UNIVERSITAT DE BARCELONA

FUNDAMENTALS OF DATA SCIENCE MASTER'S THESIS

Man-made Structures Detection from Space

Author:
Peter WEBER

Supervisor:
Dr. Jordi VITRIA

A thesis submitted in partial fulfillment of the requirements for the degree of MSc in Fundamentals of Data Science

in the

Facultat de Matemàtiques i Informàtica

June 17, 2019

UNIVERSITAT DE BARCELONA

Abstract

Facultat de Matemàtiques i Informàtica

Man-made Structures Detection from Space

by Peter WEBER

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Acknowledgements

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

Report structure:

1. Chapter 1: Introduction
    - Problem introduction, satellites, Satellogic, motivation
    - Previous work, literature
2. Chapter 2: Building datasets
    - Existing datasets
    - Google Maps
    - USGS, land cover
3. Chapter 3: Gist approach
    - Features
    - Model
    - Results
4. Chapter 4: DL approach
    - Features
    - Model
    - Results
5. Chapter 5: Final results
    - Results
    - Cost
    - Environment impact
6. Chapter 6: Conclusion

Contents

Chapter 1

Building Datasets

In this chapter, we will give an overview of existing (labelled) aerial imagery datasets and outline the reasons why none of them is suitable for our investigation. Following this discussion, we will describe two approaches for obtaining our own labelled dataset.

1.1 Requirements and Considerations

Before we go into the presentation of existing labelled datasets we discuss the requirements that the dataset needs to fulfill in order to serve for the investigation in this thesis project. As a refresher, we want to detect human impact on aerial images and determine the dependency on resolution per pixel of a chosen evaluation metric. Ideally, the range for the resolutions should scale from a few tens of centimeters to a few tens of meters, whereas the images with low resolution can be generated from the high resolution images by downsampling. Having in mind previous arguments, we mainly need to consider three aspects.

First, we need to have imagery data with labels that can be used to clearly distinguish between existing and non-existing human impact, respectively. This impact might be classified pixel wise, or as binary classification for the entire image, or as multi-class classification that can be translated into binary labelling. Second, we need a balanced dataset of approximately the same number of images for both labels, and variations of the images as large as possible with respect to different terrains. Third, the images need to have a resolution per pixel which is equal or better than 1m. Also, the height and width of the images should measure at least 500×500 pixels, so that one has enough room for downsampling.

1.2 Existing Datasets

In table 1.1 we have summarized the most relevant remote sensing datasets with ground truth labels, that can be found in literature. The table lists the name of the dataset together with the bibliographic reference. It also details the data source for the images. Further it contains a description about the number of images, the resolution of the images, the size (in pixel) of the images where images are squared, and the number of categories.

The datasets were collected using different publicly available data sources. These range from pure low resolution satellite imagery (Sentinel-2) to high-resolution images taken with an aircraft (USGS) to a mix of different image sources (Google Earth).

The satellite images have a resolution of equal or larger than 10 m and they are collected with the Sentinel-2 satellites of the European Earth observation program Copernicus. Although the datasets from this source (BigEarthNet and EuroSat) are

comparatively large, they do not suffice for our purpose, because the resolution is not good enough and the images are too small.

name	source	images	resolution (m)	size (pixel)	categories
BigEarthNet [5]	Sentinel-2	590,326	10, 20, 60	120, 60, 20	~ 50
EuroSAT [4]	Sentinel-2	27,000	10	64	10
UCMerced [7]	USGS	2100	0.3	256	21
DeepSat [1]	USGS	405,000	1	28	6
AID [6]	Google Earth	10,000	0.5 - 8	600	30
PatternNet [8]	Google Earth	30,400	0.06 - 4.69	256	38

TABLE 1.1: Publicly available remote sensing datasets with labels.

The USGS National Map Urban Area Imagery collection (*see link*) was utilized to collect remote sensing datasets in the two works UCMerced and DeepSat, where the former is the dataset that comes closest to our requirements. It has 21 categories of which only 2 belong to images without human impact, while the other 19 show human impact. The DeepSat dataset unfortunately consists of image patches which are only 28×28 large, so that we aren't able to study these images as a function of resolution.

The datasets using Google Earth as data source are collected using either the Google Earth or the Google Maps API. These images vary in resolution as well as in their original data provider since Google accesses several data sources. Both datasets, the AID and the PatternNet dataset, have about 30 categories with several images in each category. Here, different categories have different resolutions per pixel, and again most of the categories relate to urban areas so that we do not have sufficient images without human impact. Even the categories that in principle should not show human influence contain images that break this rule.

Overall, the main issue with these datasets stems from the fact that non of them was collected with the purpose to analyze the human footprint and therefore they are very unbalanced, and do not contain sufficient variety of images for the classes without human influence. Therefore, we decided to collect and label images by ourselfes. In our first approach we used the Google Maps API, and in our final approach we used datasets from the USGS Aerial Imagery collection.

1.3 Google Maps

Google has a public API that allows for querying images from their service Google Maps. In its most basic form, the API accepts as input parameters a lattitude and longitude, a zoom, and the number of pixels to return. Given this set of parameters one can calculate the resolution per pixel (see [3]), which is given by

$$\frac{meter}{pixel} = \frac{156543.03392 \cdot cos(\frac{lattitude \cdot \pi}{180})}{2^{zoom}}. \tag{1.1}$$

1.4 USGS, Land Cover

1.4.1 Getting the Data

To be able to construct a balanced and representative dataset we first decided to focus on images of the United States, which allows for a large variety of different

terrains. We then used as data source the Aerial Imagery datasets from USGS Earthexplorer (*see link*) which we combined with information about Land Cover and Land Use available from the USGS Land Cover Viewer (*see link*).

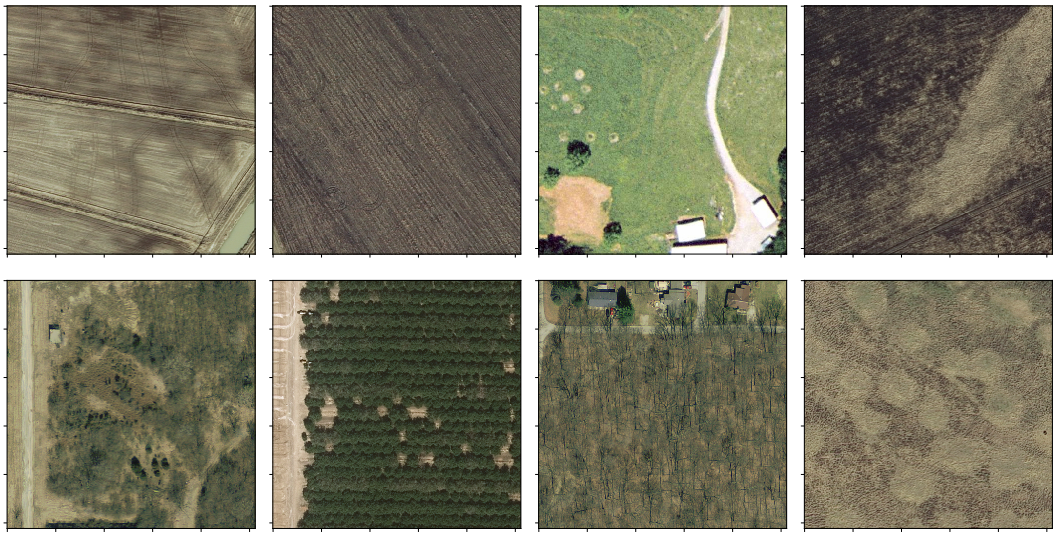

FIGURE 1.1: **Example images of category Agricultue.** All images in this figure show clear signs of human impact. The images have a size of 512 × 512 pixels and a resolution of 0.3m per pixel.

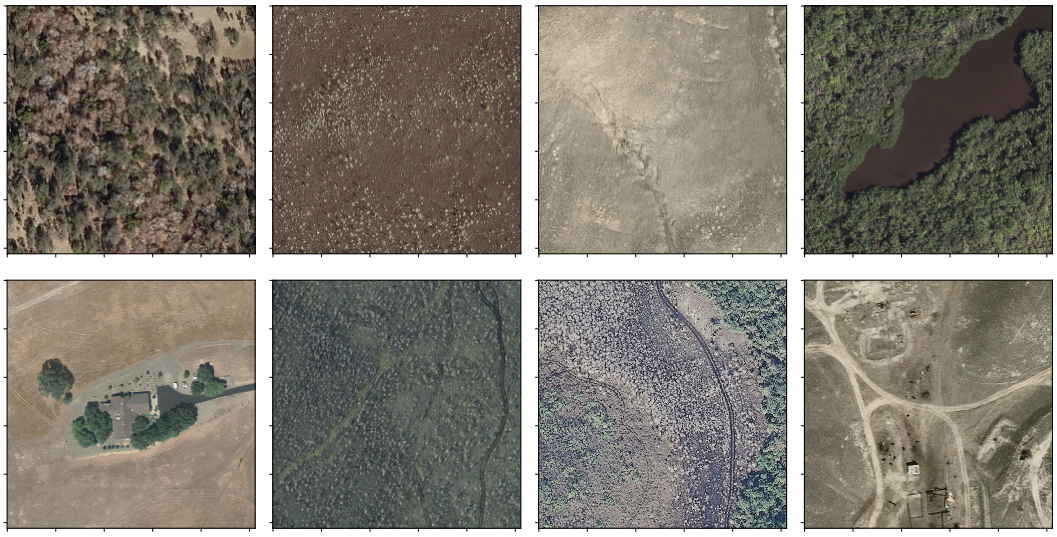

FIGURE 1.2: **Example images of category Shrubland-grassland.** The images in the first row do not contain any human influence, while the images in the second row show influence by humans. The images in this figure have a size of 512 × 512 pixels and a resolution of 0.3m per pixel.

When looking for images we excluded cities and highly developed urban areas, and instead focussed on unpopulated areas. Specifically, we limited our image search to the four Land Use categories Agriculture, Shrubland-Grassland, Semi-Desert, Forest-Woodland that can be found in the USGS Land Cover Viewer. Note

FIGURE 1.3: **Example images of category Forest-woodland.** The images in the first row do not contain any human influence, while the images in the second row show influence by humans. The images in this figure have a size of 512×512 pixels and a resolution of 0.3m per pixel.

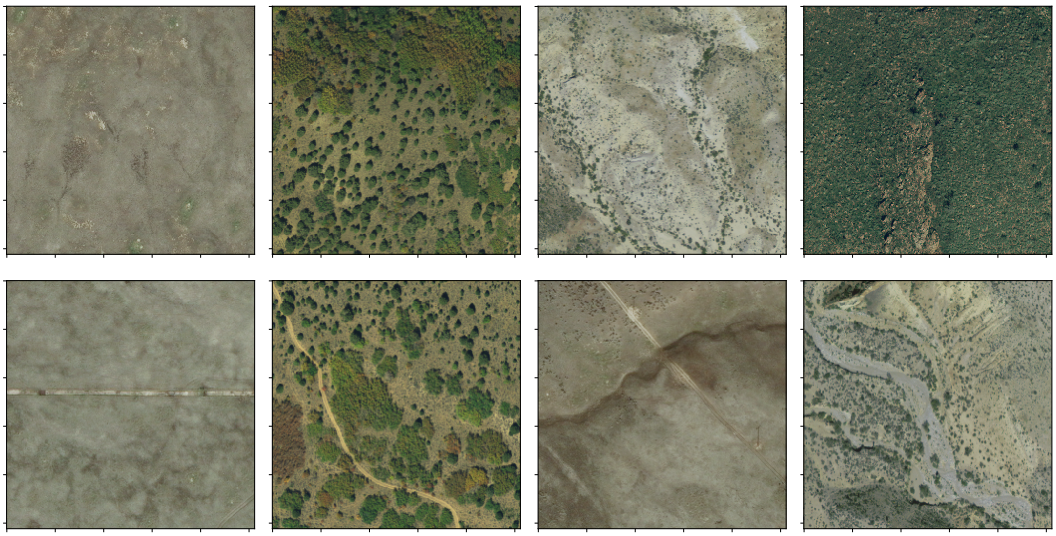

FIGURE 1.4: **Example images of category Semi-desert.** The images in the first row do not contain any human influence, while the images in the second row show influence by humans. The images in this figure have a size of 512×512 pixels and a resolution of 0.3m per pixel.

that these categories served as a rough geographic reference to pin down geolocations of interest, in order to guarantee a dataset with a good variety of different terrains. We also found that it is harder to find images without human, which is why we selected many images from national parks. However, within a given area/terrain we always tried to have images with and without impact.

Once an area was pointed out as a region of interest, we located it on USGS Earthexplorer and downloaded images from that area. In particular, we constructed two datasets with $0.3m$ and $1m$ resolution, respectively. The former was taken from the

category High Resolution Orthoimagery and the latter from the category National Agriculture Imagery Program (NAIP). Note that the images in these categories usually have a height and a width of several thousand pixels, and hence oocupy a few hundreds of Megabytes of disk space. We cropped smaller images out of the raw images, which will be discussed in more detail in the following section. Overall, we downloaded about 100 images for each dataset.

1.4.2 Data Processing and Labeling

Our data processing pipeline consists of the following steps:

- Download large raw images
- Crop images of size 512×512 pixels
- Label images with either zero (no human impact), one (minimal human impact), two (obvious human impact)
- Degrade images, i.e. reduce number of pixels and thereby resolution per pixel

Let us discuss each of these steps in more detail. Every raw image was processed, whereas the processed images of size 512×512 were saved in a folder named by its category. Note that every raw image resulted in approximately $100 - 150$ processed images, so that we ended up with more than 10,000 images for each dataset.

Within each category of the processed images we labelled a selected portion of the images, by moving them into the folder with the respective label name. The folder structure we used is the following, where pointy brackets '<parameter>' indicate a parameter and 'etc' stands for the three label folders.

```
{raw-images-}usgs-<pixels>-res<resolution>m
├── semi-desert
│   ├── label-0
│   ├── label-1
│   └── label-2
├── agriculture
│   └── label-2
├── shrubland-grassland
│   ├── label-0
│   ├── label-1
│   └── label-2
└── semi-desert
    ├── label-0
    ├── label-1
    └── label-2
```

When labelling we sticked to the following rules. First, we classified images with no human impact at all into the class with label zero, while we classified images with very clear human influence into the class with label two. Ambigious images, i.e. images with minimal human trace were classified into label one. Second, we've put major effort into creating datasets that contain images of similar texture spread across all classes. If we for example classified a set of images of a certain forest type into the class with label zero we classified another set of images with a similar forest type, but containing a building or a street, into the class with label two. The same applies for images in class with label one when they contain e.g. a small walking path. We followed the latter rule for all categories except Agriculture. The Agriculture

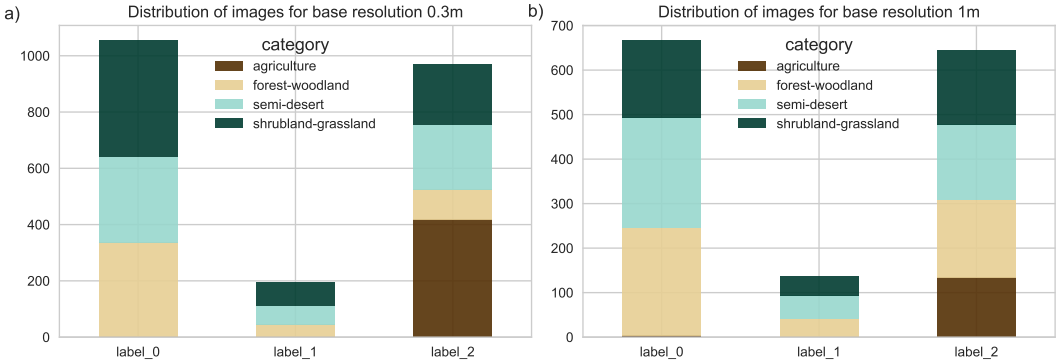

FIGURE 1.5: **Number of images per category and label.** (a) Distribution of images for dataset with resolution of 0.3m per pixel. (b) Distribution of images for dataset with resolution of 1m per pixel.

images all show human influence, and were therefore all classified with label two. By sticking to these rules, we are able to guarantee that the algorithm learns features that relate to human impact, and not to image artefacts such as color or texture.

In Figures 1.1 - 1.4 we display sample images for each of the four categories Agriculture, Shrubland-grassland, Forest-woodland and Semi-desert. These images belong to the dataset which has a resolution per pixel of 0.3m. Note that in Figs. 1.2 - 1.4 the first row represents images of label zero and the second row shows images that belong to label two. As mentioned above, the images in Fig. 1.1 all contain human influence.

The distribution of categories and labels is shown in Fig. 1.5. Overall, for the 0.3m dataset we classified about 2200 images, and for the 1m dataset we classified about 1450 images. During classification our main goal consisted in creating a balanced dataset between labels zero and two. The minority of images, roughly 10% of all classified images were assigned to label 1. These images were used at random to investigate the behaviour of the Machine Learning classifier, which is discussed in chapter **??** VERIFY.

The last step of the data processing pipeline consisted in downsampling the processed and labelled images, to obtain images with a lower resolution per pixel. We used a Lanczos filter [2] for the sampling, which is based on a sinusoidal kernel. In Fig. 1.6) we show a few selected resolutions for an example image from the agriculture category. Note that here we only schematically depict an example in order to illustrate the process. However, in our Machine Learning pipeline the images are downsampled on the fly and the result of this process is not stored on disk.

For this particular image one can observe how certain image features disappear as the image quality is decreased. Above a resolution of around 3m per pixel one is not able anymore to identify the building close to the right corner of the image. The texture of the track that leads up to the building is blurred above a resolution of around 4m per pixel. This shows how different elements in an image are not recognizable anymore once the resolution is worse than their characteristic resolution.

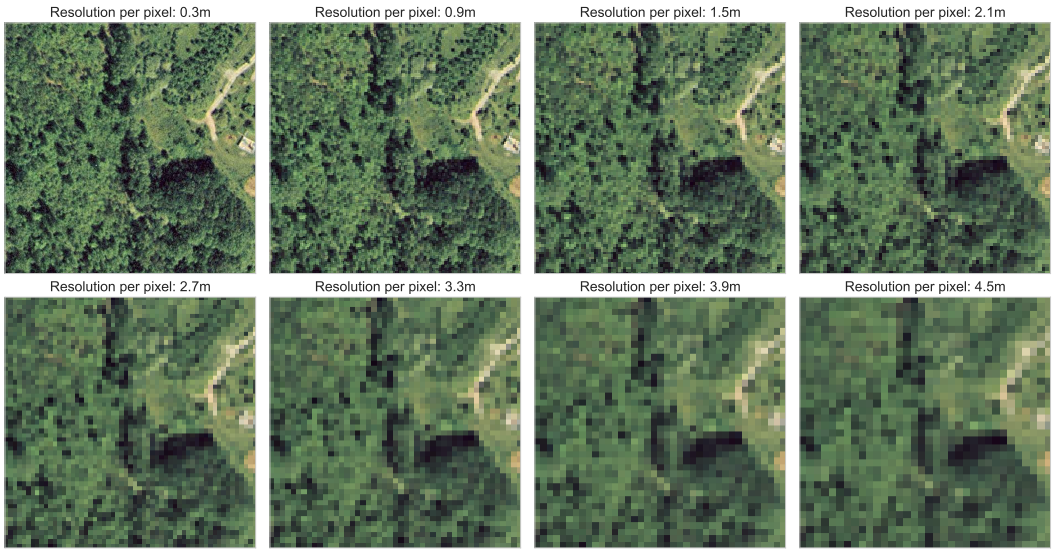

FIGURE 1.6: **Example of image downsampling**. The upper left image has the base resolution, 0.3m per pixel, and a size of 512×512 pixels whereas the lower right image has the worst resolution, 4.5m per pixel, and a size of 34×34 pixels. All intermediate images are downsampled by a factor corresponding to the resolution of the actual image divided by the base resolution. For instance, for the lower right image it is 15.

Appendix A

Frequently Asked Questions

A.1 How do I change the colors of links?

The color of links can be changed to your liking using:

`\hypersetup{urlcolor=red}`, or
`\hypersetup{citecolor=green}`, or
`\hypersetup{allcolor=blue}`.

If you want to completely hide the links, you can use:

`\hypersetup{allcolors=.}`, or even better:
`\hypersetup{hidelinks}`.

If you want to have obvious links in the PDF but not the printed text, use:

`\hypersetup{colorlinks=false}`.

Bibliography

[1] Saikat Basu et al. “DeepSat - A Learning framework for Satellite Imagery”. In: *CoRR* abs/1509.03602 (2015). arXiv: 1509.03602. URL: http://arxiv.org/abs/1509.03602.

[2] Claude E. Duchon. “Lanczos Filtering in One and Two Dimensions”. In: *Journal of Applied Meteorology* 18, 1016-1022 (1979). URL: https://icess.eri.ucsb.edu/gem/Duchon_1979_JAM_Lanczos.pdf.

[3] *Google Maps: Zoom to meters*. https://groups.google.com/forum/!topic/google-maps-js-api-v3/hDRO4oHVSeM.

[4] Patrick Helber et al. “EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification”. In: *CoRR* abs/1709.00029 (2017). arXiv: 1709.00029. URL: http://arxiv.org/abs/1709.00029.

[5] Gencer Sumbul et al. “BigEarthNet: A Large-Scale Benchmark Archive For Remote Sensing Image Understanding”. In: *CoRR* abs/1902.06148 (2019). arXiv: 1902.06148. URL: http://arxiv.org/abs/1902.06148.

[6] Gui-Song Xia et al. “AID: A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification”. In: *CoRR* abs/1608.05167 (2016). arXiv: 1608.05167. URL: http://arxiv.org/abs/1608.05167.

[7] Yi Yang and Shawn Newsam. “Bag-of-visual-words and spatial extensions for land-use classification”. In: Jan. 2010, pp. 270–279. DOI: 10.1145/1869790.1869829.

[8] Weixun Zhou et al. “PatternNet: A Benchmark Dataset for Performance Evaluation of Remote Sensing Image Retrieval”. In: *CoRR* abs/1706.03424 (2017). arXiv: 1706.03424. URL: http://arxiv.org/abs/1706.03424.