Tree object detection using airborne images and LiDAR point clouds

Alexandre Bry

École polytechnique, Geodan B.V.

alexandre.bry.21@polytechnique.org

2024-07-15

Abstract   This is the abstract. It can be on multiple lines and contain **Markdown**.

Table of contents

1 Introduction

The goal of the internship was to study the possibility of combining LiDAR point clouds and aerial images in a deep learning model to perform instance segmentation of trees. The two types of data are indeed complementary, as point clouds capture the shape of the worlds, while images capture the colors. However, combining them into a format that allows a model to handle them simultaneously is not an easy task because they inherently have a very different spatial repartition and encoding.

The second major topic of the internship was to acquire a proper dataset matching all the criteria required for the project. Most of the datasets containing tree annotations only used either RGB images or LiDAR point clouds, but not both. Therefore, I had to create such a dataset by myself, using the openly available images and point clouds in the Netherlands, by

annotating trees by hand to properly train and evaluate the methods.

2 State-of-the-art

2.1 Computer vision tasks related to trees

Before talking about models and datasets, let's define properly the task that this project focused on, in the midst of all the various computer vision tasks, and specifically those related to tree detection.

The first main differentiation between tree recognition tasks comes from the acquisition of the data. There are some very different tasks and methods using either ground data or aerial/satellite data. This is especially true when focusing on urban trees as shown in T. Arevalo-Ramirez *et al.* [1], since a lot of street view data is available.

This leads to the second variation, which is related to the kind of tree area that we are interested in. There are mainly three types of area, which among other things, influence the organization of the trees in space: urban areas, tree plantations and forests. This is important, because the tasks and the difficulty depends on the type of area. Tree plantations are much easier to work with than completely wild forests, while urban areas contain various levels of difficulty ranging from alignment trees to private and disorganized gardens and parks. For this project, we mainly focused on urban areas, but everything should still be applicable to tree plantations and forests.

Then, the four fundamental computer vision tasks have their application when dealing with trees, as explained in [2]:

- Classification, although this is quite rare for airborne tree applications since there are multiple trees on each image most of the time
- Detection, which consists in detecting objects and placing boxes around them
- Semantic segmentation, which consists in associating a label to every pixel of an image
- Instance segmentation, which consists in adding a layer of complexity to semantic segmentation by also differentiating between the different instances of each class

These generic tasks can be extended by trying to get more information about the trees. The most common information are the species and the height.

2.2 Datasets

2.2.a Requirements

Before presenting the different promising datasets and the reasons why they were not fully usable for the project, let's enumerate the different conditions and requirements for the tree instance segmentation task:

- Multiple types of data:
  ‣ Aerial RGB images
  ‣ LiDAR point clouds (preferably aerial)
  ‣ (Optional) Aerial infrared images
- Tree crown annotations or bounding boxes
- High-enough resolution:
  ‣ For images, about 25 cm
  ‣ For point clouds, about 10 cm

Here are the explanations for these requirements. As for the types of data, RGB images and point clouds are required to experiment on the ability of the model to combine the two very different kinds of information they hold. Having infrared data as well could be beneficial, but it was not necessary. Regarding tree annotations, it was necessary to have a way to spatially identify them individually, using crown contours or simply bounding boxes. Since the model outputs bounding boxes, any

kind of other format could easily be transformed to bounding boxes. Finally, the resolution had to be high enough to identify individual trees and be able to really use the data. For the point clouds especially, the whole idea was to see if and how the topology of the trees could be learnt, using at least the trunks and even the biggest branches if possible. Therefore, even if they are not really comparable, this is the reason why the required resolution is more precise for the point clouds.

Unfortunately, none of the datasets that I found matched all these criteria. Furthermore, I didn't find any overlapping datasets that I could merge to create a dataset with all the required types of data. In the next parts, I will go through the different kinds of datasets that exist, the reasons why they did not really fit for the project and the ideas I got when searching for a way to use them.

2.2.b Existing tree datasets

As explained above, there were quite a lot of requirements to fulfill to have a complete dataset usable for the task. This means that almost all the available datasets were missing something, as they were mainly focusing on using one kind of data and trying to make the most out of it, instead of trying to use all the types of data together.

The most comprehensive list of tree annotations datasets was published in A. Ouaknine, T. Kattenborn, E. Laliberté, and D. Rolnick [3]. N. I. Bountos, A. Ouaknine, and D. Rolnick [4] also lists several interesting datasets, even though most of them can also be found in A. Ouaknine, T. Kattenborn, E. Laliberté, and D. Rolnick [3]. Without enumerating all of them, there were multiple kinds of datasets that all have their own flaws regarding the requirements I was looking for.

Firstly, there are the forest inventories. T. Jucker *et al.* [5] is probably the most interesting one in this category, because it contains a lot of spatial information about almost 500K trees, with their locations, their crown radii and their heights. Therefore, everything needed to localize trees is in the dataset. However, I didn't manage to find RGB images or LiDAR point clouds of the areas where the trees are located, making it impossible to use these annotations to train tree detection.

Secondly, there are the RGB datasets. G. Reiersen *et al.* [6] and B. Weinstein [7] are two of them and the quality of their images are high. The only drawback of these datasets is obviously that they don't provide any kind of point cloud, which make them unsuitable for the task.

Thirdly, there are the LiDAR datasets, such as E. Kalinicheva, L. Landrieu, C. Mallet, and N. Chehata [8] and S. Puliti *et al.* [9]. Similarly to RGB datasets, they lack one of the data source for the task I worked on. But unlike them, they have the advantage that the missing data could be much easier to acquire from another source, since RGB aerial or satellite images are much more common than LiDAR point clouds. However, this solution was abandoned for two main reasons. First it is quite challenging to find the exact locations where the point clouds were acquired. Then, even when the location is known, it is often in the middle of a forest where the quality of satellite imagery is very low.

Finally, I also found two datasets that had RGB and LiDAR components. The first one is J. Hu *et al.* [10]. This benchmark dataset encompasses RGB images, hyperspectral images and Digital Surface Models (DSM). There were however two major flaws. The obvious one was that this dataset was created with land semantic segmentation tasks in mind, so there was

no tree annotations. The less obvious one was that a DSM is not a point cloud, even though it is some kind of 3D information and was often created using a LiDAR point cloud. As a consequence, I would have been very limited in my ability to use the point cloud.

The only real dataset with RGB and LiDAR was B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White [11]. This dataset contains exactly all the data I was looking for, with RGB images, hyperspectral images and LiDAR point clouds. With 30975 tree annotations, it is also a quite large dataset, spanning across multiple various forests. The reason why I decided not to use it despite all this is that at the beginning of the project, I thought that the quality of the images and the point clouds was too low. Looking back on this decision, I think that I probably could have worked with this dataset and gotten great results. This would have saved me the time spent annotating the trees for my own dataset, which I will talk more about later. My decision was also influenced by the quality of the images and the point clouds available in the Netherlands, which I will talk about in the next section.

2.2.c Public data

After rejecting all the available datasets I had found, the only solution I had left was to create my own dataset. I won't dive too much in this process that I will explain in Section 3. I just want to mention all the publicly available datasets that I used or could have used to create this custom dataset.

For practical reasons, the two countries where I mostly searched for available data are France and the Netherlands. I was looking for three different data types independently:

- RGB (and eventually infrared) images
- LiDAR point clouds
- Tree annotations

These three types of data are available in similar ways in both countries, although the Netherlands have a small edge over France. RGB images are really easy to find in France (Institut national de l'information géographique et forestière (IGN) [12]) and in the Netherlands (Beeldmateriaal Nederland [13]), but the resolution is better in the Netherlands (8 cm vs 20 cm). Hyperspectral images are also available in both countries, although for those the resolution is only 25 cm in the Netherlands.

As for LiDAR point clouds, the Netherlands have a small edge over France, because they are at their forth version covering the whole country with Actueel Hoogtebestand Nederland [14], and are already working on the fifth version. In France, data acquisition for the first LiDAR point cloud covering the whole country (Institut national de l'information géographique et forestière (IGN) [15]) started a few years ago. It is not yet finished, even though data is already available for half of the country. The other advantage of the data from Netherlands regarding LiDAR point clouds is that all flights are performed during winter, which allows light beams to penetrate more deeply in trees and reach trunks and branches. This is not the case in France.

The part that is missing in both countries is related to tree annotations. Many municipalities have datasets containing information about all the public trees they handle. This is for example the case for Gemeente Amsterdam [16] and Bordeaux Métropole [17]. However, these datasets cannot really be used as ground truth for a custom dataset for several reasons. First, many of them do not contain coordinates indicating the position of each tree in the city. Then, even those that contain coordinates are most of the time missing any kind of information allowing to deduce a bounding box for the tree crowns. Finally, even if they did contain

everything, they only focus on public trees, and are missing every single tree located in a private area. Since public and private areas are obviously imbricated in all cities, it means that any area we try to train the model on would be missing all the private trees, making the training process impossible because we cannot have only a partial annotation of images.

The other tree annotation source that we could have used is Boomregister(Coöperatief Boomregister U.A. [18]). This work covers the whole of the Netherlands, including public and private trees. However, the precision of the masks is far from perfect, and many trees are missing or incorrectly segmented, especially when they are less than 9 m heigh or have a crown diameter smaller than 4 m. Therefore, even it is a very impressive piece of work, we thought that it could not be used as training data for a deep learning models due to its biases and imperfections.

2.2.d Dataset augmentation techniques

2.3 Models

3 Dataset

4 Model

5 Results

Beyond mAP: R. Jena *et al.* [19].

6 Conclusion

Blablabla

Bibliography

[1] T. Arevalo-Ramirez *et al.*, "Challenges for computer vision as a tool for screening urban trees through street-view images," *Urban Forestry & Urban Greening*, vol. 95, p. 128316–128317, 2024, doi: https://doi.org/10.1016/j.ufug.2024.128316.

[2] A. Safonova, E. Guirado, Y. Maglinets, D. Alcaraz-Segura, and S. Tabik, "Olive Tree Biovolume from UAV Multi-Resolution Image Segmentation with Mask R-CNN," *Sensors*, vol. 21, no. 5, p. 1617–1618, 2021, doi: 10.3390/s21051617.

[3] A. Ouaknine, T. Kattenborn, E. Laliberté, and D. Rolnick, "OpenForest: A data catalogue for machine learning in forest monitoring." 2023.

[4] N. I. Bountos, A. Ouaknine, and D. Rolnick, "FoMo-Bench: a multi-modal, multi-scale and multi-task Forest Monitoring Benchmark for remote sensing foundation models," *arXiv preprint arXiv:2312.10114*, 2023, [Online]. Available: https://arxiv.org/abs/2312.10114

[5] T. Jucker *et al.*, "Tallo: A global tree allometry and crown architecture database," *Global Change Biology*, vol. 28, no. 17, pp. 5254–5268, 2022, doi: https://doi.org/10.1111/gcb.16302.

[6] G. Reiersen *et al.*, "ReforesTree: A Dataset for Estimating Tropical Forest Carbon Stock with Deep Learning and Aerial Imagery." [Online]. Available: https://arxiv.org/abs/2201.11192

[7] B. Weinstein, "MillionTrees." Accessed: Jul. 08, 2024. [Online]. Available: https://milliontrees.idtrees.org/

[8] E. Kalinicheva, L. Landrieu, C. Mallet, and N. Chehata, "Multi-Layer Modeling of Dense Vegetation from Aerial LiDAR Scans." [Online]. Available: https://arxiv.org/abs/2204.11620

[9] S. Puliti *et al.*, "FOR-instance: a UAV laser scanning benchmark dataset for seman-

tic and instance segmentation of individual trees." [Online]. Available: https://arxiv.org/abs/2309.01279

[10] J. Hu *et al.*, "MDAS: a new multimodal benchmark dataset for remote sensing," *Earth System Science Data*, vol. 15, no. 1, pp. 113–131, 2023, doi: 10.5194/essd-15-113-2023.

[11] B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White, "Individual Tree-Crown Detection in RGB Imagery Using Semi-Supervised Deep Learning Neural Networks," *Remote Sensing*, vol. 11, no. 11, 2019, doi: 10.3390/rs11111309.

[12] Institut national de l'information géographique et forestière (IGN), "BD ORTHO." [Online]. Available: https://geoservices.ign.fr/bdortho

[13] Beeldmateriaal Nederland, "Luchtfoto's (Aerial Photographs)." [Online]. Available: https://www.beeldmateriaal.nl/luchtfotos

[14] Actueel Hoogtebestand Nederland, "AHN4 - Actual Height Model of the Netherlands." [Online]. Available: https://www.ahn.nl/

[15] Institut national de l'information géographique et forestière (IGN), "LiDAR HD." [Online]. Available: https://geoservices.ign.fr/lidarhd

[16] Gemeente Amsterdam, "Bomenbestand Amsterdam (Amsterdam Tree Dataset)." [Online]. Available: https://maps.amsterdam.nl/open_geodata/?k=505

[17] Bordeaux Métropole, "Patrimoine arboré de Bordeaux Métropole (Tree Heritage of Bordeaux Metropole)." [Online]. Available: https://opendata.bordeaux-metropole.fr/explore/dataset/ec_arbre_p/information/?disjunctive.insee

[18] Coöperatief Boomregister U.A., "Boom Register (Tree Register)." [Online]. Available: https://boomregister.nl/

[19] R. Jena *et al.*, "Beyond mAP: Towards better evaluation of instance segmentation," *CVPR*, 2023.