

## *Arealstatistik* Deep Learning (ADELE)

Organisation:	Federal statistics office, Switzerland
Author(s):	Claudio Facchinetti
Date:	19.06.2020
Version:	1.0

### 1. Background and why and how this study was initiated

The FSO's land use (LU) and coverage (LC) statistics are an invaluable tool for long-term spatial observation with a survey period that has been gradually reduced from 12 years (in 1979) to 6 years today. At present, internal resources are almost entirely allocated to visual interpretation for classification into 73 categories (27 for land coverage and 46 for land use), at the expense of other activities. In light of this, certain artificial intelligence algorithms could help though by facilitating land use and cover classification and by improving change detection. The objective of the pilot project "Arealstatistik Deep Learning" (ADELE) is to gain a better understanding of this field in preparation for the gradual development of a computer prototype. The results and knowledge obtained so far demonstrated the innovation potential for the FSO in using artificial intelligence to process images.

This project is led by the Geoinformation team of the Territory and environment division. It is part of the FSO's experimental statistics produced using new methods and/or new data sources, in line with the data innovation strategy and the Swiss Confederation's multi-annual programme for federal statistics.

### 2. Data

#### 2.1 Input Data (short description)

See diagram in §3.2 for details where input data is used in the pipeline.

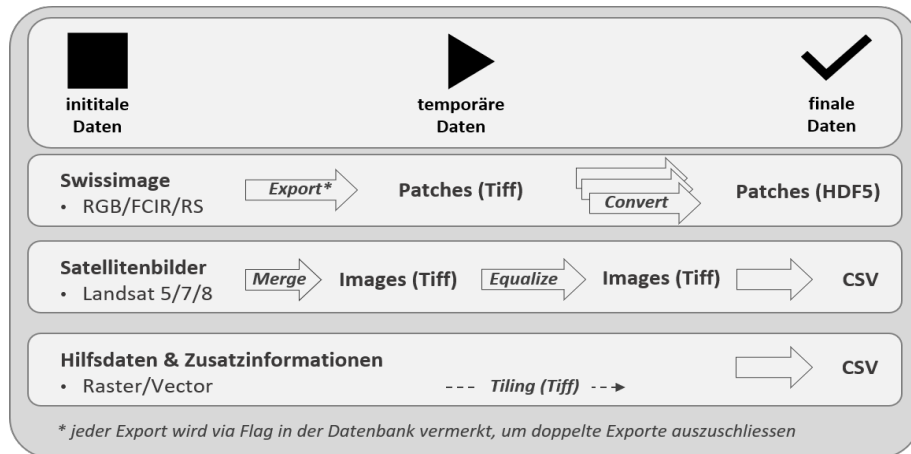
- SWISSIMAGE aerial images (25cm RGB 3 x 8-bit)
- Landsat 8 satellite images
- Digital Elevation Model (DEM)
- Canopy Height Model (CHM)
- Registers (land, agriculture, etc)

#### 2.2 Data Preparation

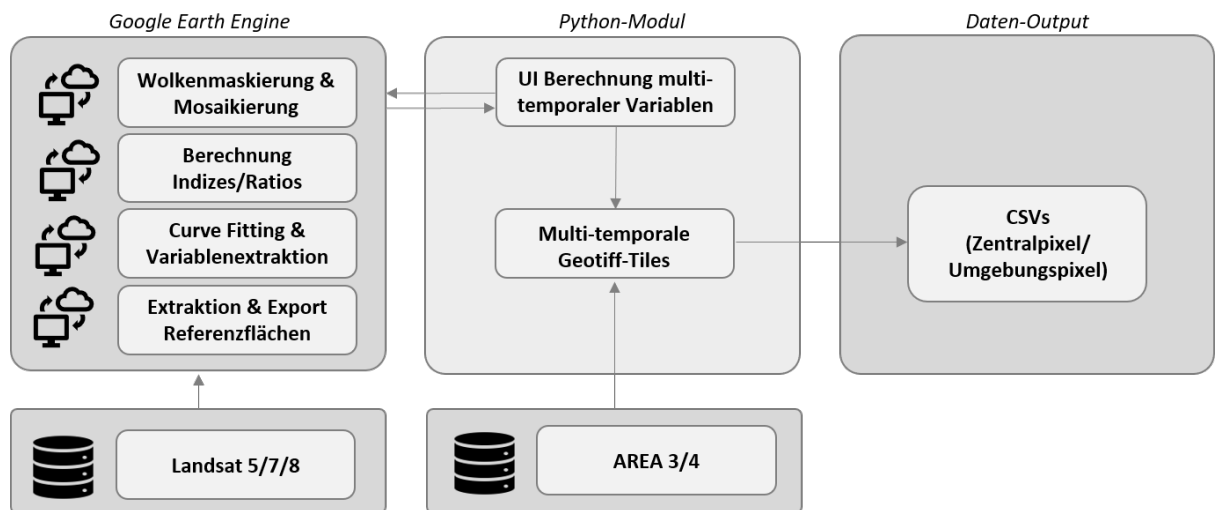
Individual modules can be accessed and parameterized via a configuration file. The modules then link the corresponding initial basic, auxiliary and additional data with the table/database of the area statistics, carry out specific pre-processing steps and generate the final data for the subsequent classification steps. Each process and the respective parameterisation are documented in a log file.

All modules are implemented in the programming language Python, whereby some processing steps concerning the satellite data were carried out in the Google Earth Engine (but this was also done via a Python interface). The respective modules partly generate temporary files, often used several

times for certain processing steps and ready-to-use data and data structures for the CNN/RF modules.

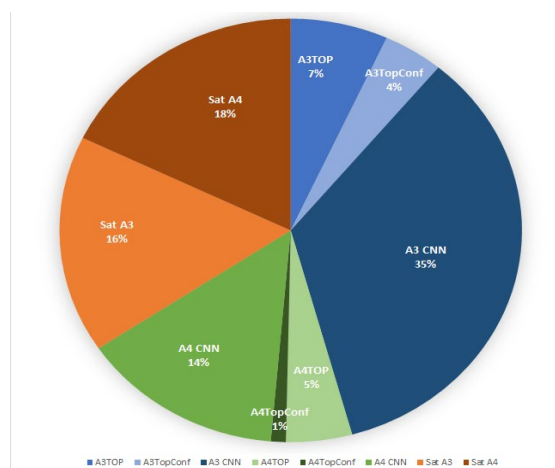


Satellite Time Series are prepared differently, as shown in the diagram below.



### 2.3 Feature Selection

Variables of importance can be extracted from the trained Random Forest models, as shown in the example below.



### 2.4 Output data

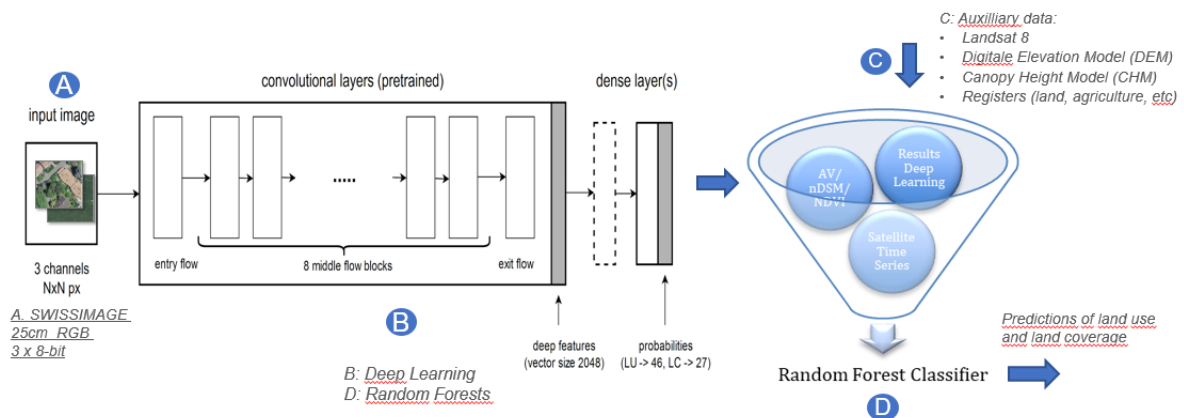
The output of the prediction step consists of land use and land cover classification vectors with probabilities and confidence metrics.

The output of the validation step consist of a large number of standards statistics, mostly derived from confusion matrices.

### 3. Machine Learning Solution

#### 3.1 Models tried

Among the learning methods that have emerged from AI for image recognition, Convolutional Neuronal Networks (CNN – e.g. Deep Learning) are particularly adapted to statistics on land use and cover for which learning data are available in very large quantities. Models such as Xception are pre-trained on ImageNet and freely available, but limited to 3 input channels. Our algorithm is hence extended using a Random forest model for post-classification, able to merge auxiliary data. This approach helps significantly in improving the prediction accuracy.



#### 3.2 Model(s) finally selected and the criterion

The optimal hyperparameters for the CNN Implementation are as follows:

- Up to 10'000 samples per category
- 80-10-10% train/val/test split
- LU: ~ 230'000 images in train set (46 categories)
- LC: ~ 190'000 images in train set (27 categories)
- Tile size: 400px/100m, 300px/75m, **200px/50m**, 100px/25m (25cm per px)
- Extra dense layer after convolutional layers (optimal without)
- Amount of trainable conv. blocks: all to none (optimal: 5-7 middle flow blocks and exit flow)
- Augmentations (LC – horizontal flips; LU – none)

#### 3.3 Hardware used

For the development of the prototype AI, consumer-grade single-user computers were used. The further development of existing CNN networks was done at the FHNW on a Ubuntu/Fedora-Linux Dell Precision Tower 7910, CPU Intel Xeon E5-2630 2.2GHz 10-Core, 32GB RAM, GPU Nvidia GTX 1080Ti 11GB DDR5 with 3584 CUDA Cores. The network to be used for the prototype AI was built under Windows 10 with a consumer grade Dell Precision Tower 5810, CPU Intel Xeon E5-1660 3.0GHz Octacore, 32GB RAM, GPU Nvidia Quadro K5200 8GB DDR5 with 2304 CUDA cores. An Nvidia Quadro K4000 3GB DDR5 with 768 CUDA cores is considered the minimum graphics solution; however, compared to the K5200, two to three times longer training times for DL CNN can be

expected. In return, RF primarily profits from a fast CPU. As the ranger library works with a highly efficient multithreading algorithm, a larger number of main processors (e.g. 16-core instead of Octacore) favors the computation time.

### 3.4 Runtime to train the model

The table below lists estimates for the respective modules of aerial image processing, satellite image processing and processing of auxiliary and additional data.

Modul	Prozess	Prozessierungszeit	Speicherplatz
Modul Luftbilder	Extraktion Referenzflächen	ca. 10 Tage	ca. 5 TB
	(für swissimage FCIR, 50x50m)	ca. 4 Tage	ca. 2.5 TB
	HDF5-Konvertierung (für swissimage FCIR, 50x50m)		
Modul Satellitenbilder	Google Earth Engine	ca. 4 Stunden	-
	(für Landsat)	ca. 10 Tage	ca. 750 GB
	Extraktion Referenzflächen (für Landsat)		
Modul Hilfs- & Zusatzdaten	Extraktion Referenzflächen	ca. 4 Tage	ca. 500 GB
	(für einen Raster-Datensatz 2 x 2 m)		
	Räumliche Verschneidung (für einen Vektordatensatz)	ca. 2 - 8 Stunden	ca. 5 GB

## 4. Results

A feasibility<sup>1</sup> study conducted in 2017 followed by a proof-of-concept in 2019 provided the preliminary results over ~200'000 sample points, summarized as follows:

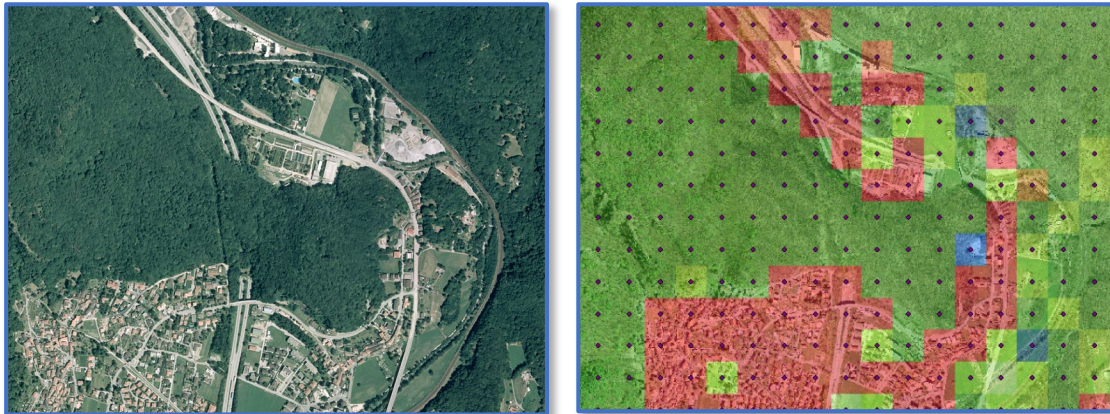
- Land cover (LC):
  - Precision  $\geq 90\%$  for 5 classes out of 27 :
    - › Bodies of water, herbaceous vegetation, trees, glaciers
  - Representing **81%** of all sample points
- Land use (LU)
  - Precision  $\geq 90\%$  for 10 classes out of 46 :

<sup>1</sup> See full report : <https://www.bfs.admin.ch/bfs/de/home/statistiken/raum-umwelt/erhebungen/area.assetdetail.5687737.html>

- › Viticulture, arable land, mountain pastures, forests, lakes
- Representing **44%** of all sample points

A few classes have very low accuracy and tend to lower the global average. This is due to high assymetry in the distribution of points.

An illustration of early results is given below, for an area unknown to the algorithm.



*Left: Aerial image of Toricella-Taverne (Tessin, Switzerland); Right: classified area points*

- Wooded areas
- Agriculture
- Housing and infrastr.
- Improductive areas

*Top levels of the classification nomenclature*

## 5. Code/programming language

The prototype AI is developed under Windows 10. The programming language used is Python in version 3.6. External dependencies (e.g. specific Py modules and libraries) are encapsulated, making the prototype AI work as stand-alone software. For the Deep Learning module "Convolutional Neural Network" (CNN) the Tensorflow environment version 1.10 is used, but controlled via the high-level API Keras (version 2.2.4). For the calculation with Nvidia GPUs, the CUDA framework (CUDA Toolkit 9.0) and the cuDNN framework (cuDNN v7.0) are used to support parallelizable machine learning operations on CUDA cores. For the "Random Forest" (RF) module, the parallelizable and highly efficient algorithm of the library Ranger is used.

## 6. Evolution of this study inside the organisation

We are currently focusing on the validation of the predictions accuracies and the characterization of thresholds on prediction probabilities (i.e. the result reliability self-estimated by the algorithm), above which we can accept results for publication of the statistics.

## 7. Is it a proof of concept or is it already used in production?

The Proof of Concept (pilot project ADELE) provided end of 2019 promising results, helping us to move forward with a concept integrating visual interpretation (by experts) with automatic classification and change detection. The latter is currently under development and will be reported at a later stage.

## 7.1 What is now doable which was not doable before?

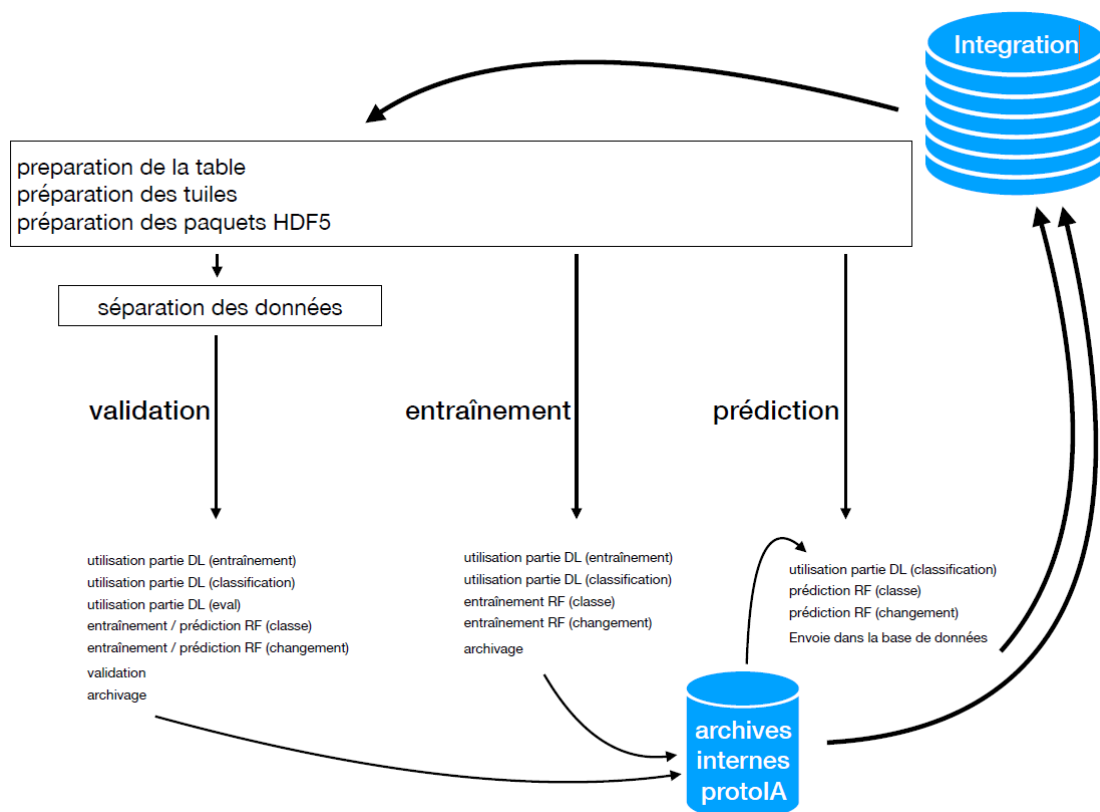
Our prototype (i.e. scripts) can execute data preparation, training, prediction and validation (i.e. production of statistical results out of test data such as accuracies, confusion matrices, etc), in an integrated way.

## 7.2 Is there already a roadmap/service journey available how to implement this?

We are still fine-tuning the algorithm, both for CNN and RF (hyper parameters, input data, etc), but at the same time we need to conceptualize the integration into production.

The illustration below presents the various steps towards a transfer into production that we intend to achieve:

- Validation (Q2 2019 - Q2 2020): fine-tuning of all parameters and measurement of performance, such as (e.g. accuracies, confusion matrices, calibration of prediction probabilities).
- Training (Q3 – Q1 2020): the models trained during validation will be fed with more data, but no change in the parameters, in order to improve them.
- Prediction (Q2 2021 onwards): change detection and classifications will be predicted based on new data, with an acceptance of the results based on prediction probabilities thresholds.



## 7.3 Who are the stakeholders?

Consumer of land use and land coverage statistics across Switzerland and abroad.

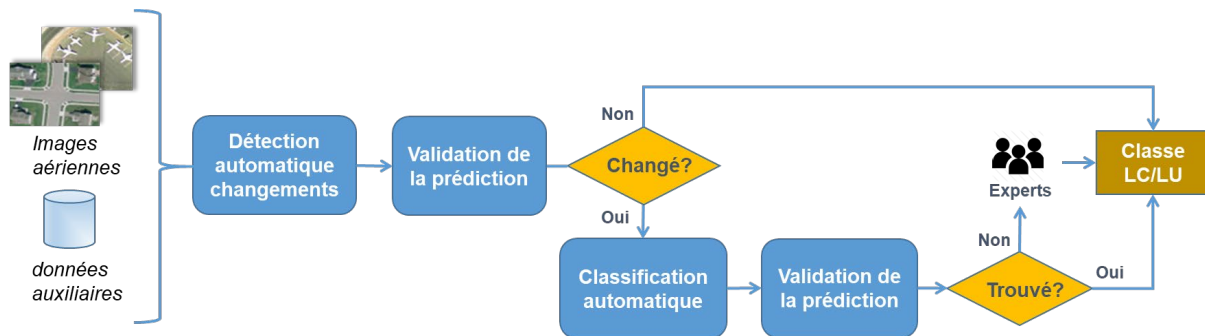
A non-exhaustive list for stakeholders in Switzerland is as follows:

- Federal Administration (ARE, BAFU, swisstopo, BLW)
- Regional statistics (CORSTAT)
- Regional geoinformation centres (KKGEO)
- Regional spatial planning offices (KPK)
- Science & Research (WSL)

A representative group of these stakeholders is being consulted twice a year, since 2019 and until Q3 2021 (transfer into production).

## 7.4 Fall Back

A visual interpretation by human experts is foreseen in case of failure of the automatic classification.

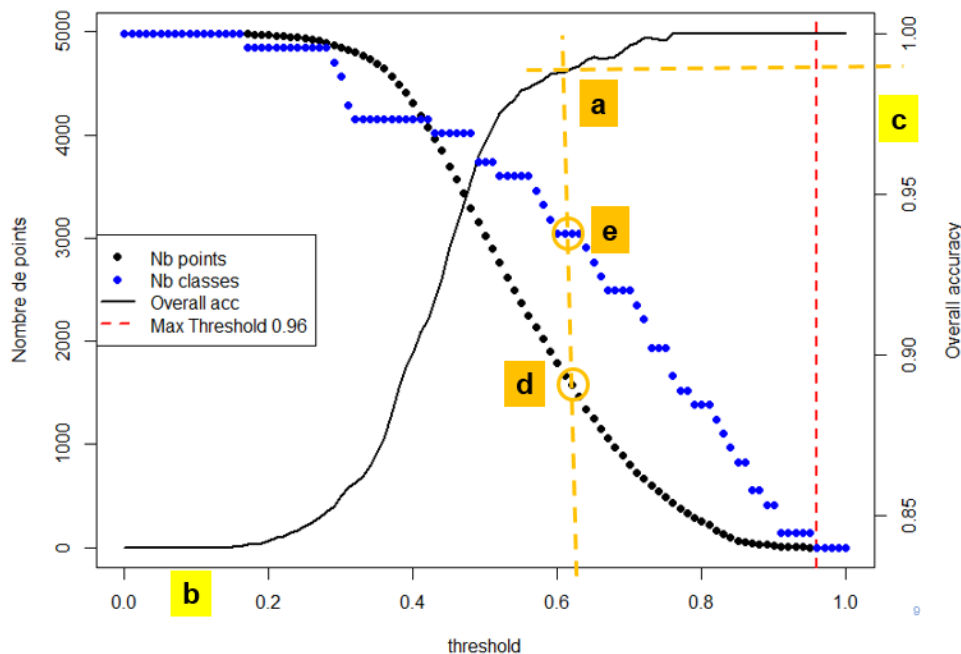


A quality system for regular monitoring and improvement of the automatic classification and change detection is being setup. It will be reported on at a later stage.

## 7.5 Robustness

Validation of the trained model is currently ongoing.

An example of threshold analysis is presented below for a small trial dataset, for which ground truth is available out of the standard visual interpretation process.



Based on the analysis of these 5000 points, if one were to set a prediction probability threshold at 0.65 (axis "b"), meaning that a classification prediction with a self-estimated probability equal or

above 0.65 would be accepted in production, one would expect that 1500 / 5000 (see “d”) of the sample points would not require visual interpretation (with a good reduction in cost/time), but also accept that approximately 3% (100% - 97% - see “c”) could be classified inaccurately and that only a subset of the best performing classes are considered (see “e”).

## 8. Conclusions and lessons learned

This project helps to prepare the team participants for future challenges, providing them with the required knowledge in machine learning methods, in particular the many training parameters and validation criteria of the Deep Learning and Rand Forest algorithms. The exchange with external partners was very beneficial in this respect. The pilot project demonstrated the innovative potential for the FSO in the application of AI in image processing. Future resources and means required can now be properly dimensioned.

We see as improvement potential the following key points:

- Plan well ahead provisioning and storage of the large volume of data to be processed (especially aerial images)
- Validate the benefit of other data sources (e.g. 10m satellite images and topographic model)
- Time adjustment between data is important
- Optimization of CNN by category must be investigated (e.g. vegetation)
- The potential of change detection must be investigated, as already identified in the feasibility study

## 9. Potential organisation risk if ML solution not implemented

Risk management is part of the project management method (internal report).

## 10. Has there been collaboration with other NSIs, universities, etc?

- Fachhochschule Nordwestschweiz, Switzerland (FHNW)
- University of Neuchâtel, Switzerland

## 11. Next Steps

- Investigate the potential improvements identified so far
- Statistical validation of reliability and reproducibility of the predictions
- Description of the new process merging classical methodology and AI (business process modelling)
- Technology Watch: the state of the art of AI is growing rapidly