URBAN FABRIC SEGMENTATION ON HIGH RESOLUTION SATELLITE IMAGES USING FULLY CONVOLUTIONAL NEURAL NETWORKS

El Mendili Lamiae [a], Chougrad Mehdi [a], Sebari Imane [a], Puissant Anne [b]

[a] Cartography–Photogrammetry department, School of Geomatics and Surveying Engineering, Hassan II Institute of Agronomy and Veterinary Medicine

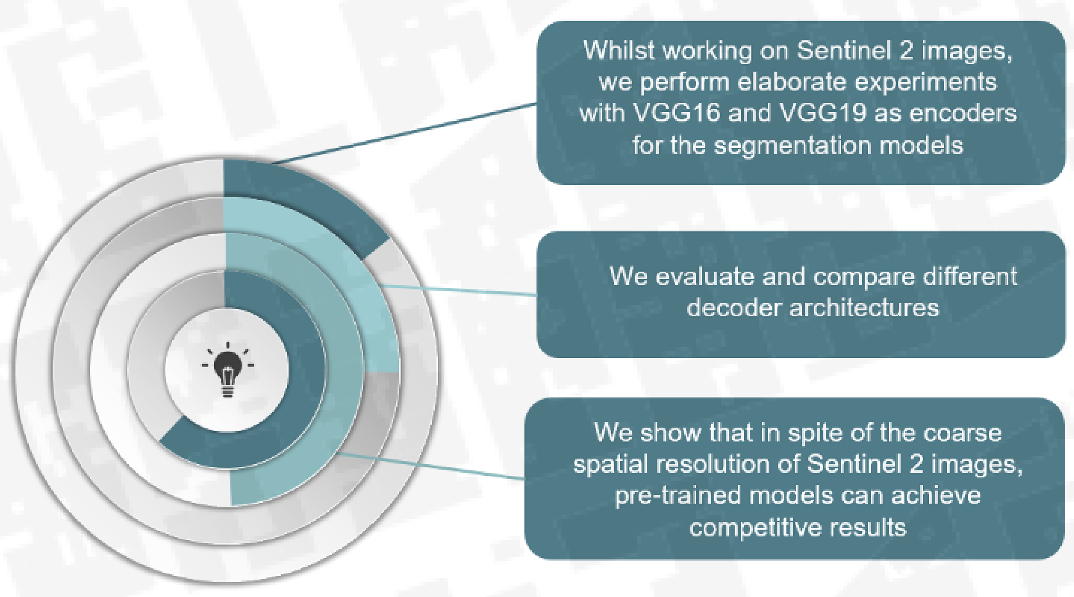
[b] Image City Environment Lab University of Strasbourg

lamiae.elmendili@gmail.com, mehdi.chougrad11@gmail.com, i.sebari@iav.ac.ma , anne.puissant@live-cnrs.unistra.fr

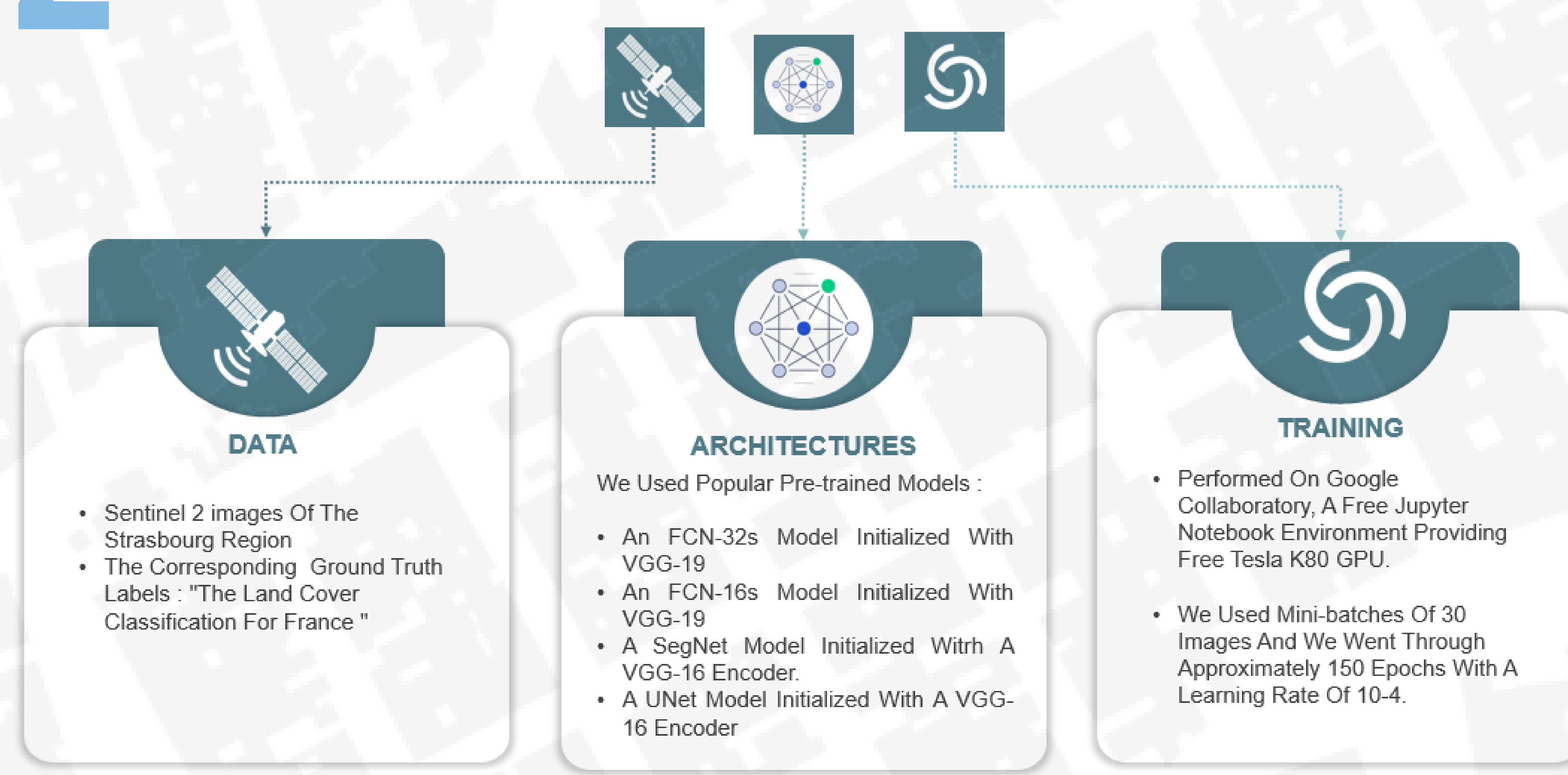
Morocco IndabaX

1 INTRODUCTION

Gathering information about the surface of the earth and particularly its urban component is one of the core fields of remote sensing. The acquired data allows for a better understanding of our living space and usually supports the decision-making process in numerous domains. The complexity of the urban fabric, paired with environmental and socioeconomic challenges has made enhancing our knowledge base about urban forms an increasingly urgent task. **Our main objective** through this work is to test deep learning methods for urban fabric mapping and for the recognition and the monitoring of urban forms. **The contributions** of this work can be summarized as follows :



2 EXPERIMENTAL SETUP



We compared our results with those of [1] who also used different models for the segmentation of buildings footprints from very high resolution satellite images.

	mean IoU	Acc. (Pixel)
Baseline FCN [2]	53.82%	92.79 %
Baseline FCN + MLP[2]	64.67%	94.42 %
FCN (VGG16 encoder)	66.21%	94.54 %
FCN + MLP (VGG16 encoder)	68.17%	94.95 %
SegNet (VGG16 encoder)	**70.14%**	**95.17** %

Figure 7 : The prediction accuracies obtained by [1] on the INRIA aerial image labeling dataset

Our overall accuracy for the SegNet based architecture is slightly inferior to the authors findings, this can be attributed to the coarse resolution of Sentinel 2 images (10 m) in comparison with the Inria Aerial Image Labeling Dataset used by the authors (0.3 m). On the other hand, the UNet architecture improved the overall accuracy by about 1.06 %.

We present below the predictions outputted by our models on some validation data:

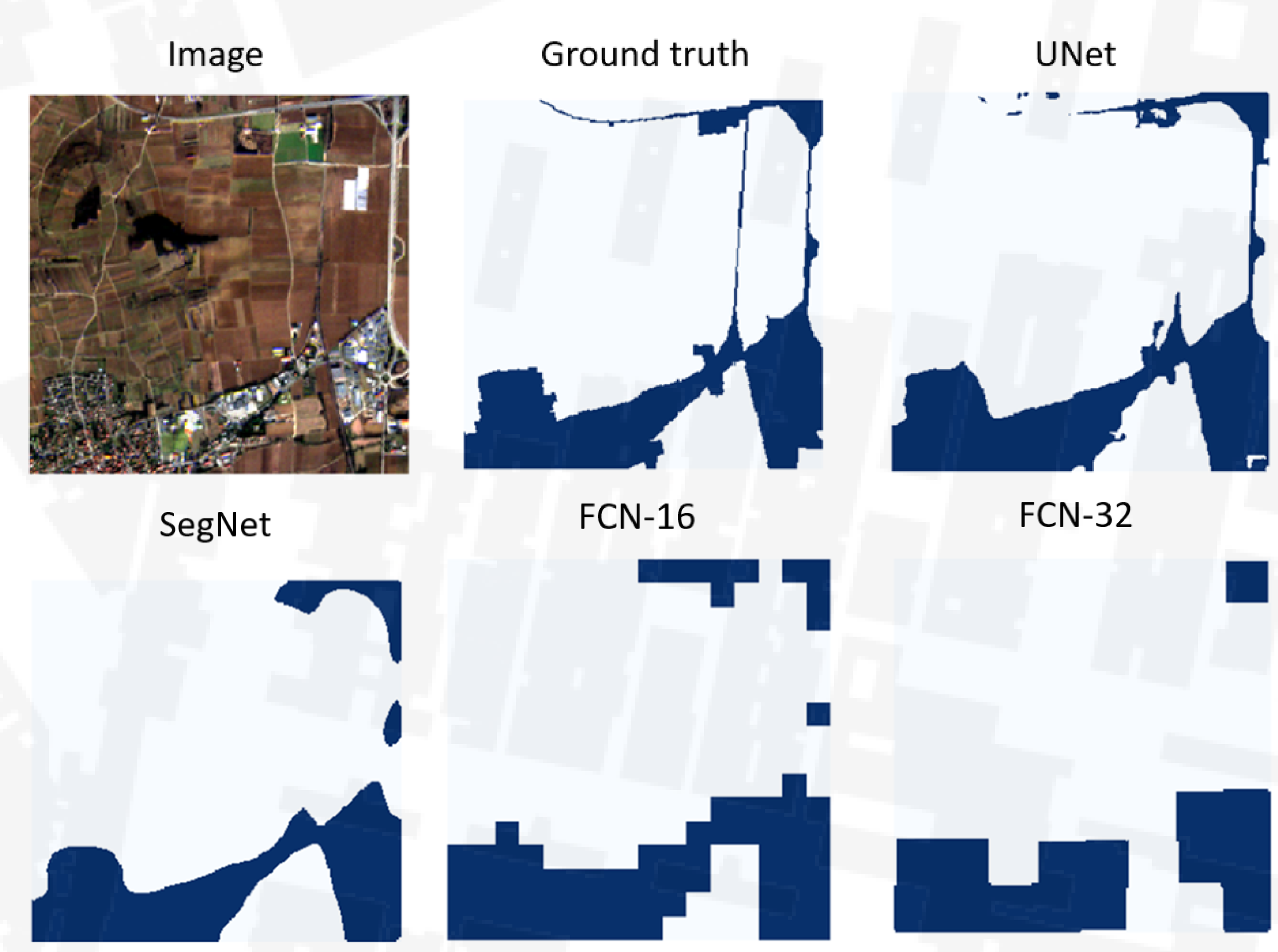


Figure 8 : One example input image with different output predictions

3 RESULTS

The metric used to evaluate the score of our training was the Dice Coefficient. The loss was set to the opposite of the dice coefficient.

We present below the learning curve of our models during training for both the training set and the development set.

FCN-32S

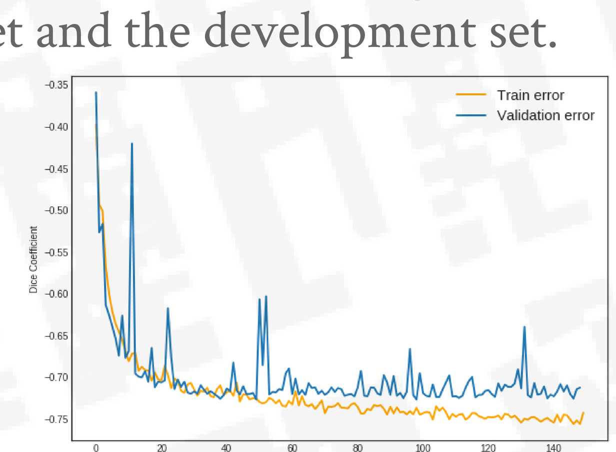


Figure 1 : Learning curve during training for FCN-32S

UNET

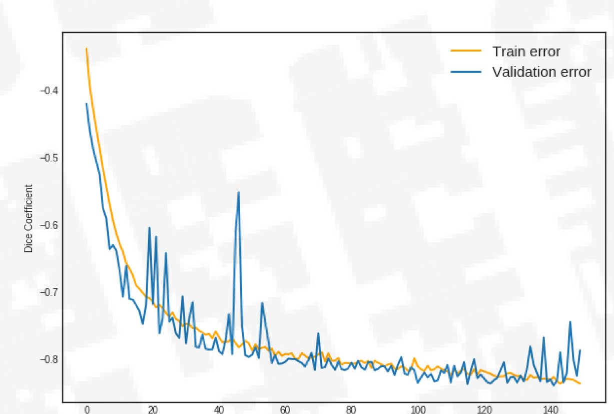


Figure 2 : Learning curve during training for UNET

SEGNET

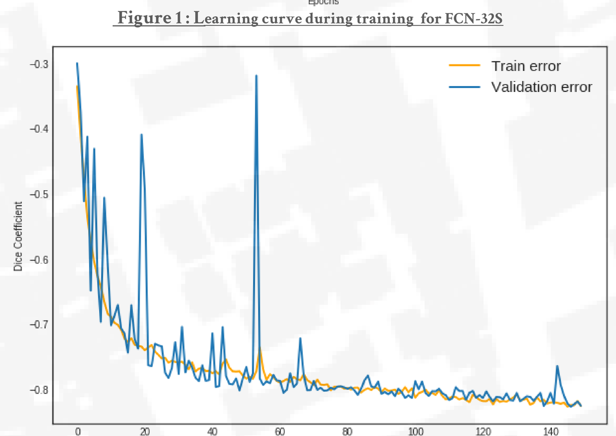


Figure 3 : Learning curve during training for SFGNET

FCN-16S

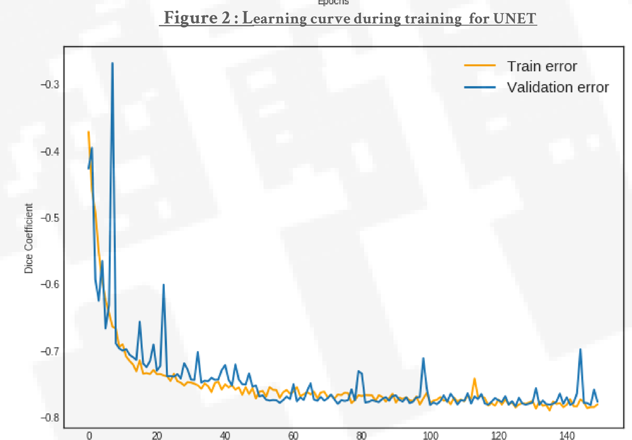


Figure 4 : Learning curve during training for FCN-16S

As shown below, the UNet architecture achieved the best results converging towards a dice coefficient of approximately 0.85 for both the training set and the development set .

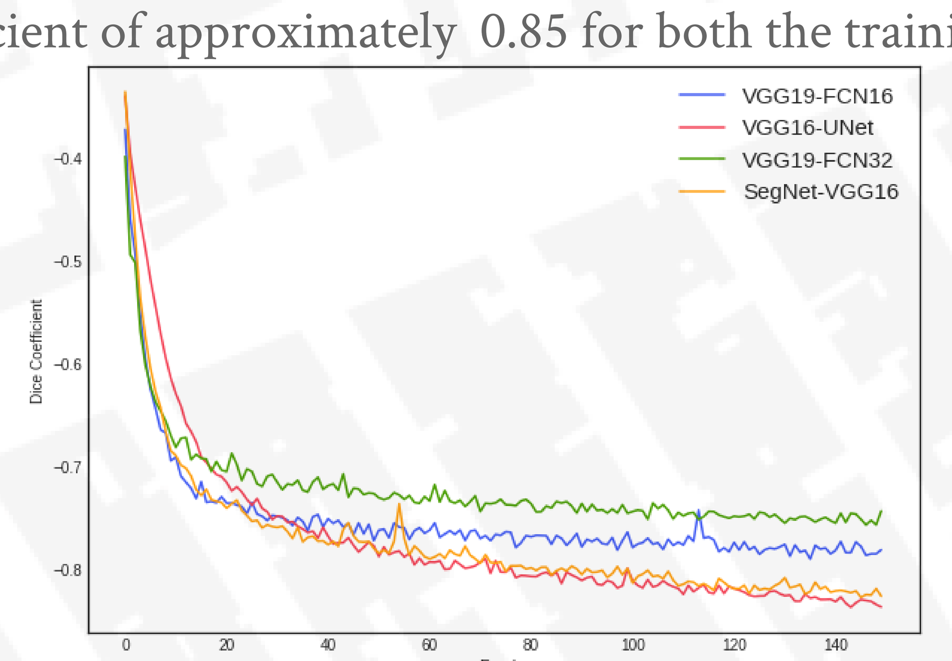


Figure 5 : Comparison of learning curves for different models

Evaluating our trained models on the development set, we obtained the following results :

Modèle	Overall Accuracy	F1 score	Intersection over union	Kappa Coefficient
FCN32-VGG19	91.98 %	76.65 %	62.15 %	71.82 %
FCN16-VGG19	92.02 %	79.70 %	66.26 %	74.79 %
SegNet-VGG16	94.84 %	83.10 %	71.09 %	80.00 %
UNet-VGG16	96.23 %	88.24 %	78.96 %	86.00 %

Figure 6 : Validation results for differeFinal results

4 FUTURE WORK :

In this work, we investigate deep neural networks for the urban fabric semantic labeling using Sentinel 2 data.

Building upon these findings, we plan to extend our models to multiple classes of urban morphology . In this context we also plan to make use of temporal cues together with spectral cues to help recognize more accurately different classes of urban forms. This way, we intend to quantify the improvements that multi-temporal images bring to the above problem.

REFERENCES

[1] : Bischke.B, Helber. P, Folz.J, Borth.B , Danjel.A . (2017).Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks. eprint arXiv:1709.05932CoRR abs/ 1709.05932 (2017): n. pag.[[2Marmanis, D., WegneDr, J.., Galliani, S.,

[2] : Schindler, K., Datcu, M., & Stilla, U. (2016). SEMANTIC SEGMENTATION OF AERIAL IMAGES WITH AN ENSEMBLE OF CNNS.

[3] : Iglovikov, Vladimir & Mushinskiy, Sergey & Osin, Vladimir. (2017). Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition.

[4] : Audebert, Nicolas & Le Saux, Bertrand & Lefèvre, Sébastien. (2017). Beyond RGB: Very High Resolution Urban Remote Sensing With Multimodal Deep Networks. ISPRS Journal of Photogrammetry and Remote Sensing. 10.1016/ j.isprsjprs.2017.11.011.

[5] : Kampffmeyer, Michael & Salberg, Arnt-Børre & Jenssen, Robert. (2016). Semantic Segmentation of Small Objects and Modeling of Uncertainty in Urban Remote Sensing Images Using Deep Convolutional Neural Networks. 10.1109/ CVPRW.2016.90.

[6] : Mahdianpari, Masoud & Salehi, Bahram & Rezaee, Mohammad & Mohammadimanesh, Fariba & Zhang, Yun. (2018). Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. Remote Sensing. 10. 1119. 10.3390/ rs10071119.