# Using Satellite Imagery and Deep Learning Algorithms for Population Census in Overseas Departments


**Clément Guillo[1], Gaëtan Carrere[2]**

*[1]Head of Statistical Service at the Interregional Directorate for the French West Indies and Guiana of INSEE (National Institute of Statistics and Economic Studies)*

*[2]Intern at Interregional Directorate for the French West Indies and Guiana of INSEE (National Institute of Statistics and Economic Studies)*



## Abstract

The credibility of the figures produced by INSEE is often questioned in the Overseas Departments. Population census figures, in particular, are heavily criticized by local elected officials in Guyana and Mayotte, prompting INSEE to work on innovative data sources to support its statistical production. The use of satellite imagery allows, on one hand, to complement the estimates produced by INSEE by aligning the observed evolution of buildings on the images with the population estimates produced by the population census. On the other hand, it supports the operation of locating housing on the ground by anticipating in advance, through the images, the areas where creation or destruction movements are most significant.

Deep Learning algorithms trained on these images are able to automatically detect building outlines in the Overseas Departments very accurately. The algorithm showing the best performance has been wrapped into a web application for office agents, allowing them to make decisions based on the areas displayed by the algorithm and the raw satellite images. The entire processing chain, from image retrieval to the decision support web application, including the training of Deep Learning algorithms, requires a variety of skills and a high level of technical expertise to be maintained.



**Keywords:** satellite data, census, segmentation models, building detection


## 1. Introduction

In French Guiana and Mayotte, the figures from the population census (RP) are often questioned by politicians and the population. Therefore, confirming the observed changes by INSEE through external sources is crucial. This paper aims to present the work carried out at INSEE within the "satellite data" project and to explain the potential contribution of earth observation data. The project involves using satellite images to identify, through artificial intelligence, the location of housing units. This is done to more precisely direct the human resources deployed for cartographic surveys and to support the estimates produced by INSEE's official population census operation.

This project mobilizes diverse skills, and its complexity requires a general understanding of different subjects: satellite image manipulation, understanding the structure of an image from a computer science perspective, mastery of deep learning tools and algorithm training methods, documentation on the progress of artificial intelligence research, processing skills, and making results available via an application. We will therefore attempt to address and detail all aspects of this project. Initially, we will recall the context of establishing this working group around satellite data, precisely outlining the needs it aims to address. Secondly, we will delve into the available data, satellite images, and their preprocessing before any analysis. We will then recall the principle of deep learning segmentation algorithms, justifying both their relevance in this specific context and presenting how we applied them to this problem and these data. This will allow us to detail the training of the algorithms and present the results obtained. Since these results are not immediately exploitable, we will present the processing performed. Finally, this paper will present avenues for improvement and continuation of the project, as well as future utilization prospects.

**1.1 Context**

Every year, the population census mobilizes around a hundred of pollsters in the Overseas Departments (DOM). Before this data collection phase, the inventory of buildings located in the DOM is updated through a preliminary cartographic survey. This inventory must contain a comprehensive list of geolocated housing units in the DOM, from which housing units are selected for enumeration in the current year's census. This survey is specific to the DOM because the usual administrative databases are not reliable enough to solely populate this inventory.

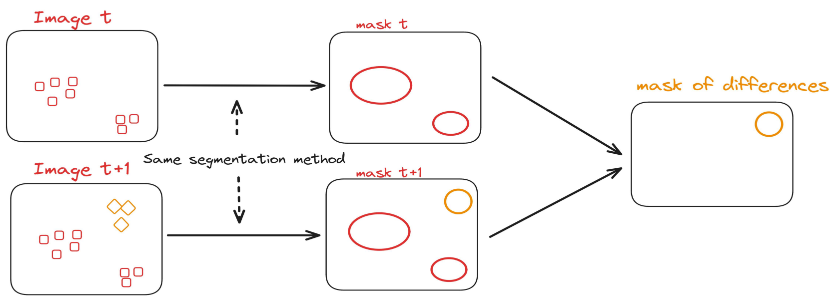
The cartographic survey helps to feed the directory of located buildings (RIL). A high-quality RIL allows pollsters to more easily locate housing units to be surveyed in a given year. The impact of a good RIL on the quality of estimates produced by INSEE is therefore considerable. The use of satellite imagery, particularly methods for detecting housing units in these images, could, for example, improve the organization of the cartographic survey, particularly by calibrating efficiently the interviewers' workload in the territories surveyed.

Moreover, in response to the demands of Mayotte's elected officials, INSEE replaced general censuses with Annual Census Surveys (EAR), leading to a significant data gap since 2017 and delaying population updates until 2025, making the use of satellite data crucial for improving the accuracy of census information and calibrating EAR preparation.

**1.2 Global structure of the project**

Schematically, we want to be able to automatically detect changes from shots of the same territory at two different dates. To do this, we train an algorithm capable ultimately of producing housing masks from a given image, that is to say, a layer of polygons representing inhabited buildings on a map. By analyzing the differences between two housing masks produced for the same territory at two different times, we can try to deduce the main movements between these two dates, notably the creation and destruction of housing (cf. Figure 1). Measuring these evolutions can then extend past population estimates made by INSEE and also help identify areas where cartographic survey should focus.

Figure 1: Segmentation Strategy for change detection




Training these housing detection algorithms requires the creation of a large number of pairs (images, masks) where:

- Images are divided into tiles small enough to be absorbed by the deep learning model. It is also necessary to ensure the absence of cloud cover, making analysis impossible.
- Masks are arrays of the same dimension as the image, drawing the presence of housing with values of 0 or 1, aggregating into polygons drawing the buildings. These masks are created from data from INSEE, notably the topographic database (BDTOPO) provided by the french national geographic institute (IGN). These examples are built from past data and will be used for detection on more recent shots for which these annotations are not available.

Thus, the project as it exists today can be broken down into several distinct parts:

- A first processing chain includes the creation of pairs (images, mask) from IGN data which will feed the algorithm.
- Another processing chain completely automates the training of algorithms from the data created in the upstream chain. The use of services such as Argo-workflow and MLFlow allows for controlled production deployment, training history, and rigorous reproducibility of the training context.
- The last processing chain consists of analyzing the predictions made by the algorithm and providing visual and statistical results for investigators and INSEE office staff wishing to compare these predictions with population figures established by INSEE through the population census.

## 2. Data preparation

### 2.1 Satellite images

Several sources of image data were considered for this work. In satellite imagery, a distinction is made between high-resolution data and very high-resolution data. In the following, we will focus on PLEIADES data with very high resolution. These data are produced by the company Airbus and are retrieved and concatenated by IGN. Thus, this organization provides us with complete coverage of the Antilles territories every year. Two characteristics are very important when it comes to satellite images. The first is spatial resolution, which is the surface covered by a pixel: the higher the spatial resolution, the smaller the surface covered by a pixel. For PLEIADES data, the spatial resolution is 0.5m. For comparison, the spatial resolution of Sentinel 2 images is 10 m, which is 20 times lower. The second characteristic is temporal resolution, *i.e.*, the frequency at which a photograph of the ground in a given area can be obtained. The higher this is, the more recent and therefore relevant the available images will be for the desired use case. The images obtained are via optical measurements, which implies that excessive cloud cover during satellite imaging will delay acquisition for the concerned territory. On average, 8 months are therefore needed to have a complete cloud-free acquisition of these territories through PLEIADES satellite imaging, which leads to a gap between the reality of the terrain and that photographed by the satellites.

### 2.2 Annotations

The objective here is to create, from a set of images covering the DOM territories, housing masks associated with the images. It is worth noting here that manual labelling is the best solution from the perspective of the quality of the generated masks, since the time gap between the date of constituting the databases used to annotate the images and the date of capturing these images will necessarily lead to imperfectly synchronized masks with the images. The cost of such manual labelling is prohibitive given the size of the project team and the available working time. Therefore, automatic labelling was performed using the IGN's topographic database (BDTOPO). This database precisely locates the outline of buildings each year with polygons, obtained through a combination of aerial photography processing and manual annotations by IGN agents.

However, BDTOPO cannot directly address the use cases mentioned in the introduction for the following reasons:

- The yearly BDTOPO produced aims to represent the spread of buildings for a given year. However, polygons outlining buildings for a given year may appear different between different versions without this evolution being linked to real creation. This situation may occur if the methods for detecting buildings by IGN improve from one year to another, and a building previously undetected is eventually detected.
- A given territory is covered by satellite image tiles, which are obtained gradually as the satellite passes over the territories, depending on cloud cover and the tilt of optical radars. These tiles are thus obtained at different times of the year. Consequently, the representation of a territory by satellite imagery is a mosaic of images obtained at different times. Therefore, it is hardly conceivable that BDTOPO will perfectly coincide with these images, so it cannot be used as is.
- The updating of BDTOPO by IGN is not guaranteed, so INSEE must internalize this building detection process.

An example of the annotation obtained via the BDTOPO is shown in Figure 2. Subsequently, we will train an algorithm whose generalization capacity will not suffer from the above-mentioned annotations mistakes.

Figure 2: Annotations obtained with BDTOPO on a PLEIADES example

It should be noted that the masks produced automatically using BDTOPO are not limited to the concept of housing but rather to that of buildings. This is problematic because an algorithm trained on these masks will also only be able to detect buildings.

### 2.3 Segmentation Models

Segmentation models can be seen as algorithms that classify each pixel in the image individually. In the context of building detection, a segmentation algorithm takes an image as input and assigns a probability of building presence to each pixel (0,1). The predicted mask is then obtained by thresholding this probability and classifying all pixels in the image for which the probability exceeds a given threshold as buildings.

Segmentation models are often based on a U-shaped structure composed of:

- A descending part, the encoder, which transforms the input image into a reduced-size numerical vector compared to the initial number of pixels in the image. The encoder part is essentially a standard convolutional neural network as presented in the previous section.
- An ascending part, the decoder, which starts from the obtained vector (also called embedding) and reverses the process through operations known as inverse convolutions to output an image of the same dimension as the input.

The quality of the segmentation process by the algorithm is closely related to the quality of the encoder, which aims to send the images into a space that is expressive and interpretable by the decoder.

Both parts (encoder and decoder) are parameterized and are thus improved during training. Many layers eventually separate the input image from the produced mask. Due to successive Max Pooling operations, whose primary purpose is to reduce the number of parameters in the network, local information is lost through the layers of the network, and the vectorized information output by the encoder no longer sufficiently reflects the local phenomena, smoothed by this aggregation operation. In Visin et al. 2016 and Jégou et al. 2017, the authors present segmentation model structures that address this issue. In these models, the elements output by certain layers serve as input to several of the following layers, visually bypassing certain layers. The U-net, introduced by Ronneberger, Fischer, and Brox in 2015, extends this logic by connecting layers from the contracting part to the expansive part. Figure 16 illustrates the structure of the U-net. Other architectures, such as those presented in Chen et al. 2017, rely on specific convolutional forms (such as atrous convolution) aimed at minimizing the summarizing effect of Max Pooling operations. These structures are very complex, and several simplified versions have been produced subsequently.

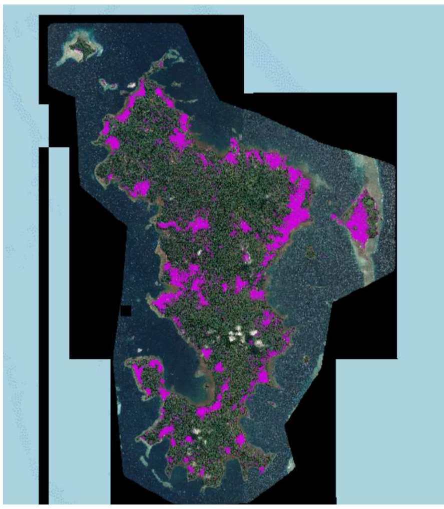
More recently, the construction of segmentation models has been greatly inspired by Large Language Models (e.g., GPT), whose effectiveness is well established. By analogy with word sequences, if we consider images as sequences in two dimensions, we can apply "transformer" structures (cf. Vaswani et al. 2017) to them. Thus, in Dosovitskiy et al. 2021, the authors demonstrate that a classification model can be trained using only transformers (as opposed to convolutional neural networks). In Xie et al. 2021, the authors generalize this approach to segmentation models using a decoder based on a transformer structure. The main advantage of using such structures is the efficiency gain (performance at fixed number of parameters), which contrasts with the convolutional neural network-based structures presented earlier. The training described below will rely on a Segformer-type model.

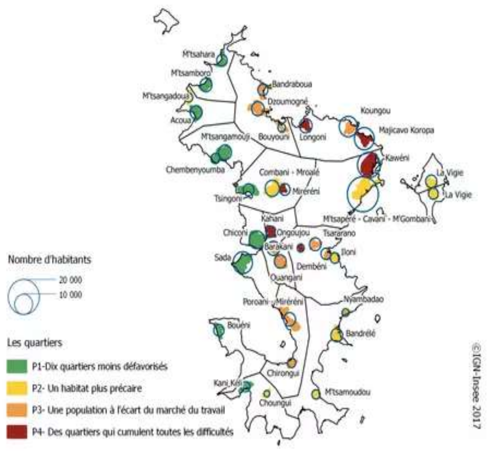
### 2.4 Results

Training of a Segformer network was conducted using PLEIADES images covering Martinique and Guadeloupe in 2022. To annotate these images, we used the available versions of BDTOPO produced by IGN, accepting the inevitable divergences due to the different temporalities of image captures and topographic database construction. The algorithms trained from this training set can only detect buildings. A Segformer-type segmentation model was trained on the obtained pairs. Training lasts an average of ten hours.

To evaluate the results of our algorithm during training, some correctly labelled areas of Mayotte were selected. We calculate the average Intersection Over Union (IOU) obtained on the considered dataset. IOU measures the overlap of the algorithm's predictions with known annotations, ranging between 0 (no overlap) and 1 (perfect overlap). At the end of the training, the algorithm's performance on the test set was 75 %. A comparison of 2023 predictions for the island of Mayotte and a map of favored or non-favored neighbourhoods in Mayotte is shown in figure 3. The built and inhabited areas are highlighted on the map on the left. This general view of the outputs helps to justify the relevance of the analysis: with the trained model, we are able to delineate the built-up areas visible on satellite images.

Figure 3: Comparison of model predictions for Mayotte with statistics produced by INSEE






Focusing on more specific areas, i.e., zooming in on the cartographic representation of polygons, we observe that the model is capable of automatically drawing the buildings. The example Figure 5 presents predictions made on images from 2023. These predictions superimposed on 2020 images show built-up areas in 2023 that were not built-up in 2020.

Figure 4: 2023 Predictions on a 2020 background (L) and on a 2023 background (R)

The results are satisfactory, as we highlight the appearance of new buildings and underscore the need to allocate human resources for the mapping of this particular area. Moreover, increasing the precision and completeness of the predictions requires training new, either more efficient models or models trained on other images (see the previous section). Secondly, it seems impossible to ask surveyors to examine the entire island by zooming in so much. Therefore, for a proper dissemination of the results, it is necessary to develop tools that allow for both quicker and easier reading. The work presented in Appendix A presents a method for extracting the main movements in building creation or destruction from predictions made in 2020 and 2023 by the algorithm.

## 3. Discussion

The final uses of the presented work (population estimates and workload planning) depend entirely on the quality of the predictions made by the algorithm. There are numerous possible directions to try to improve these predictions, particularly at the level of the training data. Currently, the algorithm is only trained on the year 2022 and on Martinique and Guadeloupe. It is necessary to test the algorithm's generalization ability by reducing or expanding, temporally or geographically, the training set that feeds its learning.

The masks built from satellite images are masks of buildings, not masks of dwellings, which means the algorithm cannot distinguish between housing and buildings. Manual labeling work could correct the masks produced via BDTOPO in this regard. Other data sources, outside of INSEE, could be explored to create masks, such as OpenStreetMap.

At the image level itself, there are multiple possibilities: other sources of satellite imagery are available, such as Sentinel images or Spot satellite images. These sources have lower spatial resolutions than PLEIADES images, but their temporal resolution is higher, which means they could be used to make provisional estimates while waiting for complete PLEIADES coverage. Work on Sentinel 2 imagery is presented in the Nabec 2023 report and shows that a very satisfactory level of building prediction can be obtained from these lower-resolution images. Sources of stereoscopic images add additional data on the altitude of the visualized buildings and also deserve to be explored.

Reflections on the selected algorithm are equally important. Indeed, the scientific literature is abundant on segmentation models, so it is necessary to conduct more frequent technical monitoring on the subject. Some models can theoretically adapt to images of different resolutions. Thus, the training of these algorithms could be augmented by images from multiple sources, including pre-annotated datasets made available by academic research efforts. Appendix B provides details of the technical stack and skills required to develop this work. A working document produced at INSEE details the work presented in this paper and is available here.

## References


Berova, Raya (2023). Détection automatisée de changements des bâtiments en France d'Outre-Mer à partir d'images satellites. Rapport de stage. Direction Régionale de l'Insee Martinique, INSEE. https://minio.lab.sspcloud.fr/cguillo/rapport_stage/Rapport_de_stage_3A-2.pdf

Chabennet, Quentin (2021). Détection automatique de la couverture des sols à partir d'images satellites à l'aide de méthodes de segmentation sémantique. Mémoire de stage. ECE engineering school. https://minio.lab.sspcloud.fr/cguillo/rapport_stage/rapport_stage_quentin_dmrg-1.pdf

Chen, Liang-Chieh et et al. (déc. 2017). « Rethinking Atrous Convolution for Semantic Image Segmentation ». In: arXiv. arXiv: 1706.05587 [cs.CV]. url: http://arxiv.org/abs/1706.05587.

Berova R., Carrere G., Faria T., Guillo C.,.Seimandi T., (2024). Usage de l'imagerie Satellitaire et des algorithmes de Deep Learning au service du Recensement de la population dans les DOMS. *Working Document, INSEE.* https://minio.lab.sspcloud.fr/cguillo/rapport_stage/document_travail.pdf

Dosovitskiy, Alexey et al. (juin 2021). « An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale ». In: arXiv preprint arXiv: 2010.11929. url: http://arxiv.org/abs/2010.11929.

Nabec, Judith (2023). Mise en place d'une méthode de détection automatique des évolutions de bâti en Outre-Mer sur des images satellites. Rapport de stage. Direction Générale de l'Insee. https://minio.lab.sspcloud.fr/cguillo/rapport_stage/Rapport_de_stage_2023-3.pdf

Ronneberger, Olaf, Philipp Fischer et Thomas Brox (mai 2015). « U-Net: Convolutional Networks for Biomedical Image Segmentation ». In: arXiv: 1505.04597 [cs]. arXiv: 1505.04597. url http://arxiv.org/abs/1505.04597.

Vaswani, Ashish et al. (2017). « Attention Is All You Need ». In: Advances in Neural Information Processing Systems. T. 30. Curran Associates, Inc. url: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

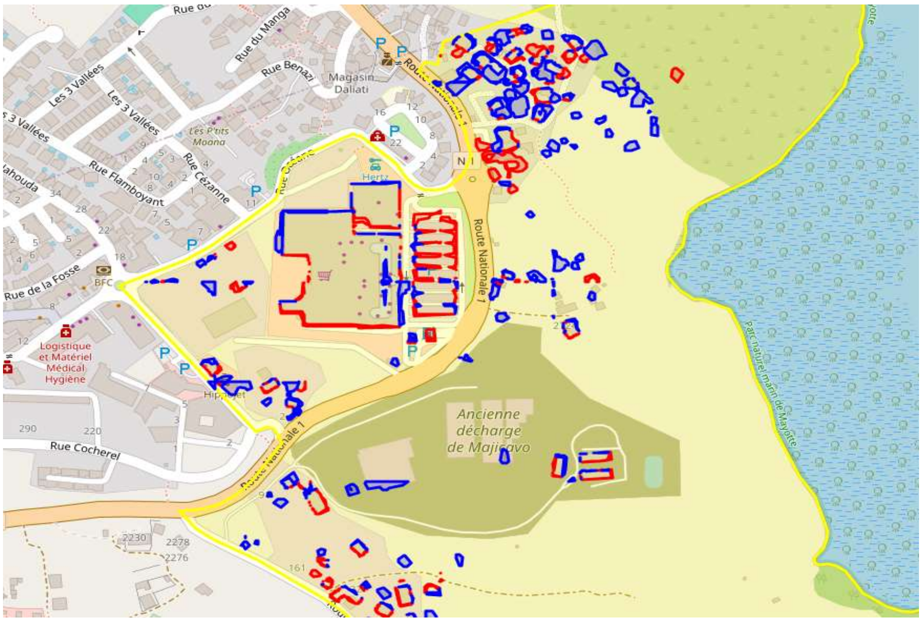
Visin, Francesco et al. (mai 2016). « ReSeg: A Recurrent Neural Network-based Model for Semantic Segmentation ». In: arXiv: 1511.07053 [cs]. arXiv: 1511.07053. url: http://arxiv.org/abs/1511.07053.

Xie, Enze et al. (oct. 2021). « SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers ». In: arXiv preprint arXiv: 2105.15203. url: http://arxiv.org/abs/2105.15203.

## APPENDIX A: Cleaning of the Subtraction

In this section, we present cleaning operations, which a posteriori make it possible to display changes based on the difference in masks at the output of the algorithm. We start by naively differentiating between the 2020 and 2023 predictions obtained using the algorithm on Mayotte. Figure 1 shows the resulting polygons. Due to the imperfect superposition of predictions from two different years, some building contours remain as residuals. We notice that the outlines of the buildings do not necessarily have the smallest area, but they have a characteristic shape: they are particularly elongated. The compactness index calculates a ratio between the perimeter and the area of a polygon. It thus varies between 0 and 1. The value 0 is a perfectly elongated line, while the value 1 corresponds to a perfect circle.

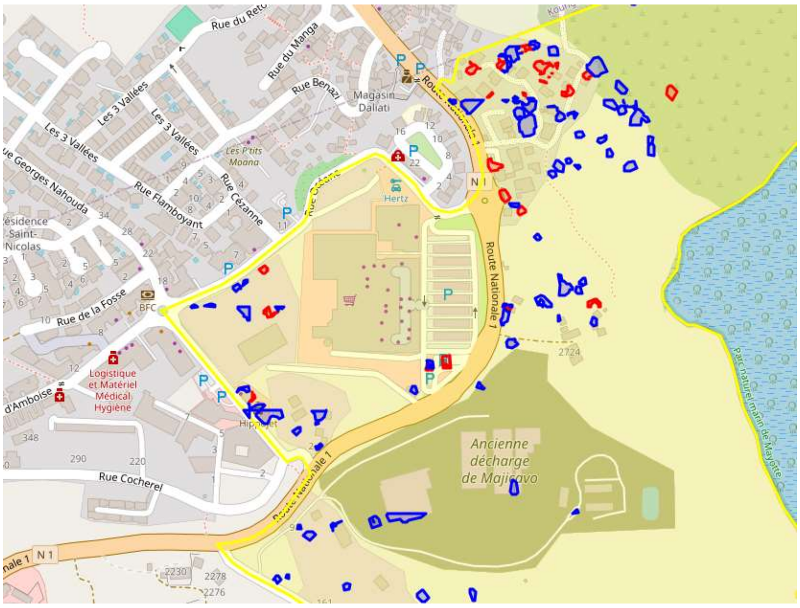
Figure 1: Mask of differences obtained by subtracting algorithm' s 2020 predictions from 2023 predictions on an example in Mayotte




Before filtering the polygons, a threshold must be determined. Here, we choose a threshold of 0.1. This decision is not simple and was approached gradually, by progressively observing the state of the polygons displayed on the map, and by visually and manually comparing them to the images of 2020 and 2023.

Finally, we can filter the results of this subtraction and obtain the visualization (Figure 2) of the real modifications to the state of the buildings, which precisely corresponds to the objectives of the project.

Figure 2: Mask of differences obtained after eliminating residues with too low a compactness index




These "cleaned" polygons now support the census figures in Mayotte, in particular, by providing a database on housing, on which a population estimate can rely and corroborate the figures produced by the annual surveys. They can also direct and support investigators for the cartographic survey.

However, regarding this second task, the data can be exploited to produce statistics and cartographic visualization by block, for example. Thus, by highlighting the geographical areas that are evolving the fastest, by ranking the blocks according to their importance as cartographic support, we can provide the most precise and useful information possible. These are statistical works that we are able to do.

## APPENDIX B: Technical Stack/ Technical Debt

Here, we highlight all the tools and skills required for the existence, maintenance, and future development of this project. Firstly, a solid expertise in the INSEE databases is indispensable. A good understanding of databases and geographic information systems is crucial for manipulating building polygons, housing coordinates (x,y), and geolocated image pixels.

Additionally, the Python package [astrovision](https://github.com/InseeFrLab/astrovision) (https://github.com/InseeFrLab/astrovision) has been developed to facilitate the manipulation of images and associated masks. Basic skills in statistical learning are also required to avoid significant overfitting errors. Furthermore, a slightly deeper understanding of deep learning algorithms is necessary, along with the ability to comprehend and reproduce models presented in the latest research articles on the subject.

Lastly, considerable effort has been made to capitalize on all trainings, ensure their reproducibility, and facilitate their execution, particularly through the use of and expertise in monitoring tools such as MLFlow (Figure 1) or task programming tools like Argoworkflow.

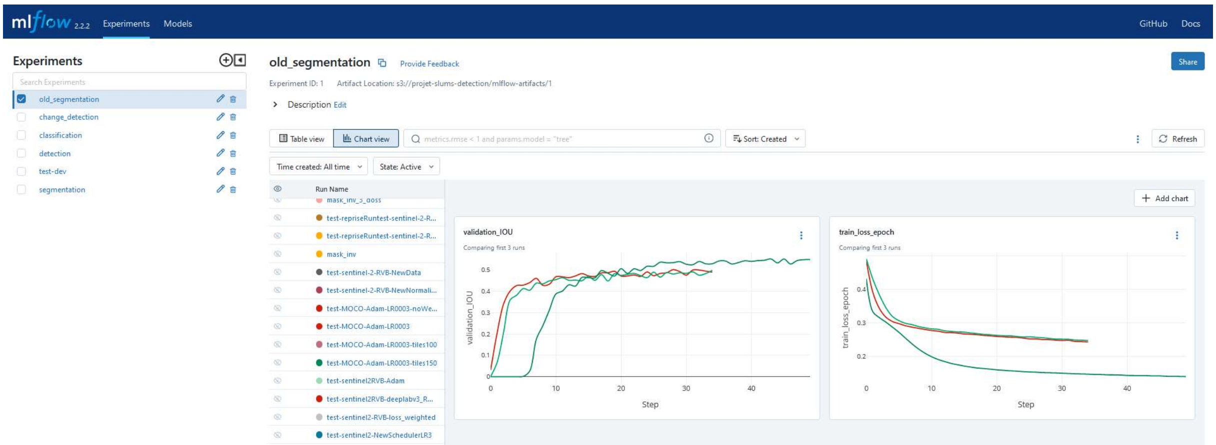
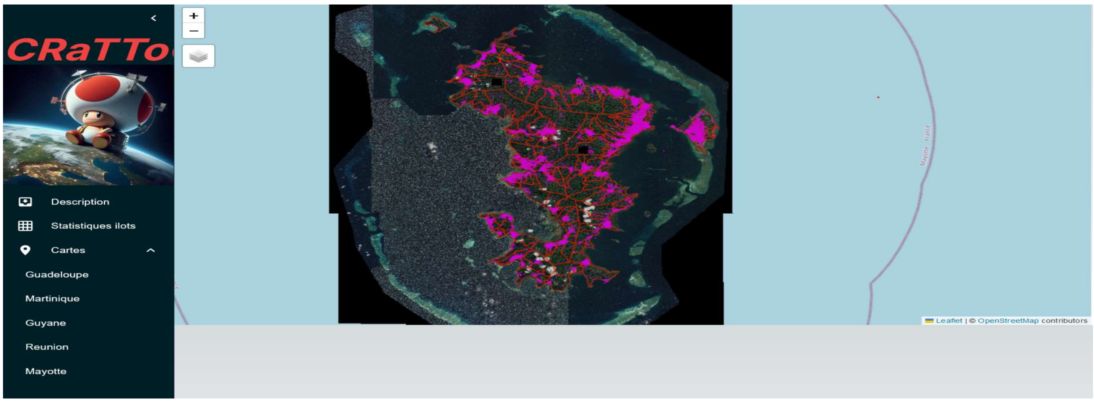



Figure 1: Example of an interface on MLFlow for monitoring algorithm training.

Regarding the management of geographic data (images, polygons, administrative has been implemented to dynamically serve image tiles and predictions generated by the algorithms over multiple years and territories.

Finally, a web application developed in React is used to highlight the algorithm's results and work on the difference of the aforementioned masks to allow validation by INSEE workers in the office, plan the cartographic survey, and retrieve produced evolution statistics. This application is also very useful for the project team to assess at a glance and on a large scale the relevance of our algorithms.

Figure 2: Web Application querying the results of the algorithm and the images hosted on the GeoServer




All the technical blocks presented here mentioned here are hosted on INSEE servers, and their integration requires mastery of application deployments via the Kubernetes container management tool. Hence, significant technical debt has accumulated on each link of the processing chain, and maintaining the entire system would already require a full-time team of several people. The diagram in Figure 3 summarizes all the tools and skills mobilized on the project.