E. Colin Koeniguer,
G. Le Besnerais, A. Chan Hon Tong, B. Le Saux, A. Boulch,
P. Trouvé, R. Caye Daudt,
N. Audebert, G. Brigot, P. Godet,
B. Le Teurnier, M. Carvalho,
J. Castillo-Navarro
(ONERA)

E-mail: elise.koeniguer@onera.fr

DOI: 10.12762/2020.AL15-08

Recent Examples of Deep Learning Contributions for Earth Observation Issues

The purpose of this article is to take stock of the progress made in remote sensing thanks to the recent development of deep learning techniques. This assessment is made by means of a systematic presentation of the various activities carried out at ONERA in remote sensing imagery using deep learning methods. It covers a large part of the observation problems: filtering, object detection, land-use classification, change detection, and biomass estimation. In light of these activities, we highlight the practical challenges of deep learning, mainly physical feature definition and training database construction. Some directions for future research are also given, such as the development and use of dedicated remote sensing platforms, hybrid supervised/unsupervised strategies, and the further exploitation of multimodal/multitemporal data.

Introduction

The rise of open-data makes it possible to promote the use of deep-learning in many areas. Also, many experts around the world are wondering about the progress that deep-learning methods will allow in the coming decades, and seek to predict when the performances of automatic algorithms will exceed those of humans, in several tasks: for translating; writing text, driving vehicles, performing surgical operations, etc. It would therefore seem that the more the field corresponds to a broad need, the more deep-learning allows rapid improvements.

In this context, the field of remote sensing is a much smaller field today in terms of the number of users and needs, than more conventional domains such as language, classical images, or text. However, even though the concerned audience is less extended, remote sensing data processing is undergoing the same revolution as other sectors of big data. The number and the diversity of sensors increase very quickly, at two levels:

- The rise of open data, in particular through data from the European Copernicus observation program, which delivers free images acquired since the end of 2014, and produces a petabyte of data every six months [39].
- The development of commercial-type data and the democratization of satellite systems, with more and more launching constellations of micro or even nano-satellites [15, 11].

This entrance into the big data movement promises exciting developments of deep learning methods for remote sensing images. All the more so since remote sensing data are mostly images, and since the most significant improvements in deep learning were made recently on images in computer vision through the ImageNet dataset [29].

Implementing deep learning methods requires a certain number of choices to be made: which databases and how to access them, which servers or computing power, and which learning architectures for a given function.

First, to successfully implement a deep learning application, it is necessary to have access to massive datasets. These are essential to the performance of artificial intelligence systems and, therefore, highly strategic. The quantity and nature of these data will guide the technological choices. Similarly, if the data are subject to protection constraints, such as military data, they will have to be processed on dedicated servers, which then directs the technological choices.

In any case, the data-processing tool remains the core of a learning method. Many companies compete in this market: big names like Amazon; Google with TensorFlow, Caffe (a project initiated at the University of California at Berkeley), and Torch (widely used and improved by Facebook engineers), can be used for remote sensing images.

However, the functions deduced from the remote sensing images that we are trying to learn are of a very different nature from that of those for which these large classical networks have been developed. For example, we can mention some specificities of remote sensing images compared to the standard images of the web:

- A remote sensing image contains multiple objects, spatially distributed and organized on several pixels. When we deal with remote sensing classification, it is not a label of a single image that we are trying to find, but rather different labels for all of the pixels or group of pixels in the image.
- Remote sensing images do not look like conventional images. Full-wave LIDAR images, for example, correspond to a complete profile describing the vertical structure of the object for each pixel. Radar images or SAR images are subject to particular statistics due to speckle, and can highlight particular physical effects that are invisible in optics (for example, humidity effects). Hyperspectral images contain spectral information and are therefore datacube images. Remote sensing time series are more and more easily accessible, when several acquisitions of the same scene are repeated over time.
- Finally, some functions are specific to this field. This, for example, is the case of the detection of changes between two images, to determine the differences between two acquisitions of the same scene, or the transfer of modalities: we try to move from one type of images to another when the information that they contain is related to the same underlying physical phenomena. Note that these functions are also important in other domains, such as the medical domain, where deep learning methods have been well developed.

ONERA conducts a large number of works in remote sensing imagery, on the one hand, and in artificial intelligence on the other. This dual competence makes it possible to ensure expertise in the use of deep learning for remote sensing applications by identifying the particular difficulties and also the opportunities.

The purpose of this article is twofold:

- First of all, to show, for a certain number of different cases studied at ONERA, how machine learning allows significant improvements in performance on functions based on the use of remote sensing images.
- Then to analyze, for each case, the major difficulties encountered in the implementation of these AI methods.

Compared with other articles summarizing the contributions of deep learning to remote sensing, this article is not intended to be exhaustive on the existing methods in the field, but rather to reveal recent and original results obtained specifically at ONERA, either in terms of methods, or in terms of the application scenario. Most of these methods have to face the challenges of the field of deep learning, such as the unsupervised or weakly-supervised paradigm, in order to prevent the need and the cost of annotation, and the issues arising from the lack of interpretability of such approaches and their perspectives with regard to the Earth observation domain.

The choice and implementation of network architectures depend on both the types of input data and the function to be implemented. Thus, in the second section, we first analyze what the state of the art is regarding the use of deep learning for the main functionalities of remote sensing. Then, the following sections are therefore organized around several fundamental studies, first by the type of function envisaged, concerning a type of specific data.

The applications are organized by hierarchical levels, from the lowest level to the highest level. Five types of functions are envisaged, for which the difficulty of the task and/or the abundance of images has been considered as a significant argument for the use of a learning method.

The successive sections of this article deal with co-registration of heterogeneous images, image quality enhancement (particularly SAR image filtering for radar images subject to speckle noise), land cover classification, vehicule detection, change detection, 3D sensing and estimation. Then concluding remarks are presented.

Related works

The creation of large-scale image databases, such as the pioneering ImageNet [29, 46] published by Stanford University, enabled an impressive shift in the way that image processing is considered. Neural network algorithms could then be applied to images. Indeed, they had made considerable progress in other fields where abundant training data was available. However, their implementation for images also relied on a recent technical advance: using Graphical Processing Units (GPUs) for general programming. Soon, deep learning [33, 38] brought a significant performance gap. Deep initially referred to the depth of the neural networks, which comprised many hidden layers. However, deep also means that the processing function is trained end-to-end, from data to expected result. From a learning point of view, this was considered much more satisfactory than previous processing pipelines designed by experts. In particular, the feature extraction was then trained, and yielded much better features than previous hand-crafted ones. In the following years, image processing underwent tremendous changes. Not only were tasks for which machine learning was already often considered then successfully addressed by deep networks, but traditional analytically-solved tasks became trainable.

Also, can we consider that remote sensing images are always like those of human vision? For traditional optical images, it is legitimate to think so. For other types of sensors, it is less obvious, because:

- Data are sometimes multi-modal,
- Data are geolocalized; they contain geographical maps rather than an object map,
- The time variable is becoming critical,
- In many cases, remote sensing is aimed at estimating geophysical parameters rather than detecting or classifying objects,
- Some images contain physical information that is different from visible information, such as SAR images or full-wave LIDARs.

Using deep learning for remote sensing came a little later than it did for computer vision; nevertheless, today, deep learning is widespread in the field and often establishes a new state of the art. Also, for each type of functionality, it is necessary to analyze what progress has been made in this area.

Up to now, the main remote sensing functions can be categorized as image processing, pixel-based classification or segmentation, target recognition, and scene understanding. A helpful review of state-of-the-art results of deep learning in remote sensing for several applications is given in [54], especially for hyperspectral image analysis for land cover/use classification and anomaly detection, SAR interpretation, high-resolution optical image interpretation, and fusion. In [53], the review focuses more specifically on classification techniques. To date, the auto-encoder (AE), the CNN, the Deep Belief Network (DBN), and Recurrent NN (RNNs) have been the four mainstream DL architectures used in the field of remote sensing. RNN is primarily used for analyzing non-stationary processes, CNN for classification tasks; and DBN or generative AEs for all other tasks, in particular unsupervised ones.

Both states of the art in [54] and [53] confirm that, as in other fields, deep learning is making remote sensing advance, even though progress is recent and further improvements can be expected. The general feeling is that, in upcoming years, we can still expect great advances in remote sensing thanks to deep learning. It also has limitations and raises new challenges: the lack of annotated data, the difficulty in deploying and transferring models under various conditions at global scales, and also the taking into account of sensor physics and purely algorithmic tasks.

Since our first deep learning works [34, 2], we have made progress on several tasks at ONERA, with research on standard tasks, such as classification or object detection, as well as on topics that have been addressed very rarely, such as co-registration, or 3D estimation. We detail below our recent advances in this area over the last 5 years.

Co-registration

Co-registration of heterogeneous images is useful in various remote sensing image fusion applications, since a gain is expected from the synergy of sensors. Relevant applications are numerous, whether for land classification, for agriculture, or forestry applications. Some applications require a pixel precision and, generally, the terrain correction applied for georeferencing is not good enough. The residual bias arises from the impact of the imprecision on orbital or DTM (Digital Terrain Model) parameters during the mapping. The influence of the relief on such registration is non-rigid and, therefore, requires the estimation of a dense motion field.

In remote sensing, deep learning methods for co-registration are not numerous. [41] is a feature-based approach that proposes an architecture derived from a Siamese Neural Network (SNN) trained to select precise and reliable points of correspondence between the two images.

Recently at ONERA, we proposed the investigation of image-based deep learning approaches taking into account all of the pixels of an image. We then proposed the adaptation of PWC-Net, a CNN already developed in Computer Vision, in order to make it efficient for heterogeneous images, such as a couple of SAR/optical images. All of the performances of our tests were compared with a reference algorithm for optical flow developed at ONERA, GeFolki.

A first significant challenge was to constitute the training base. For this purpose, we used the Google Earth Engine (GEE) data platform, able to handle both optical Sentinel-2 (S2) and radar Sentinel-1 (S1) images. We selected georeferenced images assumed to be well co-registered together. Using the platform, we were able to define large footprints common to S1 and S2. For S2 images, we choose dates with the weakest cloud cover. Around the corresponding acquisition dates, the S1 radar images have been filtered temporally to reduce the effects of speckle and increase their signal-to-noise ratio. A systematic coverage of the entire French territory has been established; this is to ensure a representative diversity of all of the landscapes encountered, such as agricultural, city, forest, and mountain areas. Then, we also artificially generated, for each pair, realistic deformations whose amplitude is modulated spatially by the relief, given by the SRTM product downloaded on the same footprint.

We have considered FlowNetS [31] and PWC-Net [48], two Convolutional Neural Networks (CNN) commonly used for optical flow estimation in computer vision, PWC-Net being state of the art.

FlowNetS can take into account any type of input image. Without modifying its architecture, it could be applied successfully to our images. It proved robust for estimating flows. On the contrary, PWC-Net is conceptually defined for two entries of the same type, in particular with a shared encoding between the entries through the application of the same encoder. Such encoding makes no sense for different images. For this reason, we propose a modification of the architecture with two independent encoders. We separate the Siamese contracting parts of PWC-Net into two different contracting parts performing two different operations. Those two contracting parts share the same architecture but can have different weights. Thus, there is a contracting part specialized for SAR images and a second one specialized for optical images. Moreover, we decided to remove the lowest resolution stage of the architecture for both the contracting and the expending parts, since it can estimate mainly large deformations. Finally, we propose to assist the training by simultaneously using three different loss functions exploiting the different combinations of contracting parts that we can use: optics/optics, radar/optics, and radar/radar. We call our new architecture PWC-Net-multimodal.

Results have been tested on a new database, and compared with that of the GeFolki algorithm [13], which is an optical flow method without machine learning. Deep learning architecture performed better, not only on data close to the training data set, but also on data acquired with different sensors, with significantly higher resolution. The expected prediction error lies between 0.7 and 1.1 pixels for the different deep learning architectures and for different scenarios, whereas it lies between 2.3 and 3.4 for GeFolki.

Concerning FlowNet, PWCNet, and PWC-Net-multimodal, we have shown that the results obtained using the three methods are close, with a better result having been achieved with our PWC-Net-multimodal method. Furthermore, the PWC-Net-multimodal method is more robust to train with excellent repeatability, while the original PWC-Net does not converge every time we try to train it with heterogeneous data.

One example of the result obtained by PWC-Net-multimodal is given in Figure 1. The first column in Figure 1 shows a mosaic of the optical and radar images before and after registration. The optical-radar junctions of the mosaic highlight some structures that are shifted before registration, and that our algorithm manages to align well. The middle column represents the norm of the flow and its direction in color. The top image gives the ground truth, and the bottom image gives estimated results. We see that the estimated flow has the right direction but still lacks spatial details. However, the absolute flow errors remain less than 2 pixels, and the relative error remains below 20%.

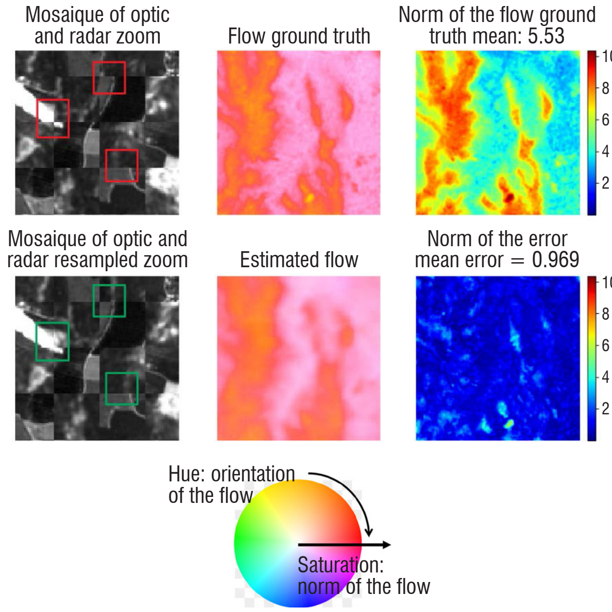


Figure 1 – Mosaic of Sentinel-1 and Sentinel-2 images of the same zone (First column). The ground truth flow and the flow estimated with PWC-Net trained with our dataset (middle column). The quadratic norm of the ground truth flow and the norm of the error of the estimation (last column).

Perspectives

Given that databases with different modalities such as SAR/optical images are still being widely developed, progress is likely to continue thanks to deep learning. Methods should take advantage of the specificity of the different modalities. For example, the different polarizations for radar images can be used together to improve efficiency. We also plan to test and improve robustness to the possible presence of clouds or changes. Finally, we have to tackle the co-registration problems for high-resolution images by considering a new formulation taking into account the 3D aspects, since the resolved 3D elements such as buildings can have different projections, without any bijective relationship between them.

Image quality enhancement

The notion of noise filtering is particularly crucial for radar images because these images have an inherent speckle noise. Many algorithms strive to remedy this noise through the speckle filtering operation. Up until now, all of the algorithms exploited spatial information. Now, as time-series become available, temporal information can also be used.

Speckle filtering methods usually fall into two categories: noise modeling and data-based approaches. The later includes machine learning methods. The amount of available data and the difficulty in modeling generic de-noising methods make the use of deep learning an already efficient solution [52, 50]. However, most proposed solutions rely on supervised methods, and thus on the description of ground truth, which is, in this case, the achieved goal at the filtering output. An essential difficulty is knowing how to describe what is meant by ground truth in this case. Obtaining training datasets by using simulation is one of the possible remedies, but transfer to real SAR data remains a significant challenge.

The originality of the work undertaken at ONERA in this area is the use of time series to avoid having to provide an objective [12]. We propose to use the redundancy of the data in the stack, and to formulate the problem as follows: given two realizations I_1 and I_2 of the same scene, let us learn to predict I_2 from I_1. The transfer function itself performs the filtering of the random part of the signal, keeping only the deterministic part.

We have tested several networks and several loss functions, and we have also compared our results with other spatiotemporal filtering methods, such as BM3D [26] and SAR-BM3D [44]. The best results are achieved using dilated convolutional networks and histogram loss, which is defined by a distance ℓ_2 on the histogram vector of a given pixel $x \in X$ in its neighborhood N_x. Then, the gradients for back-propagation are the differentiation of the previous distance. To scale up to large images, we do not feed the network with the whole image but rather with patches, with possible overlapping to prevent border effects.

The learning phase focused on an SLC Sentinel-1 image stack around Saclay, 20 km south of Paris. Figure 2 presents the results for Valencia, a scene that is not part of the auto-encoding set (red and blue channels for VV and green channel for VH). The function has removed most of the noise, e.g., around the harbor.

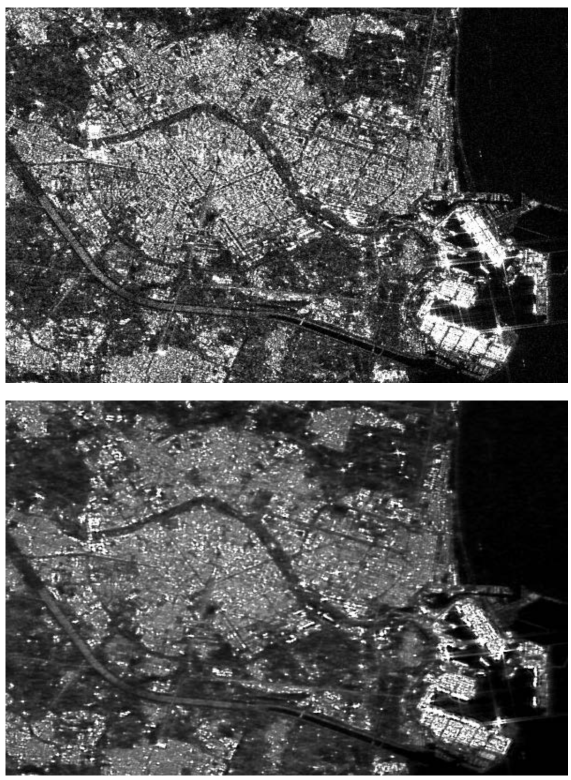
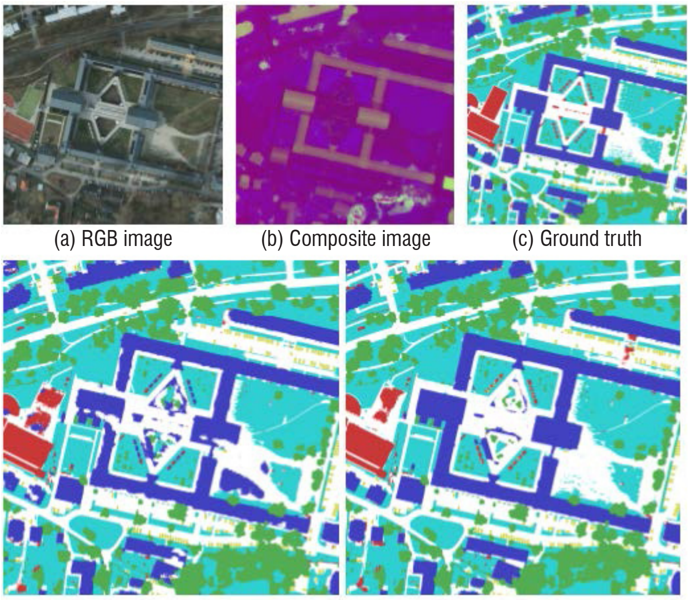


Figure 2 – Filtering results over images of Valencia with a network trained on Paris. Top: original image, bottom filtered image.



(a) RGB image (b) Composite image (c) Ground truth

(d) SegNet prediction (d) V-FuseNet prediction

Figure 3 – Semantic segmentation

Although the results are not perfect, the ability to generalize has been demonstrated, since the network has been successfully applied, having been learned on a very different site and for different resolutions.

Perspectives

The next steps are to manage the way in which knowledge of some of the metadata is included, such as polarization, incidence angle, and also the possibility of mixing this unsupervised approach with supervised ones, through the use of image databases for which speckle noise is added artificially.

Land cover classification

Classification tasks in remote sensing benefited the most from the deep learning trend [53, 54, 10]. Although image classification triggered an interest in deep neural networks in computer vision, remote sensing tasks have their peculiarities. They include a focus on pixel classification or semantic segmentation (partly tied to the fact that remote sensing images are very large, while being only small portions of the Earth Surface), the variety of imaging techniques (RGB, multispectral, LIDAR, SAR, etc.), and the variety of potential ground-truths, which lies in the existing maps of all sorts.

The abundance of remote sensing data enables us to extract relevant information through deep learning, for the automatic **semantic mapping** of the Earth from multimodal, aerial and satellite data, in urban or rural environments [40, 16, 3]. Primarily, we aim to automatically map the land cover and land use of large-scale scenes using all available data. Therefore, we have proposed new neural networks[1] to deal with highly heterogeneous multimodal data, such as LIDAR and optical acquisitions [8]. Thanks to a double-flow architecture and to the introduction of a new neuronal block called residual correction [4, 5], our model has improved upon the state of the art achieved with the ISPRS Vaihingen and Potsdam datasets [45] for various classes, such as roads, buildings, vegetation and vehicles, and with another

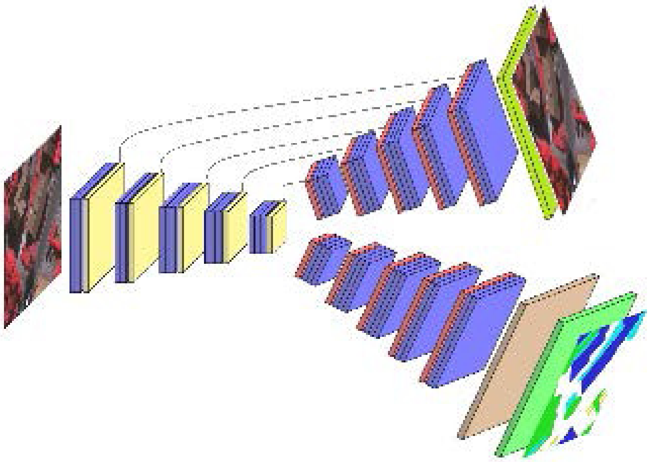


Figure 4 – BerundaNet, a multi-task neural network for semi-supervised semantic segmentation

benchmark dataset for buildings all over the world [32]. An example of a segmentation result is shown in Fig. 3 on the Postdam dataset.

In the field of **hyperspectral images** (HS), we explored models with various kinds of convolutions more suited to the specific data structure. Indeed, we consider HS images as data cubes rather than rasters of standard images. We have shown that 3D convolutions are more suited to HS data if enough da ta are available [10]. We then released DeepHyperX[2], an open-source toolbox that enables the scientific community to investigate how deep learning tools can participate in particular HS imaging classification problems. Finally, data scarcity is a common issue in vegetation or mineral studies. To enrich the databases used for training algorithms, we have proposed a method for the synthesis of realistic spectra based on generative adversarial models [9].

Geospatial data include huge volumes of ortho-rectified images and maps of different kinds (geographic but also political or theme-specific). It was necessary to find ways to leverage them for training. For instance, we proposed the inclusion of prior knowledge from OpenStreetMap in the learning process, thus showing the ability of neural networks to take advantage of heterogeneous sources of information [6]. We also proposed to encode the spatial shapes and relationships between classes through Distance Transform Regression [1]. The production of thematic maps often comes up against the high-level semantics of the expected classes. For example, in order to find solutions to characterize urban heat islands from the sky, we organized a benchmark for Local Climate Zone (LCZ, [47]) classification, spanning various cities around the world. It showed that although deep networks are efficient for quickly producing averagely-good maps, adding expert knowledge with more standard approaches like boosting or random forests were highly valuable to obtain quality maps [51].

Perspectives

We now seek to benefit from the unexploited, unlabeled data. To this end, we investigate semi-supervised architectures, such as that in Fig. 4, able to learn image characteristics from the unlabeled images available for every location to regularize land-use and land-cover classification (urban fabric, wetlands, forests, fields) [19, 20]. A related topic of interest is weakly-supervised learning, to learn with unreliable ground-truth or classes that are not visually homogeneous [23, 21].

[1] https://github.com/nshaud/DeepNetsForEO

[2] https://github.com/nshaud/DeepHyperX

Vehicle detection

Object detection in an image consists in locating all instances of the object of interest in an image. In this task, the input of the algorithm is an image, and the output is a set of locations. The quality of the algorithm output is measured by comparing the produced set of locations with the ground truth (the known set of locations of the objects). Although this problem seems very understandable, there are many ways to quantify the results, depending on the nature of the requested location. Classically, detection is aimed at placing bounding boxes on vehicles, predicting both location and scale, and ensuring one-to-one matching. This task allows, for example, vehicles to be counted.

In particular experimental settings, typically in a low-resolution context or for hard-to-understand sensors, ground truth can be obtained by monitoring vehicles on the ground, and operators and algorithms are complementary. However, most of the time, humans excel at locating vehicles in images when the resolution is higher than 20 cm per pixel. Indeed, ground truth is usually obtained just by manual inspection of the image. Also, because humans are outstanding at this task, there is no qualitative advantage of using algorithms instead of concentrated operators.

The advantage of a detection algorithm thus lies either in the ability to automate the detection of vehicles in large numbers, or to improve the performance of the detection in rare modalities among all available remote sensing data. This task could bring a real technological breakthrough, notably allowing a better town planning policy and, of course, providing valuable information for intelligence.

Deep learning is especially relevant for both accuracy and scalability. First, the performance achieved by deep learning in vehicle detection is at least as high as the performance achieved by other kinds of algorithms, such as those described in [36, 34]. Typically, given a minimal set of images, designing an *ad hoc* detector to detect vehicles for this specific context is often possible.

However, deep learning is generic and incremental: it is increasingly accurate when fed with more training data. Then, training is just linear in relation to the size of the training database, and both training and testing are very fast on hardware like GPU cards designed for deep learning. In addition, we can use a shared deep learning pipeline for multiple purposes: typically [7] offers a way to achieve detection as post-processing of land cover classification.

To reach its own opinion of the results obtained, ONERA implemented various different architectures on different datasets. In particular, ONERA has developed a manually annotated database of 20,000 vehicles on 20 cm resolution aerial images from the ORTHO HR ® produced by the IGN (National Institute of Geographic and Forest Information) in partnership with local authorities.

The first results of car detection were obtained using the approach described in [7], on these images. Figure 5 illustrates one example of detection results. We have conducted other works on Pléiades datasets at 50 cm resolution, or for aircraft type targets. This way, on highly-resolved images better than 10 cm per pixel and on large datasets, deep learning overrules the state of the art [7, 42] providing performances as high as 94% of F-score for [42] – 86% for [7]. However, today, deep learning is not sufficiently accurate to keep its promises with regard to classical remote sensing images (less than 20 cm per pixel) and, besides, suffers from a lack of large, structured, annotated and free datasets at this resolution.

Perspectives

More than on network architectures, it is on the development of better-constructed large databases that efforts are expected. The simulation of various scenes, including diverse targets under diverse lighting conditions, could also play a role in this context and help to reach that operational quality soon.

Change detection

Change detection is aimed at finding the changes between two co-registered images taken at different times [35]. It is often tackled at the pixel level by semantic segmentation approaches. It is an example of a dense classification problem, where we attempt to assign a label to each pair or sequence of corresponding pixels. Depending on the desired application, the assigned labels may be binary, change or no change, or they may contain semantic information about the changes that have happened, such as deforestation, urban expansion, or water loss.

At ONERA, we have recently achieved state-of-the-art results in change detection using state-of-the-art machine learning techniques. For this purpose, a dataset has been developed to train and benchmark various different change detection algorithms for change detection [22]: the ONERA Sentinel Change Detection dataset (OSCD). It contains several multispectral image pairs extracted from Sentinel-2 acquisitions and manually created binary change labels for all pixels in all image pairs. OSCD has also been released publicly[3] so that scientists all over the world can accurately compare their proposed algorithms quantitatively and together

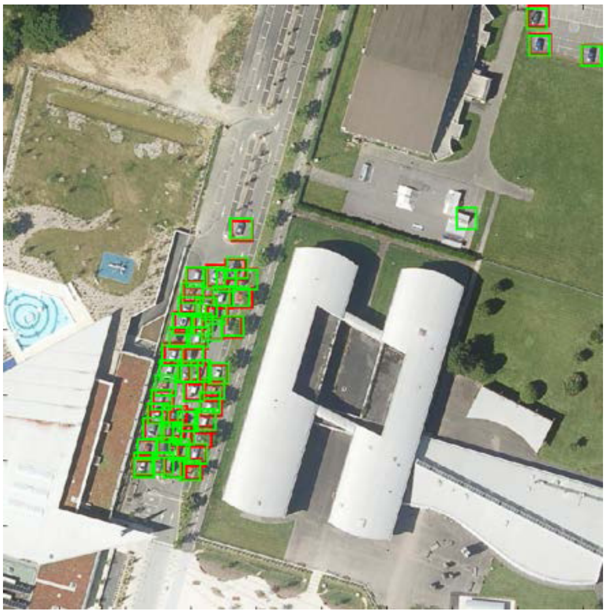
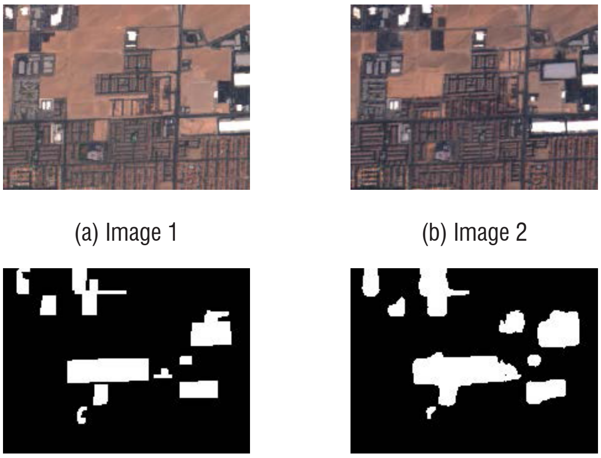


Figure 5 – Detection results obtained using the approach described in [7] implemented through a UNet, applied over aerial photography, from IGN ORTHO HR ®, at 20 cm resolution. Red boxes: detection, Green boxes: Ground Truth.

3 http://dase.grss-ieee.org/



(a) Image 1 (b) Image 2

(c) Change map (c) Estimated change map

Figure 6 – Example of a satellite image pair of Las Vegas, true change map and estimated change map using a convolutional neural network

develop ever-improving change detection methods. Example results are given in Figure 6.

The change detection methods proposed in [27] surpass related methods in both accuracy and speed. They are extensions of fully convolutional encoder-decoder networks using skip connections and, most notably, a *Siamese* extension of a traditional encoder-decoder architecture using heuristics specific to the problem of change detection achieved the best results.

More recent works have pushed the boundaries of state-of-the-art change detection methods even further. A new dataset, called High-Resolution Semantic Change Detection (HRSCD) dataset has been generated, and will also be made available to the scientific community. This dataset is more than 3000 times larger than any other change dataset openly available, and it contains high-resolution (50 cm per pixel) images. It contains not only change labels for all pixels, but also land cover information.

This dataset enables research in change detection to go much further. First, multitask learning was carried out using the HRSCD dataset, and we show that simultaneously learning to detect changes and to classify the terrain in the images led to better network performances [28]. Wea kly supervised learning techniques were also proposed to deal with the label noise inherent to automatically generated data [23]. Using a combination of iterative learning, classification filtering, and the newly proposed guided anisotropic diffusion post-processing method, an encoder-decoder fully convolutional neural network was trained to obtain excellent change detection performances.

Recently, given the availability of time-series, we have extended change detection to **activity detection** over a given period. The volume of available data for this specific problem is much smaller than for other remote sensing tasks. Indeed, in the case of optical images where clouds can be an issue, gathering a sequence of cloud-free images is much harder than finding a single image, especially in humid regions. The volume of labeled data for activity detection is also scarcer than those available for other problems, which limits the complexity of the machine learning models that we can use, since more complex models need large amounts of data to avoid overfitting. Since Sentinel-1 (S1) radar data do not suffer from cloud cover, they allow for more accessible collection and processing of stacks of images. We have developed the REACTIV algorithm on this basis: by exploiting the particular statistical properties of the radar images, the algorithm allows us to obtain unmatched change detection performance, superior to that obtained by the previous supervised deep learning approach applied to the same scenario, but on optical Sentinel-2 images [25]. We have evaluated these performances on a set of data from the Saclay plateau, chosen for a large number of construction sites present during the analysis period.

Perspectives

On the one hand, we have developed a very robust change detection algorithm with the exploitation of SAR time-series, but the interpretation of the change remains difficult using this data. On the other hand, thanks to the optical-imagery datasets that we released for the community, deep networks for change detection have emerged [27, 43]. However, a more massive dataset could improve results still further. Therefore, our next efforts will be devoted first to improving database creation by using robust automatic detection on radar images to select change key positions in optical images. Then, we will investigate more sophisticated and accurate deep learning methods for semantic change detection and high-resolution change detection based on time-series.

3D Sensing and Estimation

Obtaining 3D data through imaging is an area where the traditional deep learning used for image processing does not apply directly. However, we have invested in this area by using our expertise in the conception and processing of data from advanced sensors. Two areas of research concern machine learning for 3D: LIDAR and SAR data fusion for 3D forest structure studies; and 3D model estimation from the sky.

The goal of [14] is to predict the structural parameters of forests on a large scale using remote sensing images. LIDAR and polarimetric interferometric SAR sensors are both interesting for estimating forest biomass. However, if LIDARs offer excellent vertical accuracy, they suffer from their lack of spatial coverage. On the other hand, SAR systems have extensive coverage and ground spatial accuracy, but reduced vertical precision. Therefore, the approach is to extend the accuracy of LIDAR full waveforms to a larger area covered by polarimetric and interferometric (PolInSAR) synthetic aperture radar images using machine learning methods.

We proposed in [14] a set of PolInSAR parameters, computed for each pixel, which is likely to have strong correlations with the LIDAR density profiles on forest stands. These features were used as input data to learn a set of forest LIDAR features: the canopy height, the vertical profile type, and the canopy cover. The approach has investigated several methods of machine learning for this purpose:

- Random forest methods for class classification of vegetation profiles.
- Classical SVN methods, then perceptron methods for estimating canopy height, and canopy density.
- Finally, CNN methods were also tested for estimating canopy height.

In the latter case, the results were not better than those obtained with other traditional machine learning methods. The small number of data in the learning base can explain this. Also, the chosen input was not a simple vector, but was transformed into a two-dimensional space and interpreted artificially as an image, even though it does not correspond to real objects but rather to pure mathematical representations used to feed CNN networks.

Nevertheless, perceptrons give very encouraging results. Neural networks give the best performances in terms of RMSE on the estimated tree heights, as represented in Fig. 7, and they are the most faithful on the fidelity of the estimated statistical distributions, as shown in Fig. 8.

This work has also demonstrated the importance of the choice of input descriptors: the performances are better with descriptors judged to have the most physical meaning. This work makes us think that deep learning allows increased performance, but not necessarily based on images, even if we acquired the data in that way.

A second axis consists of scene understanding from standard optical imagery. Additionally to semantics (see Sections 5 and 6), providing the local height, for example, as with Digital Surface Models (DSMs), is useful for many applications, such as urban planning, telecommunications, aviation, and intelligent transport systems. Multi-view stereo [30] was the means of choice to obtain these products, but today deep learning approaches also offer competitive performances [37]. We address this problem by using a Multi-Task Learning (MTL) deep network that estimates both height and semantic maps simultaneously from a single aerial image [18]. Our approach is built on powerful models that we developed previously for depth prediction from a single image taken from the ground [17].

Precisely, we adapted D3-Net [17] to a multi-task architecture by adding a semantic classification decoder to the original depth estimation one. As shown in Figure 9, the contractive new decoder layers are common to both semantics and height estimation. On the contrary, layers of the decoders are specific for each objective and generate, respectively, as many channels as classes for semantics and one channel for height. We have evaluated each output with a corresponding loss function: we have adopted the absolute error ($L1$) for height regression and the cross-entropy loss (LCE) for semantics evaluation. We have also implemented various mechanisms for multi-task optimization, such as Grad-Norm [24].

Figure 10 shows maps generated from the 2018 Data Fusion Contest (DFC2018 [49]) dataset. In general, the network produces nearly

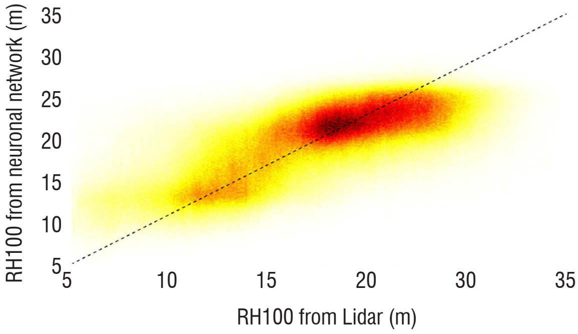


Figure 7 – Tree height estimation precision using neural network

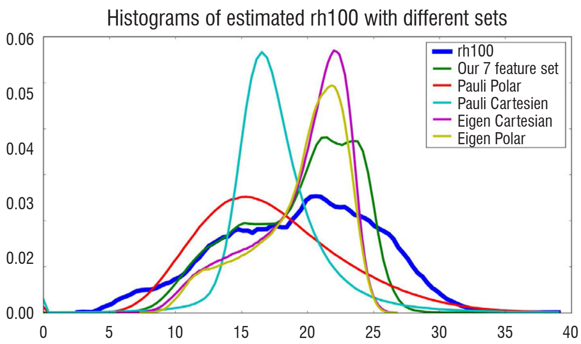


Figure 8 – Tree height estimation using different machine learning algorithms

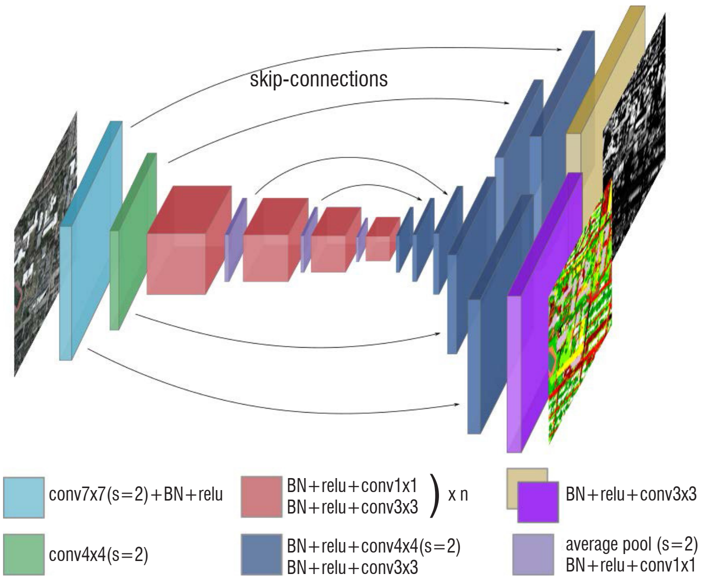


Figure 9 – Architecture of our MTL model for height regression and semantic classification. On the left most layers share parameters between all tasks and on the right most layers are task-specific.

Figure 10 – From left to right, input RGB image, semantic ground-truth and prediction. Black represents no information. Height ground-truth and prediction evaluated for DFC2018 data over Houston.

accurate heights for ground, residential buildings, and vegetation, while some structures are more challenging, like high buildings or stadiums. Indeed, these classes in bird-view images have various shapes, colors, and heights, which make precise estimation difficult. We can note that semantics are detailed, with dense cartography even when the ground truth labeled only a few objects. Following Section 5, we have shown that multi-task learning allows performance to be improved for both tasks, or in other words, that simultaneous height estimation helps classification considerably.

Perspectives

Our works on the prediction of biomass will continue through the scaling-up of this kind of algorithm by using future satellite missions, such as BIOMASS, Tandem-L, or Ni-SAR. Height-estimation deep networks are only one example of a network able to translate one modality into another: predictors of LIDAR-like point-clouds or SAR-processing networks are also on our agenda. To this end, the understanding of the physics behind the sensor benefits from every bit of available information to build a better 3D estimate. More generally, the estimation of the 3D structure with precise levels of detail using a variety of sensors is a crucial step for creating models of the world that enable environmental sustainability or development of smart cities.

Conclusion

This paper provided an overview of the performance gains obtained today in remote sensing through the use of deep learning techniques. It has demonstrated significant gains in the areas closest to those of computer vision: classification and detection of vehicles in optical images. They are probably the most impressive results because more effort has been made, and also because the transfer of techniques from computer vision to remote sensing is more comfortable to do.

For other more exotic applications, such as multitemporal change detection, LIDAR to radar data transfer for biomass estimation, and 3D estimation, deep learning methods are being developed. The first results are encouraging and prove the feasibility of the methods. Thus, today, it seems that the contribution of deep learning no longer needs to be demonstrated. The expected gains in the future relate mainly both to the development of new architectures or new learning schemes and to the way of forming the learning base.

Hence, future expectations will focus on topics that are absent today from computer vision: for example, the taking into account of complex signals such as SAR images, multimodal data, sparse data such as hyperspectral images. Moreover, taking into account the physics knowledge or the part of the signal that is useful for the intended application seems to be essential and seems essential to the quality of the result.

The other issues concern the organization and constitution of databases. This point is a real difficulty. To obtain annotated databases, in the case of the web, we can count on the annotations of millions of users. In remote sensing, this is not the case. Not to mention the difficulties due to data that are confidential and cannot necessarily benefit from these public platforms. The number of data is not necessarily the most limiting factor: in fact, the images have large dimensions of several tens of thousands of pixels, which makes it possible to process them in a large number of smaller images. Also, a promising avenue is to use approaches that mix supervised learning with unsupervised learning.

However, in any case, it is be necessary to rely on the development of dedicated platforms, and their capacity to interface with the data processing cloud type deep learning, as well as to ingest data from various sources of geographic information, to benefit from all of the advances of this sector of machine learning

Acknowledgements

The authors thank the ISPRS for making the Vaihingen and Potsdam datasets available and for organizing the semantic labeling challenge. The authors would also like to thank the National Center for Airborne Laser Mapping and the Hyperspectral Image Analysis Laboratory at the University of Houston for acquiring and providing the DFC2018 data used in this study, and the IEEE GRSS Image Analysis and Data Fusion Technical Committee.

References

[1] N. AUDEBERT, A. BOULCH, B. LE SAUX, S. LEFÈVRE - *Distance Transform Regression for Spatially-Aware Deep Semantic Segmentation*. Computer Vision and Image Understanding, 189:102809, 2019.

[2] N. AUDEBERT, A. BOULCH, H. RANDRIANARIVO, B. LE SAUX, S. LEFÈVRE, R. MARLET - *Deep Learning for Urban Remote Sensing.* Joint Urban Remote Sensing Event (JURSE). IEEE, March 2017.

[3] N. AUDEBERT, B. LE SAUX, S. LEFÈVRE - *How Useful is Region-Based Classification of Remote Sensing Images in a Deep Learning Framework?* International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2016.

[4] N. AUDEBERT, B. LE SAUX, S. LEFÈVRE - *Semantic Segmentation of Earth Observation Data Using Multimodal and Multi-Scale Deep Networks.* Asian Conference on Computer Vision (ACCV). IEEE, November 2016.

[5] N. AUDEBERT, B. LE SAUX, S. LEFÈVRE - *Fusion of Heterogeneous Data in Convolutional Networks for Urban Semantic Labeling.* Joint Urban Remote Sensing Event (JURSE). IEEE, March 2017.

[6] N. AUDEBERT, B. LE SAUX, S. LEFÈVRE - *Joint Learning from Earth Observation and Openstreetmap Data to get Faster Better Semantic Maps.* The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops / EarthVision, July 2017.

[7] N. AUDEBERT, B. LE SAUX, S. LEFÈVRE - *Segment-Before-Detect: Vehicle Detection and Classification through Semantic Segmentation of Aerial Images.* Remote Sensing, 9(4):368, 2017.

[8] N. AUDEBERT, B. LE SAUX, S. LEFÈVRE - *Beyond RGB: Very High Resolution Urban Remote Sensing with Multimodal Deep Networks.* ISPRS Journal of Photogrammetry and Remote Sensing, 140:20 – 32, 2018.

[9] N. AUDEBERT, B. LE SAUX, S. LEFÈVRE - Generative Adversarial Networks for Realistic Synthesis of Hyper-Spectral Samples. International Geo- science and Remote Sensing Symposium (IGARSS). IEEE, 2018.

[10] N. AUDEBERT, B. LE SAUX, S. LEFÈVRE - Deep Learning for Classification of Hyperspectral Data: A Comparative Review. IEEE Geoscience Remote Sensing Magazine, 7(1), 2019.

[11] W. BLACKWELL, G. ALLEN, C. GALBRAITH, T. HANCOCK, R. LESLIE, I. OSARETIN, L. RETHERFORD, M. SCARITO, C. SEMISCH, M. SHIELDS, *et al* - *Nanosatellites for Earth Environmental Monitoring: the MicroMAS Project.* Microwave Radiometry and Remote Sens- ing of the Environment (MicroRad), pages 1-4. IEEE, 2012.

[12] A. BOULCH, P. TROUVÉ, E. COLIN-KOENIGUER, F. JANEZ, B. LE SAUX - *Learning Speckle Suppression in SAR Images without Ground Truth: Application to Sentinel-1 Time-Series.* International Geoscience and Remote Sensing Symposium (IGARSS), pages 2366-2369. IEEE, 2018.

[13] G. BRIGOT, E. COLIN-KOENIGUER, A. PLYER, F. JANEZ - *Adaptation and Evaluation of an Optical Flow Method Applied to Coregistration of Forest Remote Sensing Images.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(7):2923-2939, 2016.

[14] G. BRIGOT, M. SIMARD, E. COLIN-KOENIGUER, A. BOULCH - *Retrieval of Forest Vertical Structure from Polinsar Data by Machine Learning Using Lidar-Derived Features.* Remote Sensing, 11(4):381, 2019.

[15] E. BUCHEN, D. DE-PASQUALE - *2014 Nano/Microsatellite Market Assessment.* SpaceWorks Enterprises, 12, 2014.

[16] M. CAMPOS-TABERNER, A. ROMERO-SORIANO, C. GATTA, G. CAMPS-VALLS, A. LAGRANGE, B. LE SAUX, A. BEAUPÈRE, A. BOULCH, A. CHAN-HON-TONG, S. HERBIN, H. RANDRIANARIVO, M. FERECATU, M. SHIMONI, G. MOSER, D. TUIA - *Processing of Extremely High-Resolution LIDAR and RGB Sata: Outcome of the 2015 IEEE GRSS Data Fusion Contest – Part a: 2-D Contest.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(12):5547–5559, 2016.

[17] M. CARVALHO, B. LE SAUX, P. TROUVÉ-PELOUX, A. ALMANSA, F. CHAMPAGNAT - *On Regression Losses for Deep Depth Estimation*. International Conference on Image Process- ing (ICIP), pages 2915-2919. IEEE, 2018.

[18] M. CARVALHO, B. LE SAUX, P. TROUVÉ-PELOUX, A. ALMANSA, F. CHAMPAGNAT - *Multi-Task Learning of Height and Semantics from Aerial Images.* IEEE Geosci. and Remote Sensing Letters, 2019.

[19] J. CASTILLO, B. LE SAUX, A. BOULCH, S. LEFÈVRE - *Réseaux de neurones semi-supervisés pour la segmentation sémantique en télédétection.* Colloque GRETSI, 2019.

[20] J. CASTILLO-NAVARRO, N. AUDEBERT, A. BOULCH, B. LE SAUX, S. LEFÈVRE - *What Data are Needed for Deep Learning in Earth Observation?* Joint Urban Remote Sensing Event (JURSE). IEEE, May 2019.

[21] R. CAYE DAUDT, A. CHAN-HON-TONG, B. LE SAUX, A. BOULCH - *Learning to Understand Earth-Observation Images with Weak and Unreliable Groundtruth.* International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2019.

[22] R. CAYE DAUDT, B. LE SAUX, A. BOULCH, Y. GOUSSEAU - *Urban Change Detection for Multispectral Earth Observation using Convolutional Neural Networks.* International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2018.

[23] R. CAYE DAUDT, B. LE SAUX, A. BOULCH, Y. GOUSSEAU - *Guided Anisotropic Diffusion and Iterative Learning for Weakly Supervised Change Detection.* The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops / EarthVision, June 2019.

[24] Z. CHEN, V. BADRINARAYANAN, C.-Y. LEE, A. RABINOVICH - *Gradnorm: Gradient Normalization for Adaptive Loss Balancing in Deep Multitask Networks.* International Conference on Machine Learning (ICML), 2018.

[25] E. COLIN KOENIGUER, J.-M. NICOLAS - *Change Detection in SAR Time-Series Based on the Coefficient of Variation*. arXiv preprint arXiv:1904.11335, 2019.

[26] K. DABOV, A. FOI, V. KATKOVNIK, K. EGIAZARIAN - *Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering*. IEEE Transactions on Image Processing, 16(8):2080-2095, 2007.

[27] R. CAYE DAUDT, B. LE SAUX, A. BOULCH - *Fully Convolutional Siamese Networks for Change Detection*. International Conference on Image Processing (ICIP). IEEE, 2018.

[28] R. CAYE DAUDT, B. LE SAUX, A. BOULCH, Y. GOUSSEAU - *Multitask Learning for Large-Scale Semantic Change Detection*. Computer Vision and Image Understanding, 187:102783, 2019.

[29] J. DENG, W. DONG, R. SOCHER, L.-J. LI, K. LI, L. FEI-FEI - *ImageNet: A Large-Scale Hierarchical Image Database.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 248-255. IEEE, 2009.

[30] G. FACCIOLO, C. DE FRANCHIS, E. MEINHARDT-LLOPIS - *Automatic 3D Reconstruction from Multi-Date Satellite Images.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops / EarthVision, 2015.

[31] P. FISCHER, A. DOSOVITSKIY, E. ILG, P. HÄUSSER, C. HAZIRBAŞ, V. GOLKOV, P. VAN DER SMAGT, D. CREMERS, T. BROX - *FlowNet: Learning Optical Flow with Convolutional Networks*. International Conference on Computer Vision (ICCV). IEEE, December 2015.

[32] B. HUANG, K. LU, N. AUDEBERT, A. KHALEL, Y. TARABALKA, J. MALOF, A. BOULCH, B. LE SAUX, L. COLLINS, K. BRADBURY, *et al.* - *Large-Scale Semantic Classification: Outcome of the First Year of Inria Aerial Image Labeling Benchmark.* International Geoscience and Remote Sensing Symposium (IGARSS), pages 6947-6950. IEEE, 2018.

[33] A. KRIZHEVSKY, I. SUTSKEVER, G. E. HINTON - *Imagenet Classification with Deep Convolutional Neural Networks.* Advances in Neural Information Processing Systems, pages 1097-1105, December 2012.

[34] A. LAGRANGE, B. LE SAUX, A. BEAUPERE, A. BOULCH, A. CHAN-HON-TONG, S. HERBIN, H. RANDRIANARIVO, M. FERECATU - *Benchmarking Classification of Earth-Observation Data: from Learning Explicit Features to Convolutional Networks.* International Geoscience and Remote Sensing Symposium (IGARSS), pages 4173-4176. IEEE, July 2015.

[35] B. LE SAUX, H. RANDRIANARIVO - *Urban Change Detection in SAR Images by Interactive Learning.* International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, July 2013.

[36] B. LE SAUX, M. SANFOURCHE - *Robust Vehicle Categorization from Aerial Images by 3D-Template Matching and Multiple Classifier System.* IEEE Int. Symposium on Image and Signal Processing and Analysis (ISPA), September 2011.

[37] B. LE SAUX, N. YOKOYA, R. HANSCH, M. BROWN, G. HAGER - *2019 Data Fusion Contest [Technical Committees]*. IEEE Geoscience and Remote Sensing Magazine, 7(1):103-105, 2019.

[38] Y. LECUN, Y. BENGIO, G. HINTON - Deep Learning. Nature, 521(7553):436-444, May 2015.

[39] S. LI, S. DRAGICEVIC, F. A. CASTRO, M. SESTER, S. WINTER, A. COLTEKIN, C. PETTIT, B. JIANG, J. HAWORTH, A. STEIN, *et al* - *Geospatial Big Data Handling Theory and Methods: a Review and Research Challenges.* ISPRS Journal of Photogrammetry and Remote Sensing, 115:119-133, 2016.

[40] D. MARMANIS, J. D. WEGNER, S. GALLIANI, K. SCHINDLER, M. DATCU, U. STILLA - *Semantic Segmentation of Aerial Images with an Ensemble of CNNs.* ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 3:473, 2016.

[41] N. MERKLE, W. LUO, S. AUER, R. MÜLLER, R. URTASUN - *Exploiting Deep Matching and SAR Data for the Geo-Localization Accuracy Improvement of Optical Satellite Images.* Remote Sensing, 9(6):586, 2017.

[42] T. N. MUNDHENK, G. KONJEVOD, W. A. SAKLA, K. BOAKYE - *A3 Large Contextual Dataset for Classification, Detection and Counting of Cars With Deep Learning.* European Conference on Computer Vision (ECCV), pages 785-800. IEEE/CVF, 2016.

[43] M. PAPADOMANOLAKI, S. VERMA, M. VAKALOPOULOU, S. GUPTA, K. KARANTZALOS - *Detecting Urban Changes with Recurrent Neural Networks from Multitemporal Sentinel-2 Data.* International Geoscience and Remote Sensing Symposium (IGARSS), pages 214-217. IEEE, 2019.

[44] S. PARRILLI, M. PODERICO, C. V. ANGELINO, L. VERDOLIVA - *A Nonlocal SAR Image Denoising Algorithm Based on LLMMSE Wavelet Shrinkage.* IEEE Transactions on Geoscience and Remote Sensing, 50(2):606-616, 2011.

[45] F. ROTTENSTEINER, G. SOHN, J. JUNG, M. GERKE, C. BAILLARD, S. BENITEZ, U. BREITKOPF - *The ISPRS Benchmark on Urban Object Classification and 3D Building Reconstruction.* ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, I-3(1):293-298, 2012.

[46] O. RUSSAKOVSKY, J. DENG, H. SU, J. KRAUSE, S. SATHEESH, S. MA, Z. HUANG, A. KARPATHY, A. KHOSLA, M. BERNSTEIN, *et al.* - *ImageNet Large Scale Visual Recognition Challenge*. International journal of computer vision, 115(3):211-252, 2015.

[47] I. D. STEWART, T. R. OKE - *Local Climate Zones for Urban Temperature Studies.* Bulletin of the American Meteorological Society, 93(12):1879-1900, 2012.

[48] D. SUN, X. YANG, M.-Y. LIU, J. KAUTZ - *PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume.* The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 8934-8943, 2018.

[49] Y. XU, B. DU, L. ZHANG, D. CERRA, M. PATO, E. CARMONA, S. PRASAD, N. YOKOYA, R. HÄNSCH, B. LE SAUX - *Advanced Multi-Sensor Optical Remote Sensing for Urban Land Use and Land Cover Classification: Outcome of the 2018 IEEE GRSS Data Fusion Contest.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pages 1-16, 2019.

[50] D. YANG, J. SUN - *BM3D-Net: a Convolutional Neural Network for Transform-Domain Collaborative Filtering.* IEEE Signal Processing Letters, 25(1):55-59, 2017.

[51] N. YOKOYA, P. GHAMISI, J. XIA, S. SUKHANOV, R. HEREMANS, C. DEBES, B. BECHTEL, B. LE SAUX, G. MOSER, D. TUIA - *Open Data for Global Multimodal Land Use Classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(5):1363-1377, May 2018.

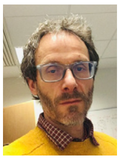
[52] K. ZHANG, W. ZUO, S. GU, L. ZHANG - *Learning Deep CNN Denoiser Prior for Image Restoration.* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3929-3938, 2017.

[53] L. ZHANG, L. ZHANG, B. DU - *Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art.* IEEE Geoscience and Remote Sensing Magazine, 4(2):22-40, 2016.

[54] X. XIANG ZHU, D. TUIA, L. MOU, G.-S. XIA, L. ZHANG, F. XU, F. FRAUNDORFER - *Deep Learning in Remote Sensing: a Comprehensive Review and List of Resources*. IEEE Geoscience and Remote Sensing Magazine, 5(4):8-36, 2017.

AUTHOR

Elise Colin Koeniguer graduated from Supelec, and received her M.Sc. in nuclear physics and particle physics in 2002, and her Ph.D. in 2005 from the University of Paris 6, with a thesis focusing on polarimetric and interferometric radar images. She obtained the degree of Habilitation à Diriger les Recherches (HDR) in 2014 from Paris-Saclay University. She is a Research Engineer at Onera. Her research fields include image processing and computer vision for remote sensing images, with applications to forest and urban areas, as well as polarimetry in the optical and microwave domains.



Guy Le Besnerais graduated from ENSTA in 1989 and obtained his PhD degree from *Université Paris Sud* in 1993. Since 1994 he has been working at ONERA, The French Aerospace Lab, currently in the Information Processing and Systems Department (DTIS) in Palaiseau (91) as Research Director. He obtained the "*Habilitation à Diriger les Recherches*" (HDR) in 2008 and is affiliated to Paris-Saclay University. His research activities include methodology for solving inverse problems, performance modeling for imaging measurement systems, and embedded vision for robotics applications.

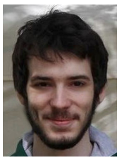
Adrien Chan-Hon-Tong received his M.Sc. from the École Polytechnique in 2011 and his Ph.D. from UPMC in 2014. His research field at the ONERA is computer vision applied to remote sensing, focusing on safety issues such as adversarial attacks, data poisoning, and performance in an operational context.

Bertrand Le Saux (bls@ieee.org) received his M. Eng. in 1999 and his M.Sc. in 1999, both from the INP Grenoble, and his Ph.D. from the University of Versailles/Inria Rocquencourt in 2003. He obtained his *Habilitation à Diriger les Recherches* in 2019 from Paris-Saclay University and is currently a research scientist in the Information Processing and Systems Department at ONERA. His research interests focus on developing machine learning and deep learning methods for remote sensing, (flying) robotics and 3D vision. He teaches machine learning, image processing, and computer vision at the *Institut d'Optique* Graduate School and ENSTA ParisTech. From 2015 to 2019, he was co-chair and later chair of the Image Analysis and Data Fusion Technical Committee of the IEEE GRSS. Since 2019, he has been an associate editor of the IEEE Geoscience and Remote Sensing Letters. He has also been a research fellow at the CNR Pisa (2004), the University of Bern (2005), and the ENS Cachan (2006-2007).

Alexandre Boulch received his M. Eng. from the École Polytechnique, his M.Sc. From the ENS Cachan, and his Ph.D. from *Université Paris Sud*). He received a Ph.D. in geometric and semantic reconstruction for buildings from the ENPC in 2014. For five years, he worked on remote sensing and computational geometry at ONERA. His research included machine learning for semantic segmentation using various data, such as Lidar and satellite imagery. He is currently a research scientist working with Valeo.ai, a research team dedicated to autonomous and assisted vehicles. He has mainly worked on 3D perception and scene analysis.

Pauline Trouvé graduated from the *Institut d'Optique* Graduate School in 2009 and obtained her Ph.D. degree in signal and image processing from the *École Centrale de Nantes* in 2012. Since 2012 she has been working in the Information Processing and Modeling Department at ONERA, the French Aerospace Lab. Her research field at ONERA concerns image processing (deblurring and denoising), optical design, and performance modeling for imaging systems. In particular, she has been working on an unconventional imaging system design using a joint digital/optical performance model. Her work is applied in computational photography and computer vision, especially in the field of 3D applications.

Rodrigo Caye Daudt received his degree in Electrical Engineering from the *Universidade de Brasília* in 2015, and his joint master's degree in Computer Vision and Robotics from the *Université de Bourgogne*, the *Universitat de Girona*, and Heriot-Watt University in 2017. Since 2017 he has been a Ph.D. student at Télécom Paris' Information Processing and Communication Laboratory (LTCI) and in ONERA's Information Processing and Systems Department (DTIS). His research studies the application of computer vision and machine learning methods to multitemporal remote sensing images.



Nicolas Audebert received his M. Eng. degree from Supélec (2015), his M.Sc. in Computer Science from *Université Paris-Sud* (2015), and his Ph.D. from *Université Bretange-Sud* (2018) in the field of deep learning for multimodal remote sensing. He is currently an assistant professor at the CNAM in the Vertigo team. His research interests focus on machine learning for image understanding, Earth observation, and data fusion.

Guillaume Brigot received his M. Eng. degree from Supélec (2014), his M.Sc. in Electromagnetism and Communication from *Université Paris-Sud* (2015), and his Ph.D. in Information Sciences from Paris-Saclay University in 2017, in the fields of machine learning and fusion of LIDAR and RADAR. He is currently an image processing engineer at MBDA.

Pierre Godet graduated from the *Institut d'Optique* Graduate School and received his Master's degree from *Université Paris-Saclay* in the field of signal and image processing. He is currently a Ph.D. student at ONERA in the Department of Information Processing and Systems, in the field of learning-based approaches for multi-frame optical flow estimation.

Benjamin Le Teurnier received his engineering degree from the *Institut d'Optique* Graduate School and his Master's degree from *Université Paris-Saclay* in the field of signal and image processing in 2019. He started a Ph.D. at the *Laboratoire Charles-Fabry* of *Université Paris-Saclay* the same year, in the field of polarimetric imaging.

Marcela Carvalho received her engineering diploma in Mechatronics in 2014 from the *Universidade de Brasilia*, followed by a Master diploma in Robotics in 2015 from the *Université de Montpellier*, and a Ph.D. degree in 2009 in the field of deep learning from Defocus. Her research activities concern robotics and computer vision. During her undergraduate studies, she participated in the creation of the Droid Team, a robotics team that participates in Brazilian national and International intelligent robotics competitions, and co-founded the Lua Startup in 2016.

Javiera Castillo Navarro holds an engineering degree from *École Centrale Paris* and from the *Universidad de Chile* (2017), and she received her M.Sc. in Applied Mathematics from the *Universidad de Chile* (2017) and her M.Sc. in Data Science from *Université Paris-Saclay* (2018). She is currently a Ph.D. student at ONERA, working on deep learning for Earth observation.