

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 15, 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

Abstract	iii
Acknowledgements	v
1 Building Datasets	5
1.1 Requirements and Considerations	5
1.2 Existing Datasets	5
1.3 Google Maps	6
1.4 USGS, Land Cover	6
1.4.1 Getting the Data	6
1.4.2 Data Processing and Labeling	7
A Frequently Asked Questions	11
A.1 How do I change the colors of links?	11
Bibliography	13

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 already 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 of a chosen evaluation metric on resolution per pixel. 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. **TODO: Explain degradation process.** 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 on the images, 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 as large as possible variation of the images 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 degradation.

1.2 Existing Datasets

In literature, there can be found several remote sensing datasets with ground truth labels. We have listed the most relevant ones in table 1.1. The table first lists the name of the dataset together with the bibliographic reference. It also details the data source for the images. Further it contains 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

name	source	images	resolution (m)	size (pixel)	categories
BigEarthNet [3]	Sentinel-2	590,326	10, 20, 60	120, 60, 20	~ 50
EuroSAT [2]	Sentinel-2	27,000	10	64	10
UCMerced [5]	USGS	2100	0.3	256	21
DeepSat [1]	USGS	405,000	1	28	6
AID [4]	Google Earth	10,000	0.5 - 8	600	30
PatternNet [6]	Google Earth	30,400	0.06 - 4.69	256	38

TABLE 1.1: Publicly available remote sensing datasets with labels.

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.

The USGS National Map Urban Area Imagery collection ([see link](#)) was utilized to collect a remote sensing dataset by 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 none of them was collected with the purpose to analyze human impact and therefore they are very unbalanced, and do not contain sufficient variety of images for the label no-human impact. Therefore, we decided to collect and label images by ourselves. In our first approach we used the Google Maps API.

1.3 Google Maps

...

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 Earth-explorer ([see link](#)) which we combined with information about Land Cover and Land Use available from the USGS Land Cover Viewer ([see link](#)).

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 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 Earth-explorer 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 occupy 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
- Labelling the images with either 0 (no human impact) or 1 (human impact)
- Degrading the images, i.e. reducing the number of pixels

TODO: mention all the scripts/notebooks for the respective parts

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. **SHOW FOLDER STRUCTURE SCHEMATICALLY**. When labelling we mainly followed two rules. First, we did not classify images that were ambiguous, i.e. images that showed minimal human impact such as a small walking path. Second, we put a lot of effort in creating datasets that contain images of similar texture spread across both classes. If we for example classified a set of images of a certain type of forest into the class with label 0 we classified another set of images with a similar type of forest, but containing a building or a street, into the class with label 1. We followed the latter rule for all categories except Agriculture. The Agriculture images all show human influence. By sticking to these rules, we are able to guarantee that the algorithm learns features that relate purely to the human impact.

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



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

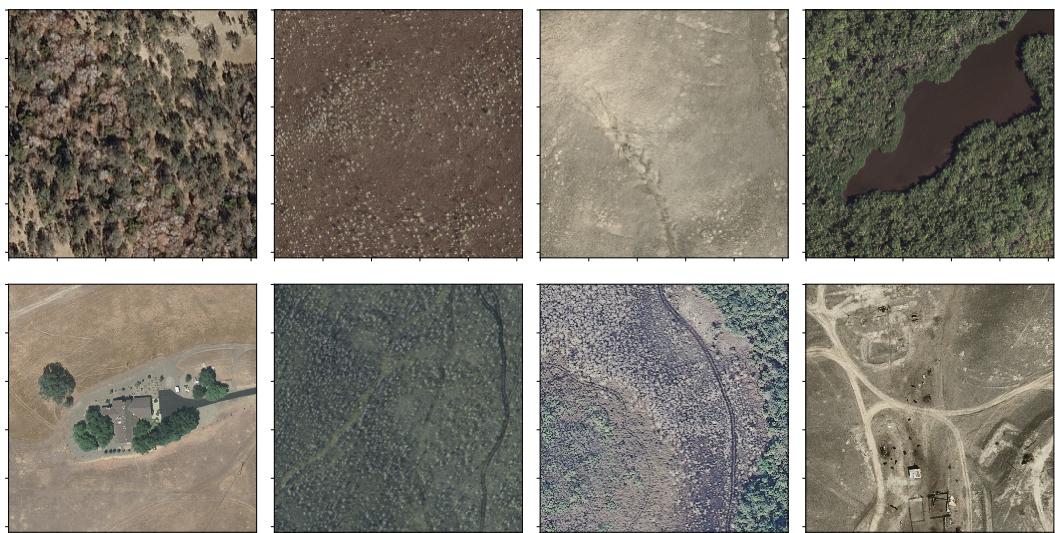


FIGURE 1.2: **Example images of category Shrubland-grassland.** The images in the first row do not contain any human influence, while the images in the second row show influence by humans. The images in this figure have a size of 512×512 pixels and a resolution of 0.3m per pixel.

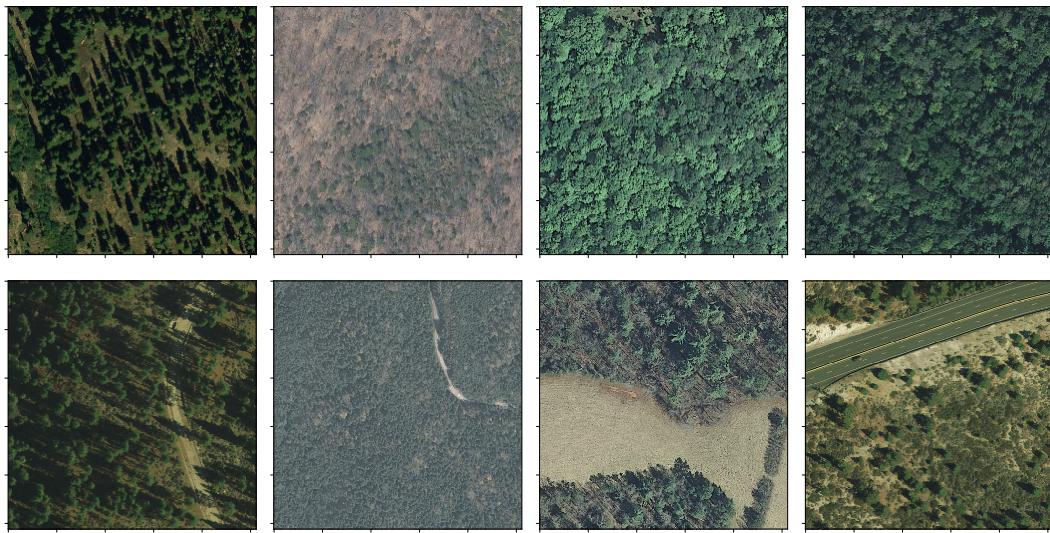


FIGURE 1.3: **Example images of category Forest-woodland.** The images in the first row do not contain any human influence, while the images in the second row show influence by humans. The images in this figure have a size of 512×512 pixels and a resolution of 0.3m per pixel.



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.

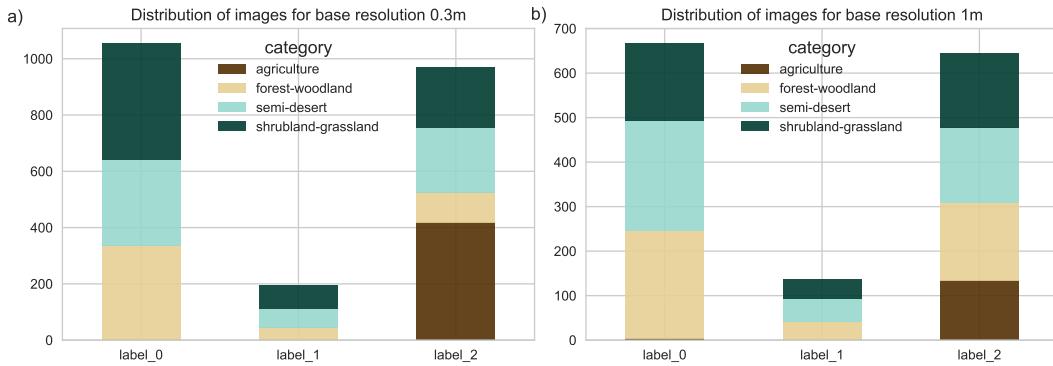


FIGURE 1.5: Image statistics.

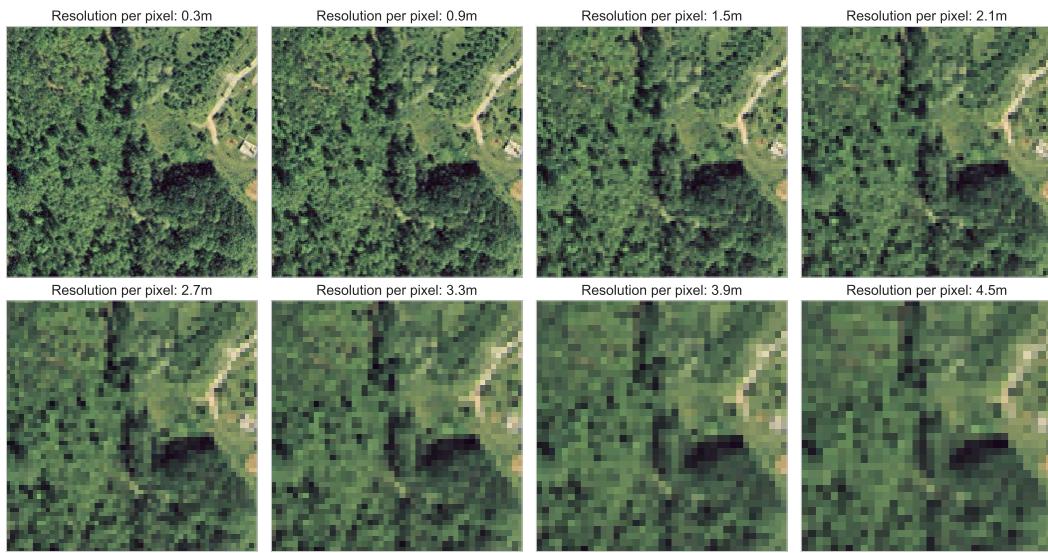


FIGURE 1.6: Demo image resampling.

belong to the dataset with 0.3m resolution per pixel. Note that in the Figs. ?? - 1.4 the first row represents images of label 0 and the second row shows images that belong to label 1. As mentioned above, the images in Fig. 1.1 all contain human influence. Overall, we classified

- Image statistics
- Image degradation

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](https://arxiv.org/abs/1509.03602). URL: <http://arxiv.org/abs/1509.03602>.
- [2] 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](https://arxiv.org/abs/1709.00029). URL: <http://arxiv.org/abs/1709.00029>.
- [3] Gencer Sumbul et al. "BigEarthNet: A Large-Scale Benchmark Archive For Remote Sensing Image Understanding". In: *CoRR* abs/1902.06148 (2019). arXiv: [1902.06148](https://arxiv.org/abs/1902.06148). URL: <http://arxiv.org/abs/1902.06148>.
- [4] 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](https://arxiv.org/abs/1608.05167). URL: <http://arxiv.org/abs/1608.05167>.
- [5] 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](https://doi.org/10.1145/1869790.1869829).
- [6] 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](https://arxiv.org/abs/1706.03424). URL: <http://arxiv.org/abs/1706.03424>.