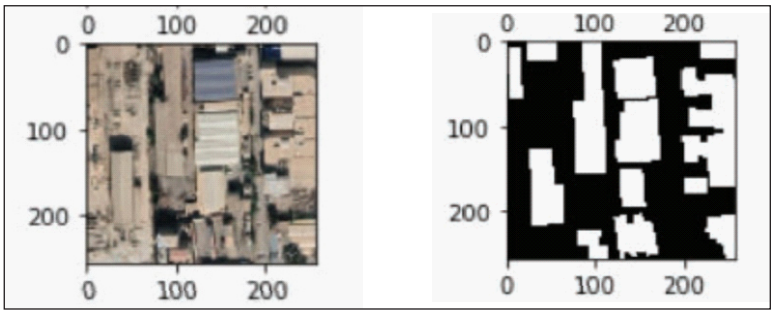

Figure 2. Definition of the sought classes.

Loss Function

This function allows to quantify the gap between the model predictions and the actual observations of the data set used in the learning. The training phase aims to find model parameters that will minimize this function.

Model Training, Validation and Testing

Training allows the model to know the task assigned to it, which is the detection of buildings in our case. The model is trained in a supervised way, it requires a large number of images and associated labels as input of the expected result in the output. Thus, the model adapts and becomes more and more accurate, thanks to an iterative training on this set of ground truth, while optimizing the margin of error between its predictions and the expected result provided by the annotations file associated with our images.

To reduce significantlythe risk of over-adjustment, the dataset is divided into three subsets. A training set on which the model will learn, a part of the data set will serve to validate the model, and a test set that will allow to verify the validation of the model on data, on which it has not performed its learning. Generally, the ground truth is randomly distributed according to these proportions: 80% of the data are used for learning and 20% for testing.

U-Net Architecture

U-Net, issued from the traditional CNN neural network, was designed for fast and accurate image segmentation[10][11]. Its architecture has been modified and extended to work with fewer training images. It is able to locate and distinguish the boundaries of the elements composing a certain image, by classifying each pixel. Its architecture is symmetrical and consists of three sections: Contraction, the choke point and expansion section, as shown in figure n°03. The first block, also called encoder, allows to retrieve the context of an image. The bridge, or the choke point, connects the encoder and the network of decoders and completes the flow of information. The second block is the decoder.

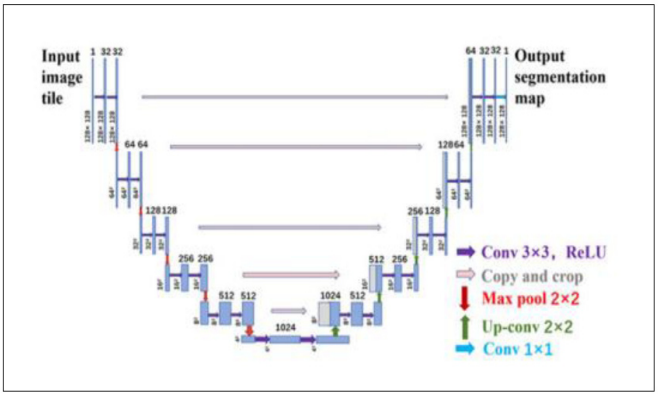

Figure 3. U-net network structure.

RESULTS AND DISCUSSIONS

In order to validate our model, we established a dataset of a total of 220 images, shared as follows: 200 for training and validationof the model, and 20 for testing. The model was tested on new images, which were not used neither for training nor for validation, but were still categorized when the dataset was created.

Once the model is trained, its effectiveness must be assessed. The visualization of the cost curve, obtained at the end of the training, gives information on the relevance of the trained network and the number of epoch from which the model achieves its best performances.The result of the cost function is presented in Figure n°04.

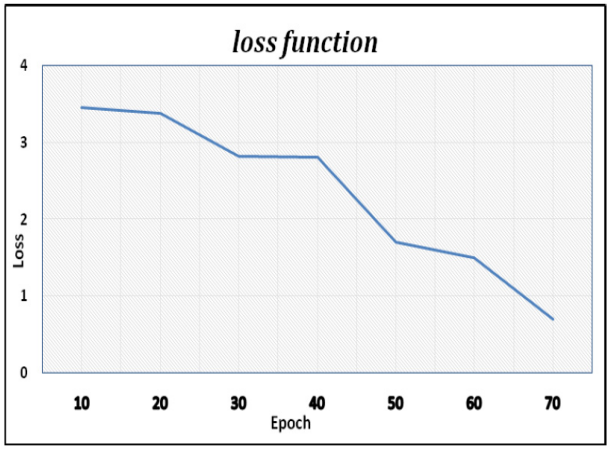

Figure 4. Loss curve.

The achievement of a best model, which responds to both training and validation data, is achieved through a series of tests and iterations, after which the parametersof the neural network are modified. The final result is presented in the form of annotated images, with the prediction of belonging to one or the other of the 2 soughtcategories: buildings/ non-buildings. Figure No. 05 shows an example of the building detection results.

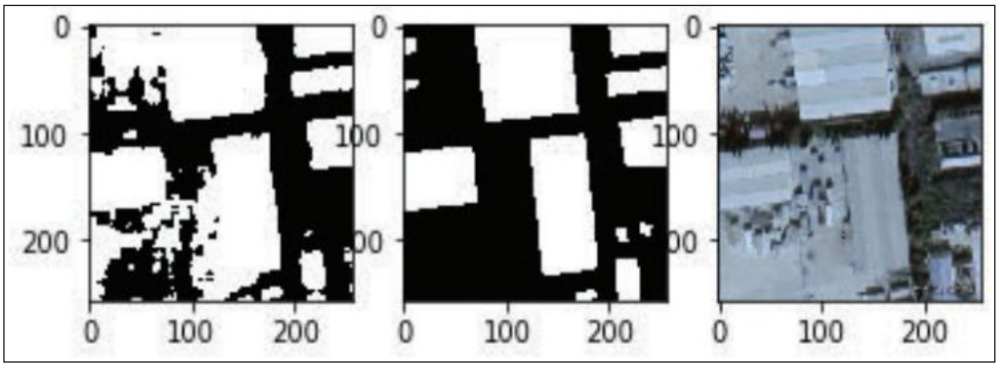

Figure 5. Exampleof building detection, from right to left: the original image, label image, and prediction result.

We observe that the model has learned to better detect the buildings in the image. Thus, the result is satisfactory with an overall accuracy rate of 84%, even if the geometry of the building outline remains unreliable.

The quality and accuracy of the result depends on certain factors. Indeed, the quality of the model that will be trained is strongly influenced by the image-label combination. For this reason, it is necessary to produce a labeled dataset of high quality.Unfortunately, these generated ground truths may have certain limitations, namely: the vagueness of labeling such as differences between the location of the building on the image and its label or omissions of elements, which is likely to generate detection errors by the model. Similarly, resolution, sharpness, image contrast and other effects on the image such as shadow can degrade the quality of the dataset and can impair the detection of the sought objects. In addition, a homogeneous dataset or one that is very specific to an architectural style or to a geographical area, will provide a very specialized detection model that is not ready for use for certain sets of ground truth.

CONCLUSION

In this study, we suggested the use of a supervised learning model, based on convolutionalneural networks, for the detection of buildings from remote sensing images of high spatial resolution.

The conducted experiment demonstrated the value of learning by transfer. Indeed, working with an insufficient amount of data (namely our case) would lead to a decrease in performances and a degradation of the model's ability to make good predictions. Thus, starting with a pre-trained model has helped us build better models. In this study, the proposed method is based on the use of a U-Net architecture model. The latter has proven its performance in the field of classification and segmentation of images. In fact, the targeted objects present on our dataset were detected with a global accuracy of 84%. Hence, we could say that the result is acceptable both in terms ofdetection and delimitation of buildings.However, there are areas for improvement. Indeed, our future works could involve the introduction of new models that have an impact on the producedresult, such as the Mask R-CNN model, which allows the instance segmentation of objects. In addition, the use of a broader and more heterogeneous dataset could further improve the quality of the expected predictions.

REFERENCES

1. Bayoudh, M. (2013). Apprentissage de connaissances structurelles pour la classification automatique d'images satellitaires dans un environnement amazonien (Doctoral dissertation, Université des Antilles et de la Guyane).
2. Belarte, B. (2014). Extraction, analyse et utilisation de relations spatiales entre objets d'intérêt pour une analyse d'images de télédétection guidée par des connaissances du domaine (Doctoral dissertation, Strasbourg).
3. You, Y., Wang, S., Ma, Y., Chen, G., Wang, B., Shen, M., & Liu, W. (2018). Building detection from VHR remote sensing imagery based on the morphological building index. Remote Sensing, 10(8), 1287.
4. Pelletier, C. (2017). Cartographie de l'occupation des sols à partir de séries temporelles d'images satellitaires à hautes résolutions: identification et traitement des données mal étiquetées (Doctoral dissertation, Université de Toulouse, Université Toulouse III-Paul Sabatier).
5. Mallet, C., Chehata, N., Le Bris, A., & Gressin, A. (2014, June). Détection de bâtiments à partir d'une image satellitaire par combinaison d'approches ascendante et descendante. In Reconnaissance de Formes et Intelligence Artificielle (RFIA) 2014.
6. Long, J., Shelhamer, E., &Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
7. Weng, Y., Zhou, T., Li, Y., & Qiu, X. (2019). Nas-unet: Neural architecture search for medical image segmentation. IEEE Access, 7, 44247-44257.
8. Pastorino, M., Moser, G., Serpico, S. B., &Zerubia, J. (2021, September). Segmentation Sémantique d'Images de Télédétection Combinant Modèles Graphiques Probabilistes Hiérarchiques et Réseaux de Neurones Convolutifs Profonds. In ORASIS 2021.
9. Castillo-Navarro, J., Le Saux, B., Boulch, A., & Lefèvre, S. (2019). Réseaux de neurones semi-supervisés pour la segmentation sémantique en télédétection. In Colloque GRETSI sur le Traitement du Signal et des Images.

10. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

11. Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical image segmentation based on u-net: A review. Journal of Imaging Science and Technology, 64, 1-12.

Citation: **Messaadi Ibtissem, Redjem Sandra, *et al.*** *Using a Deep Learning Method for the Detection of Buildings from a High Spatial Resolution Satellite Image. Int J Innov Stud Sociol Humanities. 2023;8(1): 179-184. DOI: https://doi.org/10.20431/2456-4931.080118.*