DEEP LEARNING APPROACH FOR URBAN MAPPING

C. Najjaj [1], H. Rhinane [1], A. Hilali [1]

[1] Geosciences Laboratory, Faculty of Sciences Ain Chock, University Hassan II, Casablanca 20100, Morocco

Commission 4, WG 7

KEY WORDS: Deep learning approach, Fully Convolutional Network; UNET; Satellite imagery; 25 epochs: Accuracy; ARCGIS;

ABSTRACT

Researchers in computer vision and machine learning are becoming increasingly interested in image semantic segmentation. Many methods based on convolutional neural networks (CNNs) have been proposed and have made considerable progress in the building extraction mission. This other methods can result in suboptimal segmentation outcomes. Recently, to extract buildings with a great precision, we propose a model which can recognize all the buildings and present them in mask with white and the other classes in black. This developed network, which is based on U-Net, will boost the model's sensitivity. This paper provides a deep learning approach for building detection on satellite imagery applied in Casablanca city, Firstly, to begin we describe the terminology of this field. Next, the main datasets exposed in this project which's 1000 satellite imagery. Then, we train the model UNET for 25 epochs on the training and validation datasets and testing the pretrained weight model with some unseen satellite images. Finally, the experimental results show that the proposed model offers good performance obtained as a binary mask that extract all the buildings in the region of Casablanca with a higher accuracy and entirety to achieve an average F1 score on test data of 0.91.

1. INTRODUCTION

Satellite image segmentation finds its application in urban planning, building control, forest management, meteorology and land-use agriculture, subsequently, they are the ones responsible for the identification of illegally built dwellings, and the topic of object detection on high-resolution satellite images is currently at the center of researchers. In both urban planning and population modeling, building mapping is very important. Although many techniques have been proposed, currently, some authors like Garcia from Spain in 2017 [1] proposed the fully convolutional network as a review on deep learning techniques applied to semantic segmentation which's more and more being of interest for computer vision and machine learning. Others in 2019 [3] used a novel fully convolutional network (FCN) that adopts attention based re-weighting to extract buildings from aerial imagery, furthermore Yaning Yi [2] from Chinese Academy of Sciences exposed DeepResUnet as a Deep Convolutional Neural Network of Urban Buildings from VHR Remote Sensing Imagery. Recently, Heng Liu in 2020 [6] from china prefer to use a multiloss method to solve the problem that the boundary of the segmentation region is not clear.

So, the remainder of this paper provides an Summary of various approaches to extracting buildings for detection, we used supervised learning in this project, and in particular the Unet algorithm that represents a convolutionary neural network architecture. Then the suggested method of profound learning is explained. To illustrate the feasibility and efficiency of our method, we provide insight into the experiments performed and, at the end, we present the findings. A brief discussion of the results obtained is given. Our building extraction task's purpose is not only to detect building location, but also to ensure that the precisely extracted individual buildings are complete (spatial extent). We also present a simple but successful method for initialization that uses a unet model that has been trained with extra spectral bands while still using the learned parameter values of a pre-trained model.

The obtained results show that deep CNN may be used to extract buildings reliably at scale. The knowledge gained from this operational effort will be useful in future large-scale object detection projects based on remote sensing data.

2. MATERIALS AND METHODS

2.1. Study site

Casablanca, Arabic Al-Dār al-Bayḍāʾ, or Dar al-Beïda, principal port of Morocco, on the North African Atlantic seaboard. [9] Located in the central-western part of Morocco bordering the Atlantic Ocean, it is the largest city in the Maghreb region and the eighth largest in the Arab world (Fig.1) [11]. Casablanca, located near the capital city of Rabat, is the main Atlantic seaport in Morocco [1]. Its man-made harbor has a 3,180 meter-long jetty with fishing and canning as the primary industries. With a population of approximately 3,000,000 people, Casablanca is the largest city in Morocco. Not surprisingly, it has the largest Jewish community in the country, numbering about 5,000. [8]
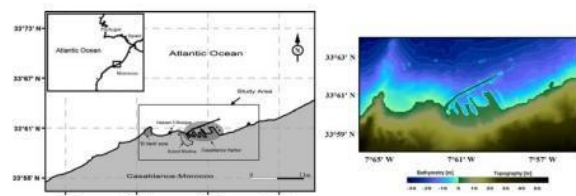

Figure 1. The study area of Casablanca-Morocco, The Bathymetry and topography for the Casablanca and its surrounding are

The city's origins are closely connected to the medieval town of Anfa, which is now a suburb of the modern metropolis [8]. The Berbers, who made Anfa their capital in the 7th century, quickly embraced Islam, but held fast to certain heretical doctrines. For example, they embraced their own prophet and developed a Quran in the Berber language. [10]. Visitors to Casablanca will enjoy the seacoast area with its fine dining spots, the famous Hassan II

Mosque, the bustling central market area, the Habous district, the Royal Palace, the Mohamed V square, and the residential area of Anfa. Located near the capital city of Rabat, Casablanca is also the site of Hassan II University (Fig.2) [9] in this Project, the extraction of buildings was focused in the city of Casablanca. Its population was 3,359,818 inhabitants during the 2014 census making it the most populous city in the country, and its agglomeration was 4,270,750 inhabitants.
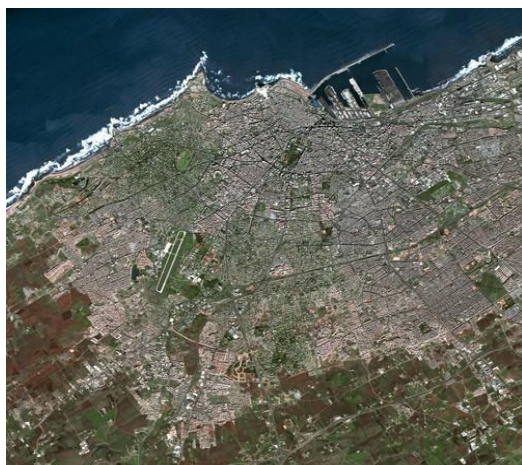

Fig. 2. Satellite Image of Casablanca city.

2.2. Images acquisition:

The important part for training and quality estimation for various machine learning algorithms is a dataset of images [3]. There are now several repositories of satellite images online. [3]

To train any neural network, a training dataset is required that directly depends on the type of neural network to be trained. In our case it is a convolutional neural network, [4] which receives an RGB synthesis of 512*512 image pixels at the input, and at the output gives a binary image. [12]

In our research the data used in this project for building detection will be 1000 satellite imagery which is taken from SaS Planet and the masks extracted using ArcGIS software to train the model Unet for 25 epochs (Tab.1). Every image of the Planet dataset has a resolution of 16384 × 16384 pixels with a spatial resolution of 0.5 m / pixel.

Model	Epoch	Batch_size	Validation_split
UNET	25	10	0,15

Tab. 1. Epoch, batch size &Validation_split of our train model. One advantage of our dataset is that the images have been labelled by humans, resulting in a fairly good precision [13]. However, a non-negligible proportion of satellite images were not update with the most recent masks, as new buildings had already been added by users. [13] (Fig.3)

Some images were also blurry.[13] Overall, this undoubtedly reduced our model's performance as it had trained partially on mislabeled and blurred training images and resulted in a suboptimal performance at test time.[13] (Fig.4)

Fig. 3. One binary building mask generation

(a)   (a')   (b)   (b')

Fig.4. The detailed comparison between two binary building mask generation.

Basically, the U-net builds upon the Fully Convolutional Network [2]. A contracting path extracts features of different levels through a sequence of convolutions, ReLU activations and max pooling. [5] The shapes of the produced building detection have smoother boundaries and have smaller size compared to the ground truth. [16] In contrast, the building extraction produced by this UNET archiecture are very similar to the ground truth. [16] (Fig.5)

Fig.5. Satellite Image and its Mask.

2.3. Images processing

The U-Net work well with only a limited number of training examples, provided one made heavy use of data augmentation. Accordingly, [13] in order to maintain a reasonable amount of images, and above all to avoid overfitting by ensuring a sufficient invariance and robustness of the network, [13] This task was done using the Keras framework which allowed us to augment the data when feeding the network with batches. [8] In order to train our model on a pre-built network, as we will see further on, we first had to resize our images and label masks to a $256 \times 256 \times 3$ size Figure14. Below is one of the satellite images from our dataset, along with its corresponding mask (Fig.6):

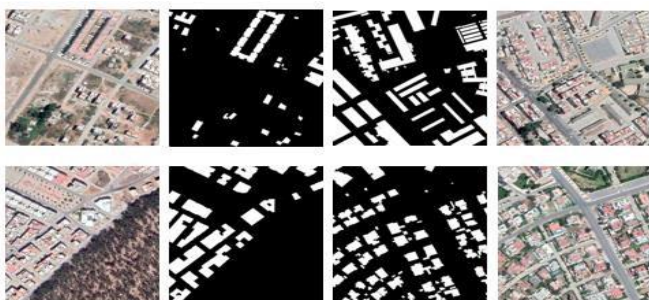**Fig. 6**: Satellite image and corresponding mask with buildings identified in white.

Fig. 7: Ground-truth masks and satellite images from our Dataset.

The (Fig.8) shows the Data processing framework in the study to detect buildings in our study area which is Casablanca's city.
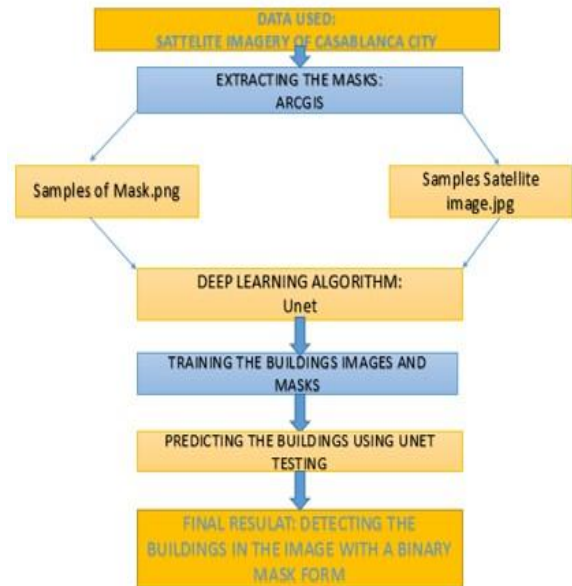

Fig. 8: The high-level architecture for the proposed building detection methodology, which contains 4 main stages: preprocessing, image preparation, model training and building prediction.

- Training set.
- Validation set.
- Test set.

Model	Training Set	Validation Set
UNET	850 Samples	150 Samples

Tab.2: The training and validation set.

2.4. Model Building

The main idea of U-Net (Fig.9) is replacing pooling operators between successive layers by up sampling operators. Those layers thus increase the resolution of the output. [9] This technique helps to locate desired objects with high accuracy because the total set of features or image depth is increased when doing the upsampling, which will help to create a proper mapping between the input and the output. Additionally, this network uses only the correct portion of each convolution that has no completely linked layers around the network. [9] Most deep neural networks including U-Net are facing the following common problem: distributions of the output of each hidden layer always depend on the result of previous layers. [6] Thus, model parameters are changed during the learning process. Each layer needs to adapt to those changes during the training process. [4] Instead of developing a model from scratch, we decided

to use an existing model of Convolutional Neural Network for image segmentation. Namely, we turned to the U-net, [13] originally developed for biomedical image segmentation. [7] Once trained, the network was able to output a pixel wise binary classification (building or not) with good accuracy. [13] (Fig.9)
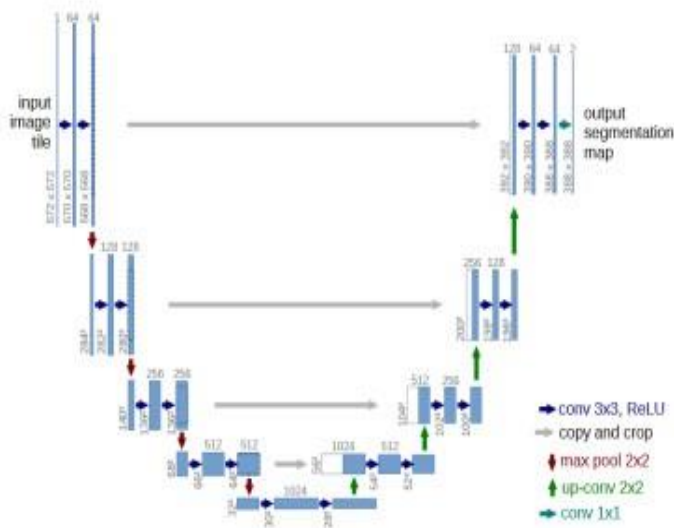

Fig.9: Original U-net for biomedical image segmentation.

2.5. Model Evaluation

There is an image set that is devoted to validating the pattern. Initial model preparation, is used in each model with 25 epochs along with 10 batch size. The end of each epoch is determined using validation dataset to measure model accuracy. To measure the similarity between labeled footprints for buildings and to predict footprints for each model, (Tab.3).

Model	Epoch	Batch_size	Accuracy (A)	Validation Loss
UNET	25	10	0.9176	0.1931

Tab.3: Overall Summary on Test Data.

Once clearly separating out whether it is false positive or true positive, model accuracy can be calculated using F score.
F score can be defined as:

$$F_1 = \left(\frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right)^{-1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

3. RESULTS ANALYSIS AND DISCUSSION

After using the deep learning algorithm Unet, all the Buildings were detected with a high accuracy and best performance and shows in a binary mask form (fig 10- 11-12). With a python code.

Fig.10: Test image 1 of Casablanca city (unseen areas)

Fig.11: Test image 2 of bouskoura (unseen areas)

Fig.12: Test image 3 of nassim Casablanca (unseen areas)

he performance of a model should not only be evaluated based on metrics, but also based on the number of free parameters. The larger his number, the longer the training of the network takes. [9] It is mportant to be aware of the respective training dataset. In general, larger training datasets can cover a greater amount of buildings and variety of building appearances, [15] allowing the network to produce scores with higher certainty and generalize better.
The example collects the history, returned from training the model and creates two charts:

- A plot of accuracy on the training and validation datasets over training epochs. (Fig.13)

- A plot of loss on the training and validation datasets over training epochs. (Fig.14)
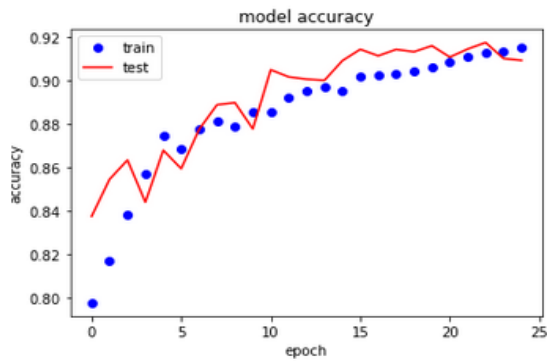
Fig.13: A plot of accuracy on the training and validation datasets over training epochs.
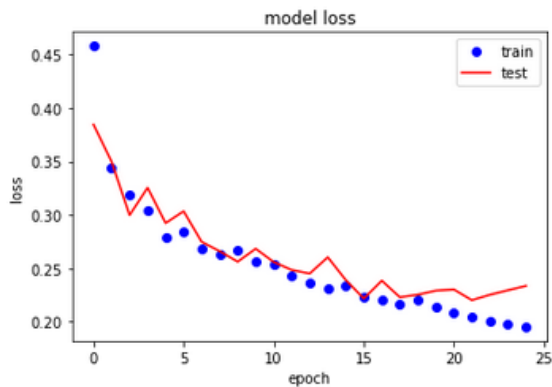

Fig.14: A plot of loss on the training and validation datasets over training epochs.

The History callback is one of the default callbacks that is registered when all deep learning models are trained. Each epoch's training metrics are recorded. This comprises the loss and accuracy (for classification tasks) as well as the loss and accuracy (if one is specified) for the validation dataset. [17] Calls to the fit () function, which is used to train the model, return the history object. [18] In the history member of the object returned, metrics are saved in a dictionary.

Accuracy is more straightforward. It measures how well our model predicts by comparing the model predictions with the true values in terms of percentage. The learning curve represents the neural network model accuracy and loss plots with respect to training and validation data set after each epoch of training are run. [18] The learning curve helps in determining the optimal value of Hyperparameters for creating the most optimal neural network in order to avoid overfitting and help achieve greater generalization in making correct predictions on unseen data.
From the plot of accuracy we can see that the model could probably be trained a little more as the trend for accuracy on both datasets is still rising for the last few epochs.

From the plot of loss, we can see that the model has comparable performance on both train and validation datasets (labeled test). If these parallel plots start to depart consistently, it might be a sign to stop training at an earlier epochs.

4. CONCLUSION

In this project, we examined a semi-semantic satellite image segmentation approach for building detection using 1000 satellites images and relatively low resolution. Our proposed solution has achieved good precision, though slightly less than the latest findings published in the literature using other CNN architectures, and without post processing. U-Net architecture performance is good and it can be used for various segmentation tasks. We have achieved average F1 score on test data is 0.91. U-Net is relatively easy to understand and implement. In our case we have seen that building boundaries are well separated by U-Net model. We adapt the Unet architecture for the task of building extraction and show that it can provide building masks of high quality. Identifying buildings from images appears to be a binary classification problem, with buildings being one class and nonbuildings being the other. Although the preceding CNNs are capable of pixel labeling segmentation, we approach our building extraction problem from a different perspective from classaware to instance-aware pixel labeling segmentation.

The performance of the proposed architecture does not depend on simple or reoccurring shapes, but segments complex and very small building structures accurately. Data augmentation techniques can increase training, Learning curve is used to plot the model training, validation accuracy and training, validation loss vs epochs.

Keras is a powerful library in Python that provides a clean interface for creating deep learning models and wraps the more technical TensorFlow, You can learn a lot about neural networks and deep learning models by observing their performance over time during training, Visualizing data is one of the best ways to humanize data to make it easy to understand and get the relevant trends from it. This activity can be crucial when we still trying to optimize the model and make it production ready. Matplotlib library offers many different tools to help in this visualization process.

REFERENCES

1 A. Garcia-Garcia, S. Orts-Escolano, S.O. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez: A review on Deep Learning Techniques Applied to Semantic Segmentation, the Department of Computer Technology, University of Alicante, Spain, and pp: 1-23, 2017.

2 Ke Li, G.Wam, G.cheng, L.Meng, J.Han, Yaning Yi: Object Detection in Optical Remote Sensing Images, Technical University of Munich. Germany & School of Automation, china pp: 1-19, 2019.

3 A. Ostrovskaya, L.Ivanovsky, V. Pavlov and V.Khryashchev: Building detection on Aerial images using UNET Neural networks. University of Russia (RUDN UNIVERSITY), (ISSN 2305-7254). 2019.

4 V. Khryashchev, L. Ivanovsky, A. Ostrovskaya and A. Semenov, "Application of satellite Image Segmentation for Urban Planning Optimization," Hong Kong, pp. 171-175, 15-17 June, 2019.

5   Qing Peng Li, Lichao Mou, Qizhi Xu, Yun Zhang, and Xiao
    Xiang Zhu: R-Net: A Deep Network for Multi-oriented Vehicle
    Detection in Aerial Images and Videos. Pp: 1-14, 16 Aug 2018.

6   Heng Liu, Mingqiang Guo, Yongyang Xu and Ying Huang:
    Building Extraction Based on U-Net with an Attention Block
    and Multiple Losses. University of Geosciences, April 2020

7   O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional
    Networks for Biomedical Image Segmentation, Computer
    Science Department University of Freiburg, Germany, pp. 200–
    255, 2015.

8   Saad ALBAWI, Tareq Abed MOHAMMED, and Saad AL-
    ZAWI: Undestanding o a covolution neural networks,
    Department of Computer Engineering Faculty of Engineering
    and Architecture Istanbul Kemerburgaz University Istanbul,
    Turkey, pp: 1-7, ICET2017.

9   Discovering Casablanca Africa-ata.org. Retrieved 17 April
    2011.

10 Casablanca, Morocco: Jewishvirtuallibrary.org. Retrieved 17
    April 2011.

11 T.Wang: Machine learning Group; Semantic segmentation,
    University of Toronto, pp: 1- 28, 2018.

12 Artur M. Gafurov, and Oleg P. Yermolayev, Automatic Gully
    Detection: Neural Networks and Computer Vision, Article in
    Remote Sensing, pp: 1-15, May 2020.

13 Guillaume Chhor, Computational and Mathematical
    Engineering, Cristian Bartolome Aramburu, Mechanical
    Engineering, and Ianis Bougdal-Lambert, Aeronautics and
    Astronautics: Satellite Image Segmentation for Building
    Detection using U-net pp: 1-6, 2018.

14 Y. Goodfellow, Y. Bengio, A. Courville, Deep Learning, the
    MIT Press, 800 p, 2016.

15 F.Milletari, N.Navab, S.Ahmad: Fully convolutional neural
    networks for volumetric medical images segmentation pp: 1-2-
    3, 15 Jun2016.

16 J.Long, E.Shelhamer, T.Darrell: Fully Convolutional Networks
    for semantic segmentation, provided by the computer vision
    foundation, pp: 1-10.

17 Pan X, Zhao J. A central-point-enhanced convolutional neural
    network for high-resolution remote-sensing image
    classification. 2017.

18 Kadhim MA, Abed MH. Convolutional Neural Network for
    Satellite Image Classification BT - Intelligent Information and
    Database Systems, College of Computer Science and IT,
    Computer Sciences Department, 2020.