# Exploring inference of a land use and land cover model trained on MultiSenGE dataset

Romain Wenger
*LIVE UMR 7362 CNRS*
*University of Strasbourg*
F-67000 Strasbourg, France
romain.wenger[at]live-cnrs.unistra.fr

Anne Puissant
*LIVE UMR 7362 CNRS*
*University of Strasbourg*
F-67000 Strasbourg, France
anne.puissant[at]live-cnrs.unistra.fr

Jonathan Weber
*IRIMAS UR 7499*
*University of Haute-Alsace*
F-68100 Mulhouse, France
jonathan.weber[at]uha.fr

Lhassane Idoumghar
*IRIMAS UR 7499*
*University of Haute-Alsace*
F-68100 Mulhouse, France
lhassane.idoumghar[at]uha.fr

Germain Forestier
*IRIMAS UR 7499*
*University of Haute-Alsace*
F-68100 Mulhouse, France
germain.forestier[at]uha.fr

***Abstract*—Land use and land cover and Urban Fabric (UF) mapping are very useful for urban modeling and simulation (growth, pollution, noise, micro-climate, mobility) in a context of global change. In recent years, due to the increase of Earth Observation data researchers built and shared datasets to the machine learning scientific community to apply and test their models for semantic segmentation. Few works have trained their deep learning model based on a geographic region and applied it to another geographical area of the same country. In this study, we explore the inference of a deep learning model pretrained on a multitemporal and multimodal dataset named 'MultiSenGE' dataset (based on a region in the East of France representing one fifth of the area of France) on five different cities over France (Toulouse, Dijon, Orléans, Lille, Rennes) away from the trained area. Results are encouraging and achieve F1-Score between 0.70 and 0.80 for the five urban fabric classes.**

***Index Terms*—inference, semantic segmentation, Sentinel-1, Sentinel-2, land cover, time series, urban fabric**

## I. Introduction

For many years, urban sprawl and global warming have been at the heart of the scientific community's concerns. Indeed, by 2050, no less than three out of four inhabitants will be living in cities [1]. This urban sprawl causes more and more pressure on ecosystems and consumes many natural areas essential for the preservation of biodiversity. In many cities, urban green areas are locate in the private and public domain and are essential to maintain urban cool islands [2] especially in a context of global change. Thus, urban planners need frequent and up-to-date land cover land use (LULC) mapping in order to detect and quantify changes at the district or city level.

International space agencies have set up Earth observation programs with the deployment of many satellites, whether in passive remote sensing (optical) but also active remote sensing (radar). This is the case of the ESA (European Space Agency) and the Copernicus program who launched in orbit Sentinel constelation which produce every day several TB of data ($\sim$ 15 TB each day).

In order to process this increasing amount of data, researchers have developed methods based on artificial intelligence and more particularly neural networks. These methods need a large amount of training data to be efficient [3].

Semantic segmentation is one of the methods in the field of artificial intelligence. It assigns a semantic class to each pixel of an image and produce a mask according to a probability of belonging to each class [4]. Also, one of the main advantage of this method is its ability to recognize a set of category that form a cluster of pixel of the same class. It reduce the salt and pepper noise resulting from classical pixel approches [5]. This method is used to produce a LULC map [6]. One of the most widely used semantic segmentation networks is the U-Net [7] which has already proven its efficiency, especially in urban fabric (UF) semantic segmentation works [8], [9].

Features fusion has been widely used by the community to perform multitemporal and/or multimodal semantic segmentation [10]. Temporal dynamics of land cover objects are retrieve from multitemporal imagery and properties and structural characteristics are provided by optical and radar [11] imagery. The synergy of these two aspects shows interesting results for natural areas [12] and urban fabric mapping [13].

In previous work [14], it has been shown that the contribution of multitemporal and multimodal imagery improve UF semantic segmentation. Thus, a convolution network has been developed and trained on the MultiSenGE [15] dataset which was built for the entire Grand-Est region in France. It takes as input multitemporal Sentinel-1 and Sentinel-2 imagery.

The objective of this work is to use this pretrained network to classify five cities in France. Through this approach, we would like to explore the genericity of deep learning models over cities located far from of the training area.

## II. Methods

This section describes the datasets and the methods used to perform semantic segmentation over five cities in France.

 DOI: https://doi.org/10.1109/JURSE57346.2023.10144156

### A. Datasets

In this work, we have chosen to use MultiSenGE [16] which is a land use/land cover (LULC) dataset developed over the entire Grand-Est region in France. It contains 8,157 multitemporal and multimodal patches ($256 \times 256$) cut from the Sentinel-1 and Sentinel-2 time series with represents 14 Sentinel-2 tiles. This dataset was developed for the year 2020, which corresponds to the year of production of the reference data, OCSGE2-GEOGRANDEST (https://www.geograndest.fr), used to match the satellite data. It has also been reprocessed to be coherent with the spatial resolution of the satellite images.

The Sentinel-2 images used by MultiSenGE have been uploaded through the Theia portal (https://www.theia-land.fr/). They are offered to users at L2A level which corresponds to a correction of the atmosphere effects, slope effects and the availability of a cloud mask. This dataset used only images containing less than 10% cloud cover and each Sentinel-2 patch is composed of a stack of bands at 10m (B2, B3, B4, B8) and 20m (B5, B6, B7, B8A, B11 and B12) spatial resolution.

Sentinel-1 data were downloaded using the *s1tiling*[1] processing chain, developed by CNES (Centre National des Etudes Spatiales). Only the Ground Range Detection (GRD) products have been kept and the Sentinel-1 patches are composed of a stack at 10m spatial resolution of the VV and VH radar bands.

In previous work, the reference data [14], initially in 14 classes, was reclassified into 10 classes by merging the least represented classes (Table I). Indeed, these classes represented less than 0.1% of the total area of the region and were not spatially homogeneous over the territory, which causes problems for the training and classification, even when applying a weighted loss.

TABLE I
SEMANTIC CLASSES FOR MULTISENGE AND 10 CLASSES SEMANTIC SEGMENTATION (ADAPTED FROM [14])

| MultiSenGE semantic classes | 10 classes |
|---|---|
| Dense Built-Up (1) | Dense Built-Up (1) |
| Sparse Built-Up (2) | Sparse Built-Up (2) |
| Specialized Built-Up Areas (3) | Specialized Built-Up Areas (3) |
| Specialized Vegetative Areas (4) | Specialized Vegetative Areas (4) |
| Large Scale Networks (5) | Large Scale Networks (5) |
| Arable Lands (6) | Arable Lands (6) |
| Vineyards (7) | Vineyards and Orchards (7) |
| Orchards (8) | |
| Grasslands (9) | Grasslands (8) |
| Groces, Hedges (10) | Forests and semi-natural areas (9) |
| Forests (11) | |
| Open Spaces, Mineral (12) | |
| Wetlands (13) | Water Surfaces (10) |
| Water Surfaces (14) | |

Five different cities composed of the same typology of urban fabrics to the Grand-East cities in France have been chosen to infered a model trained on a region located several hundred of kilometers away : Toulouse, Dijon, Orleans, Lille and Rennes (Table II).

[1]https://github.com/CNES/S1Tiling

TABLE II
FRENCH CITIES SELECTED FOR SEMANTIC SEGMENTATION

| City | Tile | Surface | Inhabitants |
|---|---|---|---|
| Toulouse | T31TCJ | 118 $km^2$ | 400,000 |
| Dijon | T31TFN | 40 $km^2$ | 160,000 |
| Orléans | T31UDP | 27 $km^2$ | 116,000 |
| Lille | T31UES | 34 $km^2$ | 235,000 |
| Rennes | T30UWU | 50 $km^2$ | 220,000 |

### B. Multitemporal and multimodal network

In this study, a multitemporal and multimodal network pretrained on MultiSenGE was used (Fig. 1). It was initially pretrained on 4 Sentinel-2 dates spaced at least 17 days apart and the first available Sentinel-1 date per patch [14].

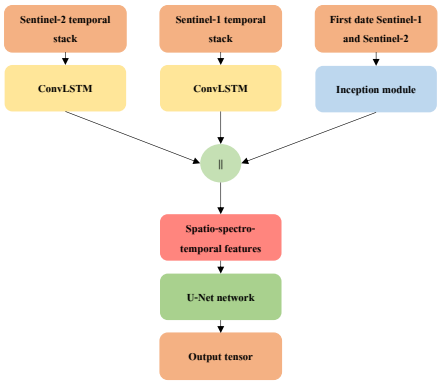


Fig. 1. Pretrained multitemporal and multimodal network (*ConvLSTM+InceptionS1S2* presented in [14])

This method consists of a multitemporal and multimodal stack of the selected Sentinel-1 and Sentinel-2 patches as well as a monodate but multimodal stack of the first Sentinel-1 and Sentinel-2 dates of each patch. Then, a spatio-temporal feature extractor (ConvLSTM) is applied to both the Sentinel-1 multitemporal series and the Sentinel-2 multitemporal series. On the other hand, an Inception module, allowing the extraction of spatio-spectral features is applied this time on the monodata stack. The computed features are stacked and passed in a U-Net to compute a LULC map. The network thus pre-trained on the MultiSenGE dataset can be applied and tested on other tiles focused on other cities in France. All the method is fully described in [14].

### C. Semantic segmentation process

We chose to classify the set of each Sentinel-2 tile that we have previously segmented into patches of $256 \times 256$ pixels, the size of the input patches of the initial network. The areas of the cities studied are then cut out for each semantic segmentation result. Sentinel-2 images were downloaded according to several criteria :

- Less than 10% cloud cover and if no images are available on the desired dates, the lowest cloud cover available

- A complete image not containing no-data values (from the 290km swath of Sentinel-2)
- One image per month by maximizing the initial selection criteria used to train the network (17 days difference between two dates for the months of July, August, September and November)

Concerning Sentinel-1, we chose the first complete tile (still without no-data) preprocessed and sliced by the *s1tiling* processing chain, which does not change from the initial training method. These images has been downloaded in GRD format like MultiSenGE dataset.

### D. Semantic segmentation evaluation

The evaluation of the semantic segmentations was primarily based on a qualitative assessment, as similar baseline data was not available at the selected study sites. Nevertheless, in order to complete with a quantitative assessment, a ground thruth is manually digitized for five urban fabrics (the sixth class correspond to the aggregation of all natural classes present in Table I) for each of the five cities. This digitization (Fig. 2) based on a Sentinel-2 image as a background is applied on five areas of 4 $km^2$ each for each city from the urban centre towards the suburbs area. Also, Urban Atlas[2] is also used for each city as an additionnel decision support to discriminate very complex areas.

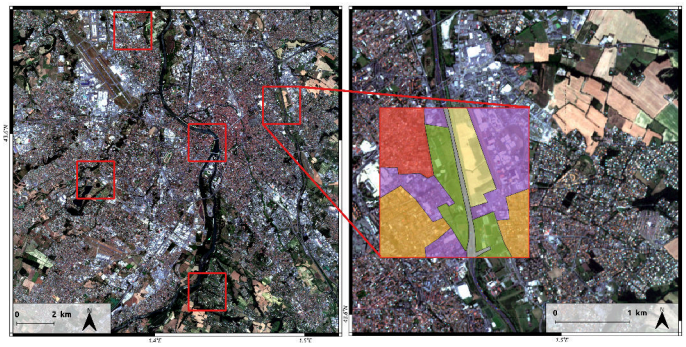

Fig. 2. Digitization areas over Toulouse, France (Legend can be seen through Table I)

The quantitative assesment is performed based on $F1_{Score}$ metric calculated for each city and for each class as it represent the harmonic mean of precision and recall.

## III. Results

This section describe the quantitative and qualitative results[3] obtained for the semantic segmentation of each city using a pretrained multitemporal and multimodal network. Results over training region (Grand-Est, France) can be seen in [14].

### A. Quantitative assessments

Weighted $F1_{Score}$ metric was computed for each city over the five subset areas (Table III). We also performed $F1_{Score}$ for each class in Table IV. As a reminder, class (6) is the concatenation of all natural classes (class (6) to (10) in Table I). As shown in Table III, Dijon is the city where the weighted $F1_{Score}$ is the higher with a score of 0.8087 following by Orléans with 0.7742. Lille has the least score with 0.6866.

TABLE III
WEIGHTED $F1_{Score}$ FOR EACH CITY.

| Cities | Score |
|---|---|
| Toulouse | 0.7029 |
| Dijon | **0.8087** |
| Orléans | 0.7742 |
| Lille | 0.6866 |
| Rennes | 0.7187 |

Concerning each class, we clearly see that Dijon has the best $F1_{Score}$ per class for 4 out of 6 classes. Also, it outperform the other cities for class (1), (3) and (5) as seen in Table III. As presented in [14], class (4) remains complex to classify, even in dense urban areas. The same observation can be applied for class (5). Large Scale Networks are complex to detect in cities as they are often less then one pixel wide.

TABLE IV
$F1_{Score}$ PER CLASS FOR EACH CITY.

| Cities selected | Classes | | | | | |
|---|---|---|---|---|---|---|
| | *(1)* | *(2)* | *(3)* | *(4)* | *(5)* | *(6)* |
| Toulouse | 0.7137 | 0.7255 | 0.7021 | **0.6072** | 0.3855 | 0.7819 |
| Dijon | **0.8005** | 0.8140 | **0.7774** | 0.5363 | **0.6154** | **0.8996** |
| Orléans | 0.6408 | **0.8462** | 0.6989 | 0.4215 | 0.4885 | 0.8355 |
| Lille | 0.6315 | 0.7864 | 0.6360 | 0.5835 | 0.4930 | 0.8082 |
| Rennes | 0.6446 | 0.6892 | 0.7012 | 0.5209 | 0.5666 | 0.8589 |

### B. Qualitative assessments

Land cover maps presented in Fig. 3 confirmed encouraging semantic segmentation results. Salt and pepper classification noise, which was often detected for pixel approaches (e.g. Random Forest) and especially for UF mapping, is almost none for these results. For each city, urban city centre is well identified and correspond to Dense Built-Up for western cities.

As showed in Fig. 3, urban boundaries are well classified on almost each city. The most important confusion is with class (4), which represents Specialized but Vegetative areas (Table I). These areas are correctly classified when they are inside dense urban areas but, at the boundaries of the city, we can relate some confusion with natural areas, especially for Orléans and Rennes. The difference in performance between the cities could be explained by the difference in vegetation between the territories but also by the morphological and spectral difference of each city. Indeed, Dijon being the closest city to the training region, it presents the best global scores.

## IV. Conclusion and Perspectives

This paper has shown the capacity of a deep learning model, trained over a specific french region, to generalize over others cities in France. The model trained over Grand-Est region in France (MultiSenGE dataset) was applied for five cities

[2]https://land.copernicus.eu/local/urban-atlas/urban-atlas-2018
[3]https://romainwenger.fr/multisenge/Cities.html

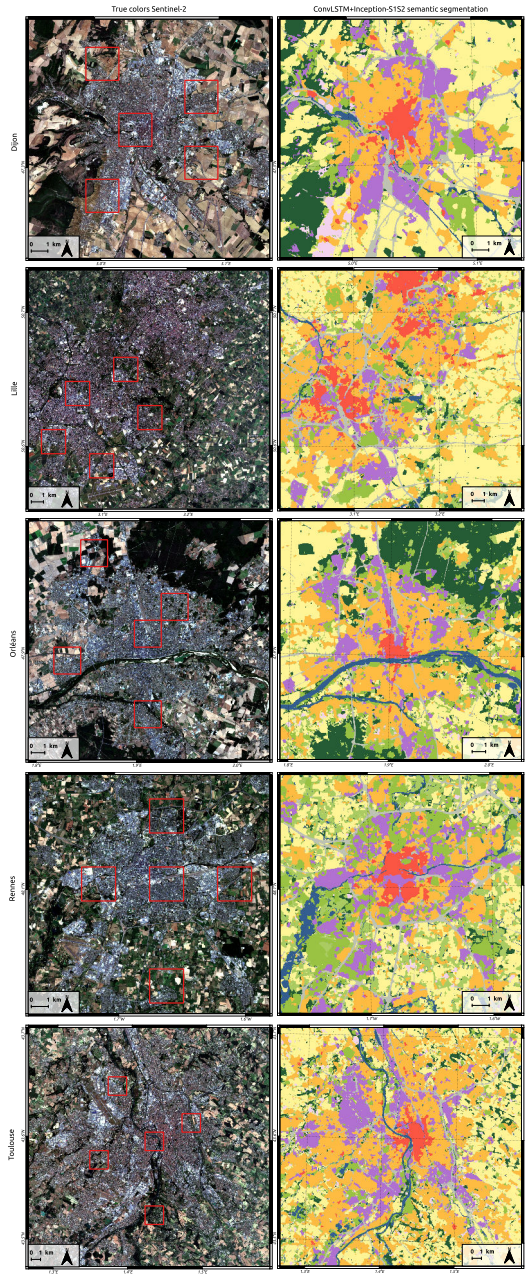


Fig. 3. Qualitative results and digitized area for the five cities studied (Legend can be seen through Table I)

in France to evaluate the genericity of a LULC multitemporal and multimodal semantic segmentation network. Results are encouraging and illustrates that models trained on small datasets can achieve great performance over other geographical areas for UF mapping. Thus, it would not be mandatory to perform transfer learning for UF mapping over France even if small convergence of the weights, as a results of a new training phase over local ground reference data, could improve current results. Work in progress is exploring this inference over european and north african cities to investigate the impact of different climate areas. Also, we are investigating temporal inference, which consists in the semantic segmentation of the same training area but for one or two different years to perform change detection.

## ACKNOWLEDGMENT

We thanks to the Spatial Data Infrastructure GeoGrandEst provided the reference data used in this study. We also would like to thanks the computer center Mesocentre for providing the calculation ressources, ANR TIMES [ANR-17-CE23-0015] and the French TOSCA project AIMCEE [CNES, 2019-2022].

## REFERENCES

[1] United Nations Department of Economic and Social Affairs Population Division. 2018. “The World’s Cities in 2018.” Accessed 9 February 2021. https://population.un.org/wup/Publications/Files/WUP2018-Report.pdf

[2] X. Yang, Y. Li, Z. Luo, and P.W. Chan, The urban cool island phenomenon in a high-rise high-density city and its mechanisms. 419 International Journal of Climatology 2017, 37, 890–90

[3] A. Anaby-Tavor, et al. ”Do not have enough data? Deep learning to the rescue!.” Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 05. 2020.

[4] L. Ma, Y. Liu, X. Zhang, Y. Ye, and G. Yin, Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing 2019, 152, 166–17

[5] H. Hidetake, C. S. Ram, T. Mizuki, and H. Keitarou, (2018): Evaluating multiple classifier system for the reduction of salt-and-pepper noise in the classification of very-high-resolution satellite images, International Journal of Remote Sensing.

[6] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, Deep Learning Classification of Land Cover and Crop Types Using Remote 477 Sensing Data. IEEE Geoscience and Remote Sensing Letters 2017, 14, 778–782

[7] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation. CoRR 2015, 462 abs/1505.04597

[8] R. Wenger, A. Puissant, J. Weber, L. Idoumghar, G. Forestier, U-Net feature fusion for multi-class semantic segmentation of urban fabrics from Sentinel-2 imagery: an application on Grand Est Region, France, International Journal of Remote Sensing, 43:6, 1983-2011.

[9] L. El Mendili, A. Puissant, M. Chougrad, and I. Sebari, Towards a Multi-Temporal Deep Learning Approach for Mapping Urban Fabric Using Sentinel 2 Images. Remote Sensing 2020, 12.

[10] J. Hu, L. Mou, A. Schmitt, X. X. Zhu, FusioNet: A two-stream convolutional neural network for urban scene classification using PolSAR and hyperspectral data. In 2017 Joint Urban Remote Sensing Event (JURSE) (pp. 1-4). IEEE.

[11] N. Clerici, C.A.V. Calderón, and J.M. Posada, Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: a case study in the lower Magdalena region, Colombia. Journal of Maps 2017, 13, 718–72

[12] J. Betbeder, M. Laslier, T. Corpetti, E. Pottier, S. Corgne, and L. Hubert-Moy, 2014. Multitemporal optical and radar data fusion for crop monitoring: application to an intensive agricultural area in brittany (France). In: 2014 IEEE Geoscience and Remote Sensing Symposium, IGARSS 2014, Quebec City, QC, Canada, July 13–18, 2014, 2014, pp. 1493–1496.

[13] G.C. Iannelli, and P. Gamba, Jointly exploiting sentinel-1 and sentinel-2 for urban mapping. In: 2018 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2018, Valencia, Spain, July 22–27, 2018, 2018, pp. 8209–8212.

[14] R. Wenger, A. Puissant, J. Weber, L. Idoumghar, and G. Forestier, “Multimodal and multitemporal land use/land cover semantic segmentation on Sentinel-1 and Sentinel-2 imagery : an application on MultiSenGE dataset”, Remote Sens. 2023, 15, 151.

[15] R. Wenger, A. Puissant, J. Weber, L. Idoumghar, G. Forestier. A new remote sensing benchmark dataset for machine learning applications : MultiSenGE (1.0) [Data set]., March 2022, Zenodo. https://doi.org/10.5281/zenodo.6375466

[16] R. Wenger, A. Puissant, J. Weber, L. Idoumghar, and G. Forestier: MultiSenGE: a multimodal and multitemporal benchmark dataset for land use/land cover remote sensing applications, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2022, 635–640.