



HAL
open science

Deep-Learning Based Multiple Land-Cover Map Translation

Luc Baudoux, Jordi Inglada, Clément Mallet

► **To cite this version:**

Luc Baudoux, Jordi Inglada, Clément Mallet. Deep-Learning Based Multiple Land-Cover Map Translation. IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Jul 2022, Kuala Lumpur, Malaysia. pp.1260-1263, 10.1109/IGARSS46834.2022.9883056 . hal-03983066

HAL Id: hal-03983066

<https://hal.science/hal-03983066v1>

Submitted on 10 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DEEP-LEARNING BASED MULTIPLE LAND-COVER MAP TRANSLATION

Luc Baudoux^{1,*}, Jordi Inglada² and Clément Mallet¹

¹ LASTIG, Univ Gustave Eiffel, IGN-ENSG

² CESBIO, Univ. de Toulouse, CNES/CNRS/IRD/INRAE/UPS

* luc.baudoux@ign.fr

ABSTRACT

This paper presents a framework for simultaneously translating multiple land-cover maps into a given one in a supervised way. Conversely to existing approaches working on 1-1 translation, we propose a multi-translation setup that increases the generalizability and translation performance, especially on land-cover maps covering restricted spatial extents. The proposed method mainly assumes that the map of interest spatially overlaps at least with one of the other maps. High performance translation is achieved with a Convolutional Neural Network (CNN) based encoder-decoder framework trained with three goals: (i) high-quality translation; (ii) self-reconstruction ability; (iii) mapping of all datasets into a common representation space. Country-scale experimental results show the method effectiveness in translating six highly heterogeneous land-cover maps, achieving significantly better results than the traditional semantic-based method and better results than CNN trained for a 1-1 translation task (+ 9.7% in Overall Accuracy (OA) and +12% in macro F1-score (mF1)).

Index Terms— Land-cover mapping, operational, translation, semantic segmentation, country-scale.

1. INTRODUCTION

Land-use/Land-cover (LULC) maps describe the Earth coverage with discrete classes (the *nomenclature*) at a defined spatial resolution. The chosen labels and resolutions highly impact the land-cover map potential usage and subsequent manipulations such as map fusion, harmonisation, comparison or update. The challenge in map-to-map translation lies in the difficult interleaved association of semantic and spatial resolutions of both maps [1]. Usually, two different LULC maps establish complex relationships between their classes, and direct one-to-one association is most of the time infeasible. The standard translation method consists of a nomenclature-level semantic association such as LCCS [2], followed by a spatial resampling strategy. However, such approaches fail to translate complex relationships, acting

as a word-by-word translation, disregarding natural multiple possible associations, by hardly assigning each class to its strongest correspondent in the other nomenclature. Moreover, by processing the nomenclature translation separately from the spatial resolution adaptation, such approaches neglect semantic consideration on pixels holding multiple classes. To relax the constraint of manually defining semantic relations, [3] presented a Latent Dirichlet Allocation solution, which defines the relation between classes by computing their spatial co-occurrence. To obtain multiple translations of each source class, they rely on the use of multiple maps, the translation of one pixel therefore depending on the composition of each multiple source class. Conversely, we proposed to integrate the map spatial context to obtain a context-based translation of each source class, thus alleviating the need for numerous maps on the same extent with close-by dates. Despite encouraging results, the adopted asymmetrical U-Net was only trained to translate one unique source map into one unique target map. Recent advances in natural language processing have shown that learning a single multiple translation model achieved comparable results with learning multiple one-to-one cases. However, the former yields more robust results on languages with few samples and has better generalisation abilities. Thus exploring multiple land-cover translation models appears relevant.

This paper presents a supervised deep learning-based procedure to learn a single model for translating multiple maps without a single remote sensing image. We train our method to translate six land-cover maps with highly varying nomenclatures and scales. We satisfactorily apply our method on the operational case of extending the spatial extent of one land-cover map using another more comprehensive map. Our experimentation, conducted at country-scale, outperforms the traditional semantic-based method, and generalises better than the multiple one-to-one models.

2. MULTI-MAP TRANSLATION FRAMEWORK

Our method includes two main steps: (i) dataset generation; (ii) training a multi-LULC translation. Our framework assumes that each map spatially overlaps with at least one of the others. We do not use additional data such as geospatial

This work is funded by the MAESTRIA project (grant ANR-18-CE23-0023) and supported by the AI4GEO project <http://www.ai4geo.eu/>.

imagery.

2.1. Dataset creation

Since no multi land-cover dataset was available, we created a new one over France, with six land-cover maps, exhibiting varying characteristics (global and continental to local scales, see Table 1). To ensure that the number of land-cover changes remains modest between our maps, we carefully selected as close-by years as possible. The generation includes: (i) a cut-off along the French borders, (ii) a projection into the official French geographical system, (iii) a rasterization step for maps available in vector format preserving the official product resolution (iv) wide $6 \times 6 \text{ km}^2$ tilling procedure to keep enough contextual information even for map with high minimum mapping unit.

The full dataset with more details is available at <https://doi.org/10.5281/zenodo.5843595>.

2.2. Multi-land cover translation

Our goal is to jointly translate the spatial resolution and the nomenclature using spatial context information. Our method relies on Convolutional Neural Networks (CNN) since they efficiently integrate the pixel semantics with its spatial context. We cast our problem as a multi-task one in which one LULC map must be translated into several others. Inspired by recent work on multi-task learning, we propose to train separate encoders (U-Net) and decoders (Spatial Pyramidal Pooling) for each map. All maps are first resampled at the maximum resolution, then go through their respective encoders to provide a joint embedding with the same dimensions (Fig. 1). The resampling back to the original resolution is performed by specified decoders. Encoders and decoders are trained with three goals: (i) translation, (ii) self-reconstruction, (iii) and defining a common latent representation for all maps. The translation strategy implies that a given map A going through its encoder and decoder B should result in map B. This is achieved by using a cross-entropy loss denoted L_{tra} , between the resulting translation and map B. The self-reconstruction goal implies that a given map A going through its encoder and decoder A should result in the original map A. This ensures that the learnt embedding keeps all information. This is achieved by using a cross-entropy loss denoted L_{rec} , between the resulting reconstruction and map A. Lastly, to ensure high generalisation ability, we enforce similar features to be mapped closely into the common representation space across all maps by adding a constraint between the encoded version of two maps covering the same areas. This constraint is enforced through the computation of the mean square error between two maps covering the same area (denoted as L_{emb}). The total loss L is then computed as :

$$L = L_{rec} + L_{tra} + L_{emb}. \quad (1)$$

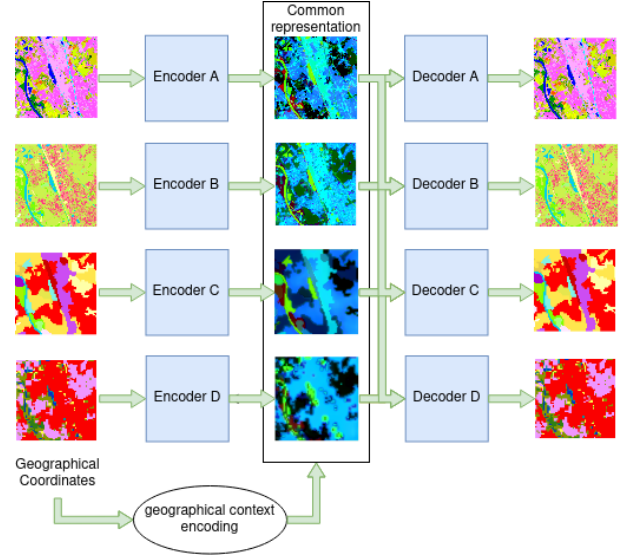


Fig. 1. A multiple cross encoder-decoder is designed to learn to simultaneously translate and self-reconstruct each land-cover map among a heterogeneous set.

As in our previous paper [1], we assume that the translation of a LULC element may co-vary with its spatial coordinates (latitude and longitude). To take into account vast geographical structures (*e.g.*, mountainous areas), we thus insert a geographical-context sub-module taking the coordinates of the patch to translate. Coordinates are encoded using a positional encoder strategy and given to a single hidden layer perceptron. The resulting coordinate embedding is then multiplied by the embedding (Fig. 1).

Source code for the implementation and training details are provided at <https://github.com/LBaudoux/MLULC.git>.

3. EXPERIMENTS

The method is trained on the 6 maps included in our dataset. Fig 2 shows translations of randomly selected pairs of source and target patches based on either: (i) an expert-based pure semantic association between nomenclatures, (ii) the statistical most probable association, (iii) a translation using a CNN explicitly trained for this source/target pair (mono-LC model [1]), (iv) and the results of our multi-LC model. The traditional semantic-based and statistic methods which do not use spatial context fail to replicate the minimum mapping unit of maps like Corine Land Cover, resulting in higher geometric accuracy than the original and predicted CLC. In contrast, mono and multi-LC translations using spatial context give a significantly higher thematic accuracy of the translated maps (visible when translating OSGE-use to CGLS-LC100 in the second column). This finer nomenclature is especially exacerbated when translating classes with no correspondence in

Name	CGLS-LC100 [4]	CLC [5]	OSO [6]	OCS-GE cover [7]	OCS-GE use [7]	MOS
Extent	World	Europe	France	West+South France	West+South France	Paris area
Selected year	2018	2018	2018	2014-2015	2014-2015	2017
Number of classes	12	44	23	14	17	11
Raster spatial resolution	1ha	25ha	0.01ha	0.01ha	0.01ha	0.02ha

Table 1. Main characteristic of the six selected land-cover maps (see references for more details).

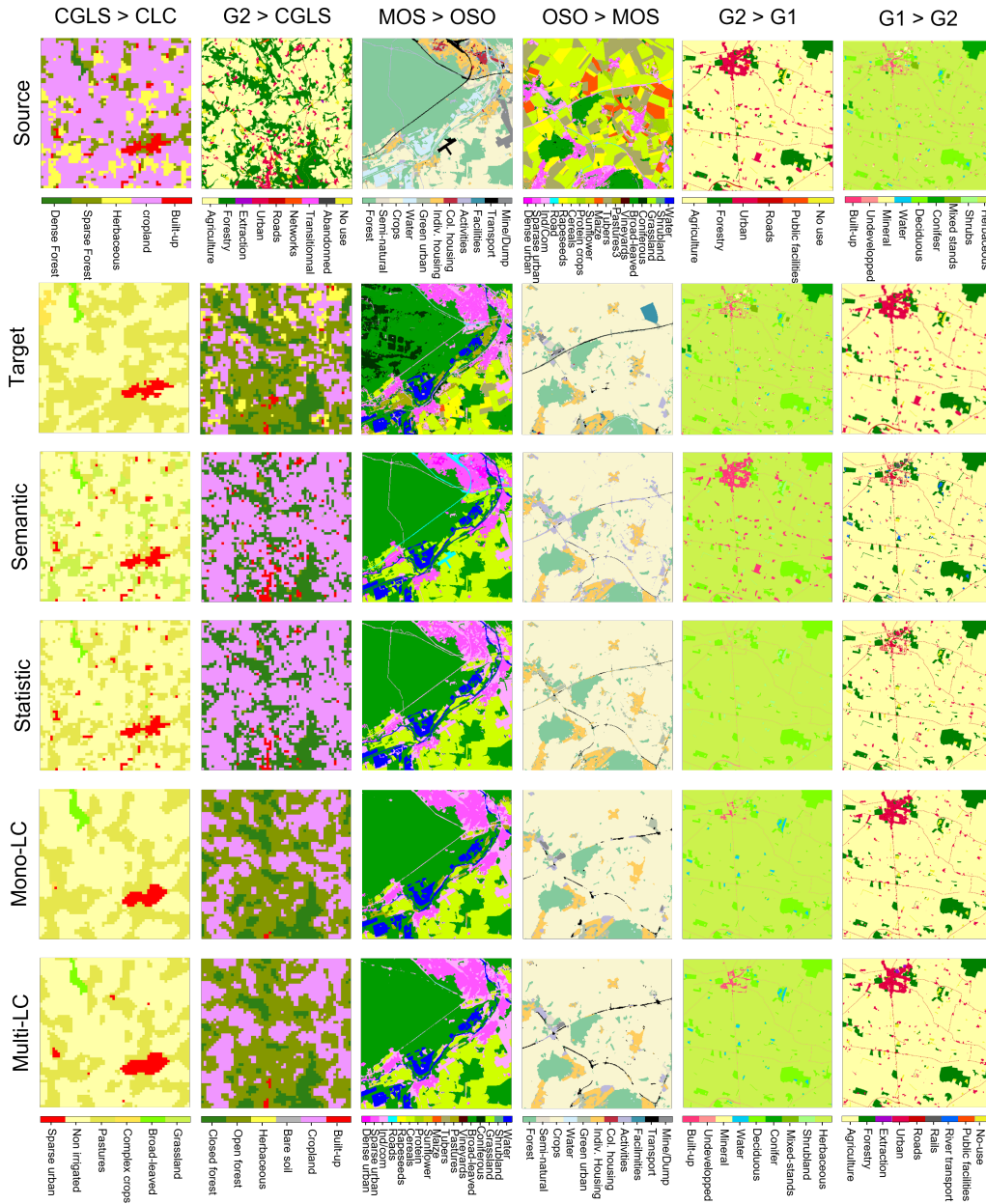


Fig. 2. Comparison between different translation methods with varying source and target maps. *Semantic* refers to a simple expert knowledge-based translation, *Statistic* refers to a translation of each source class by its most frequently associated target class in the dataset, *mono-LC* to a CNN trained on a specific source-target pair, *multi-LC* is our method. *G1* stands for OCS-GE cover and *G2* for OCS-GE use

Source	C					P					O					M			G1				G2				
Target	O	M	G1	G2	P	C	O	M	G1	G2	C	M	G1	G2	P	C	O	P	C	O	G2	P	C	O	G1	P	
OA	semantic	49	79	67	77	65	52	42	75	56	70	59	81	69	76	62	76	59	80	41	34	87	56	40	31	75	57
	statistic	55	79	71	78	68	54	44	75	65	70	61	82	73	80	65	81	62	83	44	49	89	57	40	41	78	57
	mono-LC	59	80	72	80	74	64	57	76	69	78	69	85	80	86	77	83	63	85	58	58	93	71	54	53	79	69
	multi-LC	59	80	72	81	74	64	56	77	69	78	66	85	78	86	76	84	64	86	58	58	92	70	54	53	80	69
mF1	semantic	26	42	36	28	46	13	17	24	22	15	19	38	36	20	38	19	19	38	10	17	27	27	9	8	29	20
	statistic	32	42	33	30	47	13	18	24	19	16	18	39	34	20	36	17	18	32	10	20	27	27	9	10	27	20
	mono-LC	37	41	36	31	57	30	33	31	29	20	39	53	45	26	61	30	23	45	34	31	43	52	29	25	40	52
	multi-LC	35	41	37	26	59	30	29	34	30	19	36	52	43	23	56	36	23	48	34	32	37	50	30	26	43	49

Table 2. Correspondence between our translation and the original target maps. OA: Overall accuracy, mF1: macro F1-score, C: CLC, O: OSO, M: MOS, G1: OCS-GE cover, G2: OCS-GE use, P: CGLS-LC100.

the target nomenclature or on classes defined by their spatial arrangement, such as the CLC class *heterogeneous crops*. For quantitative assessment, we compare the translated map to their respective target with the original resolution (Table 2). The first observation is that the semantic and statistical-based methods give significantly lower results than the mono and multi-LC models, both in terms of overall accuracy (+9.7% in average between multi-LC and semantic models) and mean f1-score (+12% in average). Moreover, the mono and multi-LC scenarios give comparable results but with some disparities depending on the source map used for the translation. When the source is the OSO map, results tend to be a little better for the mono-LC, especially in terms of mean f1-score. This behaviour is mainly attributed to a better ability to translate some specific erroneous classes in the OSO map (higher noise robustness). Conversely, the multi-LULC model tends to slightly outperform the mono-LC model when the source data initially covers a small spatial extent (MOS, OCS-GE use, OCSGE cover), making it more easily generalizable.

However, some classes are poorly translated by all the methods, especially those mixing land-use with land-cover such as *Artificial non-agricultural vegetated areas*. Moreover, the obtained translation performance results are below the six maps’ official expected accuracies, making the method not yet adapted to operational purposes. Lastly, further improvements are still required on geometric accuracy, especially on linear structures (e.g., fourth row of Fig. 2).

4. CONCLUSION

In this paper, a multi-LC translation method for joint nomenclature and resolution translation was presented and satisfactorily assessed at a country wide-scale over France. The fully automatic method relies neither on additional data nor expert knowledge of translated maps. The analysis of the results reveals that the proposed method gives satisfactory results even in difficult situations with no semantic relationship between the source and the target classes. The method significantly outperforms traditional baselines and gives on par results with multiple one-to-one LULC translation methods using only a single model. This paves the way for higher interoperability between land-cover maps, improving result studies on opera-

tional setups such as updating old maps with more recent ones or for map fusion tasks.

The method is also time-efficient since training the multi-LC model France for our 6 maps (resulting in 30 possible translations) is achieved in less than 24 hours with a single Nvidia-V100 and 8 CPUs. For comparison, a mono-LULC model (1 translation) is trained in 12 hours. As future work, we aim to improve the geometric accuracy and evaluate it under various operational constraints.

5. REFERENCES

- [1] L. Baudoux, J. Inglada, and C. Mallet, “Toward a yearly country-scale CORINE land-cover map without using images: A map translation approach,” *Remote Sensing*, vol. 13, no. 6, pp. 1060, 2021.
- [2] A. Di Gregorio, *Land cover classification system: classification concepts and user manual: LCCS*, vol. 2, FAO - United Nations, Rome, 2005.
- [3] Z. Li, J.C. White, M.A. Wulder, T. Hermosilla, A.M. Davidson, and A.J. Comber, “Land cover harmonization using latent dirichlet allocation,” *International Journal of Geographical Information Science*, vol. 35, no. 2, pp. 348–374, 2020.
- [4] M. Buchhorn, M. Lesiv, N.-E. Tsensbazar, M. Herold, L. Bertels, and B. Smets, “Copernicus global land cover layers—collection 2,” *Remote Sensing*, vol. 12, no. 6, pp. 1044, 2020.
- [5] A. Moiret-Guigand, G. Jaffrain, A. Pennec, and H. Dufourmont, “CLC2018 / CLCC1218 validation report,” Tech. Rep., GMES Initial Operations / Copernicus Land monitoring services, 2021.
- [6] J. Inglada, A. Vincent, M. Arias, B. Tardy, D. Morin, and I. Rodes, “Operational high resolution land cover map production at the country scale using satellite image time series,” vol. 9, no. 1, pp. 95, 2017.
- [7] French National Institute for Geographic and Forestry Information (IGN), “OCS GE version 1.1,” Tech. Rep., French IGN, 2016.