Deep building footprint extraction for urban risk assessment – Remote sensing and Deep learning based approach -

Hicham MHARZI ALAOUI [1], Hassan RADOINE [1], Jêrome CHENAL [2], Hicham Hajji [3], Hassan YAKUBU [1]

[1] School of Architecture, Planning & Design, Mohammed VI Polytechnic University, UM6P, Benguerir 43150, Morocco
[2] École Polytechnique Fédérale de Lausanne, EPFL, 1015 Lausanne, Switzerland
[3] School of Geomatics and Surveying Engineering, IAV Hassan II, Rabat, Morocco

KEY WORDS: Deep learning, building footprint, Attention-Unet, Unet, Remote sensing, Satellite imagery, risk assessment

ABSTRACT:

Mapping building footprints can play a crucial role in urban dynamics moni-toring, risk assessment and disaster management. Available free building footprints, like OpenStreetMap, provide manually annotated building foot-print information for some urban areas; however, frequently it does not en-tirely cover urban areas in many parts of the world and is not always availa-ble. The huge potential for meaningful ground information extraction from high-resolution Remote Sensing imagery can be considered as an alternative and a reliable source of data for building footprint generation. Therefore, the aim of the study is to explore the use of satellite imagery data and some of the state-of-the art deep learning tools to fully automate building footprint extraction. To better understand the usability and generalization ability of those approaches, this study proposes a comparative analysis of the perfor-mances and characteristics of two of the most recent deep learning models such as Unet and Attention-Unet for building footprint generation.

1. INTRODUCTION

In the urban age, the world urban area is growing at a considerable rate and exposing more people to a range of natural hazards and risks. Earthquakes, floods, and other extreme events affect more and more people each year. Land use change is exacerbating the negative effects of disasters and contributing to the transformation of the natural and human environment at an unprecedented rate. Moreover, future projections of climate change would add an additional vulnerability factor, with both an increase in average summer temperatures of 2 to 6 degrees and a decrease in precipitation of the order of 20% by the end of the century on average with a notable increase in the frequency of extreme events(IPCC, 2014; UN-Habitat, 2015). The risks of droughts, floods, fires in wildland-urban interface... will increase accordingly. Sea level rise will also influence the risk of storms and sea swells with higher wave heights and increase the effects of tsunami risk. The International Panel on Climate Change estimates that sea level will rise by an average of 28 to 98 cm by 2100 under different greenhouse gas emission scenarios(IPCC, 2014).

In this context marked by numerous uncertainties, ensuring the urban resilience is becoming an essential urban policy and a smart investment for cities. The urban resilience start, first and foremost, by an accurate risk and vulnerability assessment(Asmita et al., n.d.; UN-Habitat, 2015). Buildings are one of the most important elements-at-risk for risk assessment. A building houses both goods and people, the behavior and response of a building to a specific disaster determines the damage to be incurred as well as the number of people to be injured or killed.

Building footprints can ease the process of risk analysis. Building footprint is then the most important and basic information needed for evaluating the vulnerabilities of a building. It represents the total area of a building and provides a better description of its spatial characteristics compared to a point representation in terms of spatial location, form, distribution, floor space ratio, and relationship between buildings and other objects (topological, orientation, proximity, etc.). Once building footprints are available, attribute information such as building type, number of floors, use, occupancy, etc. can be added and used for vulnerability and risk analysis.

There are several ways to obtain building footprint maps for a given area of interest, either collecting from the available dataset such as a cadastral map, creating a new dataset from ground survey, or creating new dataset using remote sensing data.

Cadastral maps provide a very good and official source of building footprint data. However, this data might not be up to date and lack detailed information and sometimes there are some restrictions to access to this data due to complicated processes involving legal issues and security concerns(Luo et al., 2017). In addition, manual processes for extracting data from such maps can be very time consuming and labor-intensive work.

A ground-based surveying method involve a specific team to perform many surveying and measurement processes to record building attributes and dimensions. This method is extremely expensive and requires much harder work as numerous ground personnel and equipment are hired to perform building survey.

The most efficient way of obtaining building footprints is from the high- resolution imagery (Feng and Zhao, 2009;

Sahar et al., 2010)and remote sensing (Sahar et al., 2010)data captured by satellites, airplanes or drones. However, digitizing building footprints from imagery is a time-consuming task and is commonly done by digitizing features manually. Deep learning models and AI are highly capable of learning these complex semantics of building footprint and can produce superior results. Use this deep learning model and AI to automate the tedious manual process of extracting building footprints, reducing time and effort required significantly(Bischke et al., 2017; Zhang et al., 2019).

The aim of this paper is to explore a state of art deep convolution network, U-Net and Attention-Unet, and its use for automatic semantic image building segmentation from remote sensing imageries. We highlight how the Attention-Unet deep learning architecture can achieve good results in building segmentation.

2. MODELS ARCHITECTURES

2.1   U-net

U-Net is a convolutional neural network that was designed by Olaf Ronneberger et al. 2015 for biomedical image segmentation at the Computer Science Department of the University of Freiburg, Germany. Its architecture is in the form of a U (Figure 1) composed of two paths, a contracting path to capture context (Encoder) and a symmetric expanding path that enables precise localization (Decoder) (Badrinarayanan et al., 2016; Ronneberger et al., 2015):

- The encoder (contraction path): it is composed of 4 blocks, each block consists of i) two layers of 3x3 convolutions, each followed by a ReLU activation function and ii) a 2x2 Max Pooling operation.
  In this path, the size of the image gradually decreases while the depth increases. At each pooling step, the number of channels in the feature maps doubles, starting with 64 in the first block, and ending with 512 in the last block. The goal of the encoder is to capture the context of the input image in order to extract the maximum amount of information contained in the image.

- The decoder (expansion path) is symmetrical to the contraction path with 4 blocks, each block is composed of i) a 2 x 2 up-convolution operation. A concatenation with the corresponding feature map from the contraction path, the white rectangles in figure 1 represent the copied feature maps. The white rectangles in Figure 1 represent the copied feature maps. Two 3x3 convolutions, each followed by a ReLU activation function. The purpose of this path is to obtain general information combining the location information of the downsampling path and the contextual information of the upsampling path.
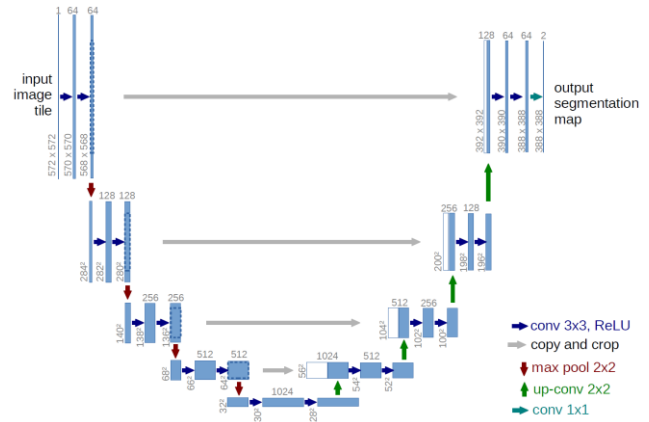

Figure 1: U-net architecture for image segmentation (Ronneberger et al., 2015)

2.2   Attention U-net

There are many advanced variants of the U-Net. The U-Net architecture can be enhanced by using attention blocks called Attention Gate as proposed by Ozan Oktay et al. in 2018. Attention Gate systematically improve the learning rate and and prediction performance of U-Net on different datasets while data while preserving computational efficiency(Oktay et al., 2018; Vaswani et al., 2017). The concept of attention mechanism in Deep Learning, is to be able to focus on parts of interest on a heterogeneous data, be it a sentence or an image, including other useless parts. It helps us to determine which regions R are more relevant to a query q, as illustrated in Figure 2, attention is given to the part of interest that is the default while the plane background receives less attention.

Experimental results show that GAs consistently improve U-Net's prediction performance on different datasets and the learning rate while preserving the computational efficiency, and it slightly outperforms the learning rate while preserving the computational efficiency, and it slightly outperforms the U-Net model(Bischke et al., 2017; Oktay et al., 2018; Vaswani et al., 2017).
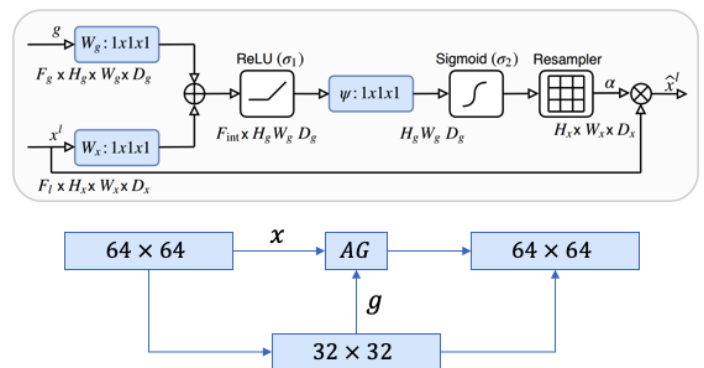

Figure 2: Attention gate (AG) schematic. Bottom: How AGs are implemented at every skip connection

3. EXPERIMENTS

3.1 Dataset

Inria Aerial Image Labeling Dataset (Maggiori et al., 2017) is the dataset used in this study, it is released to address the problem of buildings segmentation in aerial imagery and composed from 360 RGB ortho rectifed aerial images having a spatial resolution of 30 cm. The tiles are of size 5000 x 5000 px and covering a surface of 1500 x 1500 m per tile.

Dataset images are labeled with two classes: building and not building. it covers heterogeneous urban morphology and different urban densities, ranging from densely populated areas (Austin, Chicago) to alpine towns with large green area(Kitsap County, Western Tyrol) to Vienna with its unique architectural style. This variance of the buildings morphology ensures a best generalization of the model to other different urban areas around the world.
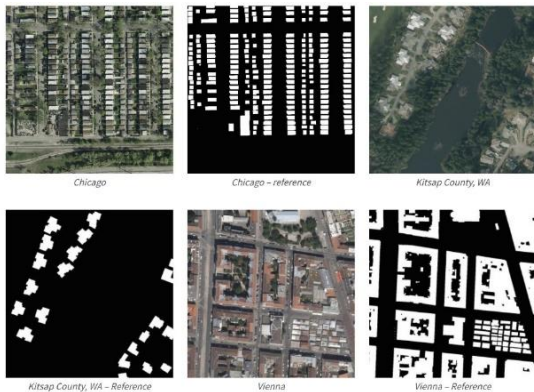

Figure 3: Example of the dataset

3.2 Data augmentation

Collecting and labeling data is a time consuming and expensive, that why when the dataset is small or insufficient, Data augmentation strategy can be used to increase the size of the dataset and generate different versions of a real dataset artificially(Shorten and Khoshgoftaar, 2019). Data augmentation can help models in learning diverse internal representations, which ultimately may lead to improved performance. Generally, 2 techniques are used offline and online augmentation, in this study, "Offline Augmentation" is used to increase the size and add sufficient invariance and robustness to the dataset, different techniques of transformations from horizontal flip, vertical flip, rotations and brightness level have been used.

4. RESULTS AND DISCUSSIONS

Three main metrics widely used to evaluate building footprint generation results have been selected in the following experiments: overall accuracy, F1 score, and intersection over union (IoU):

$$\text{Overall accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$\text{F1 score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN}$$

Where *TP* is the number of building pixels correctly detected, and *FN* denotes the missed building pixels. *FP* and *TN* are the numbers of non-building pixels in the ground reference, but detected as buildings and non-buildings in the result, respectively. The F1 score indicates a balance between precision and recall.

Table 1: Parameters of the model.

Model	IOU	Global Accuracy	Loss	Sensitivity	Specificity
U-Net	0.88	0.91	0.13	0.981	0.988
Attention U-Net	0.92	0.95	0.09	0.976	0.988

Even on small size data, the U-Net Architecture has proven its ability to generate interesting segmentation results with a good accuracy (IOU=0.88) with a reduced computation time. This achievement can be explained by the presence of the skip connection of U-Net effective for improving the accuracy of the extraction and a good segmentation map prediction. The use of the Attention mechanism has achieved better result with an accuracy of IOU=0.92 (figure 4).
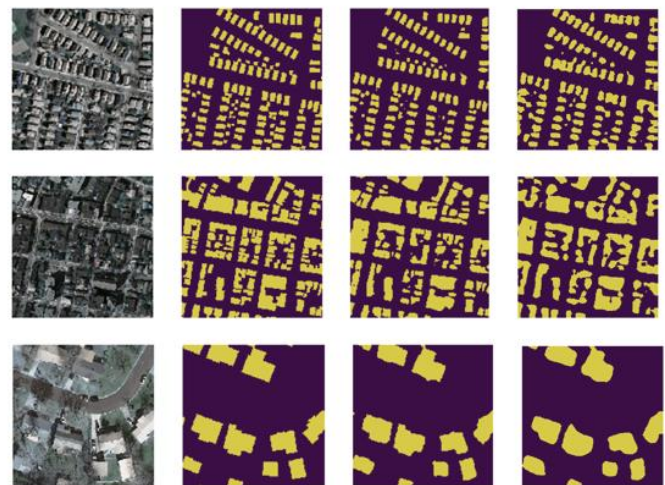

Figure 4: From left to right: input image ground-truth mask, Attention Unet, Unet,

5. CONCLUSION:

This paper presented a Deeplearning technics based on Unet and Attention Unet architecture for building extraction from high-resolution remote sensing images. The adopted architecture aims to generate segmentation maps of buildings and to discriminate the ground truths.
The experiments were conducted on the Inria aerial image labeling dataset for buildings. The experimental results show that the Unet and Attention Unet can both provide good alternative to improve the segmentation performance, while Attention Unet can further give best result and refine the segmentation with little time consumption during the testing stage. The developed model can help urban risk manager to detect with a high accuracy the building footprint as first important information to predict impact disasters on urban areas.

REFERENCES

Asmita, D., Eric,Baker, Judy L. ,Hoornweg, Daniel,Tiwari, n.d. Urban risk assessments : understanding disaster and climate risk in cities [WWW Document]. World Bank. URL https://documents.worldbank.org/en/publication/documents-reports/documentdetail/659161468182066104/Urban-risk-assessments-understanding-disaster-and-climate-risk-in-cities (accessed 9.11.22).

Badrinarayanan, V., Kendall, A., Cipolla, R., 2016. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation.

Bischke, B., Helber, P., Folz, J., Borth, D., Dengel, A., 2017. Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks.

Feng, T., Zhao, J., 2009. Review and Comparison: Building Extraction Methods Using High-Resolution Images, in: 2009 Second International Symposium on Information Science and Engineering. Presented at the 2009 International Symposium on Information Science and Engineering (ISISE), IEEE, Shanghai, China, pp. 419–422. https://doi.org/10.1109/ISISE.2009.109

IPCC, 2014. AR5 Synthesis Report: Climate Change 2014 — IPCC. URL https://www.ipcc.ch/report/ar5/syr/ (accessed 8.13.21).

Luo, X., Bennett, R., Koeva, M., Lemmen, C., Quadros, N., 2017. Quantifying the Overlap between Cadastral and Visual Boundaries: A Case Study from Vanuatu. Urban Sci. 1, 32. https://doi.org/10.3390/urbansci1040032

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark, in: IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D., 2018. Attention U-Net: Learning Where to Look for the Pancreas. https://doi.org/10.48550/ARXIV.1804.03999

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.48550/ARXIV.1505.04597

Sahar, L., Muthukumar, S., French, S.P., 2010. Using Aerial Imagery and GIS in Automated Building Footprint Extraction and Shape Recognition for Earthquake Risk Assessment of Urban Inventories. IEEE Trans. Geosci. Remote Sens. 48, 3511–3520. https://doi.org/10.1109/TGRS.2010.2047260

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0

UN-Habitat, 2015. Enterprise risk management implementation guidelines.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention Is All You Need. https://doi.org/10.48550/ARXIV.1706.03762

Zhang, W., Tang, P., Zhao, L., Huang, Q., 2019. A Comparative Study of U-Nets with Various Convolution Components for Building Extraction, in: 2019 Joint Urban Remote Sensing Event (JURSE). Presented at the 2019 Joint Urban Remote Sensing Event (JURSE), IEEE, Vannes, France, pp. 1–4. https://doi.org/10.1109/JURSE.2019.8809055