

Rome. 5-6 December 2022

WORKSHOP ON METHODOLOGIES IN OFFICIAL STATISTICS

A Deep Learning Approach to Land Cover Estimation from Satellite Imagery

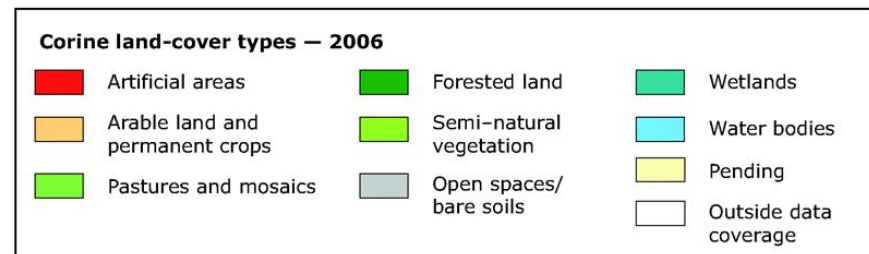
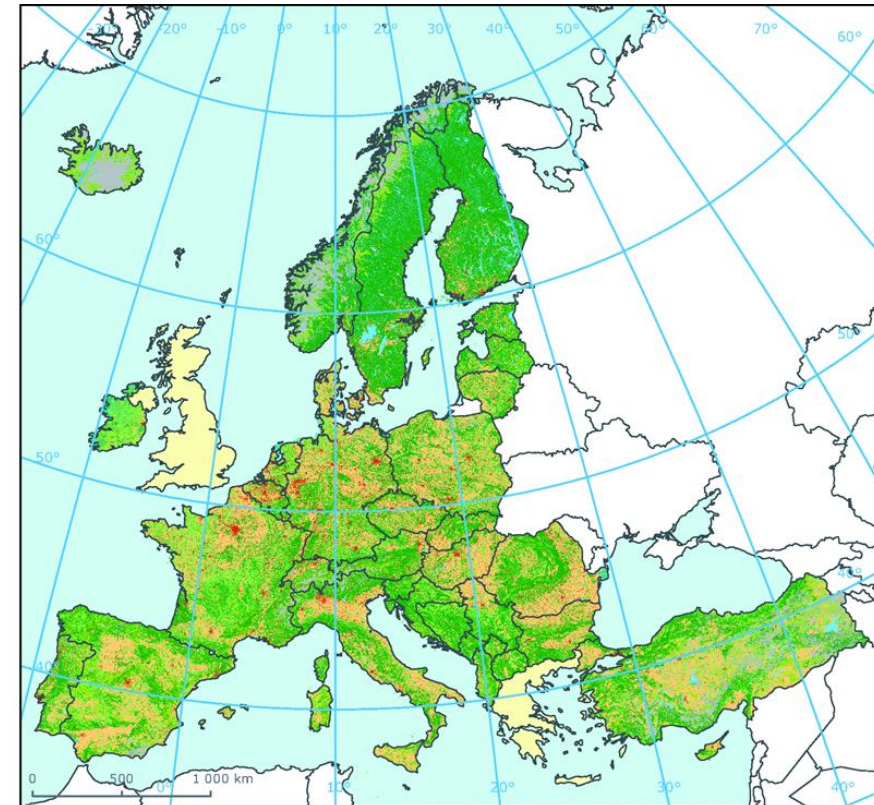
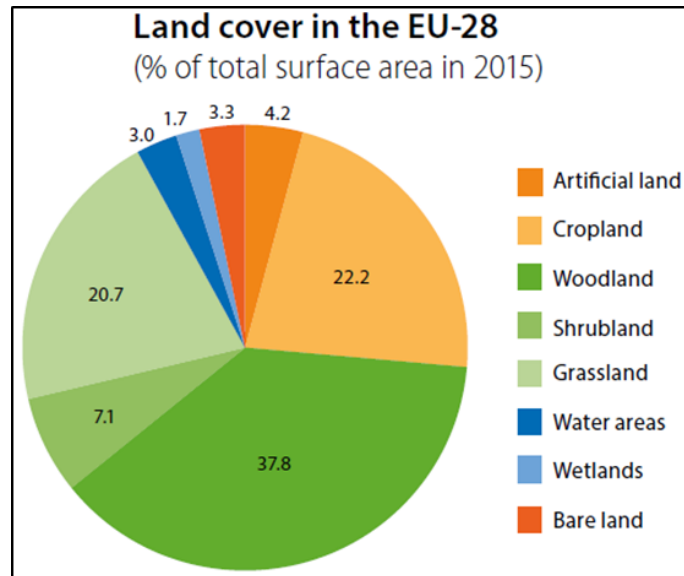
Erika Cerasti, **Fabrizio De Fausti**, Angela Pappagallo, Francesco Pugliese, Diego Zardetto

Istat | DIRECTORATE FOR METHODOLOGY AND STATISTICAL PROCESS DESIGN

Land Cover

Land cover is the observed (bio)physical cover on the **earth's surface**.

An important **aim** of LC studies is to **estimate** the area (or the percentage) of **land occupied** by a certain **category of entities**: forests, crops, herbaceous vegetation, residential areas, rivers, lakes, etc.



Land Cover EU Projects

European flagship LC projects:

- **CORINE** (run by the Copernicus Programme)
cartography (i.e. full-coverage) perspective
- **LUCAS** (managed by Eurostat)
statistical estimation (i.e. sample survey) perspective
 - very costly
 - very complex production pipelines
 - heavy human workload
 - low time frequency

Goal

Goal: given an **input satellite image** of an **area of interest**, we want an automatic system able to:

- classify the territory according to some **LC taxonomy**
- **quantify the area (or proportion)** of territory covered by each LC class

Its main benefits can be:

- to returns very **timely** land cover statistics
- to enable land cover estimation for **sub-regional area** level
- to dramatically **reduce production costs**

THE EUROPEAN COPERNICUS CONSTELLATION PROJECT



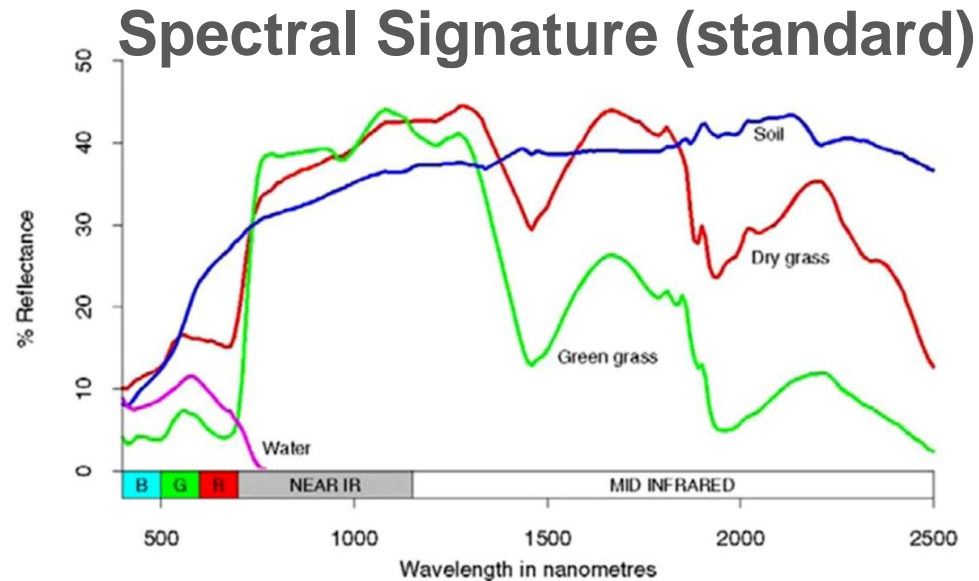
Monitoring of planet in many ecosystem

Constellation of 6 families of satellites

- Sentinel 1: land & water, SAR active radar, Revisit time 12 days
- **Sentinel 2: land & water, MSI 13 bands, Revisit time 5 days**
- Sentinel 3: water
- Sentinel 4: atmosphere
- Sentinel 5: atmosphere
- Sentinel 6: atmosphere

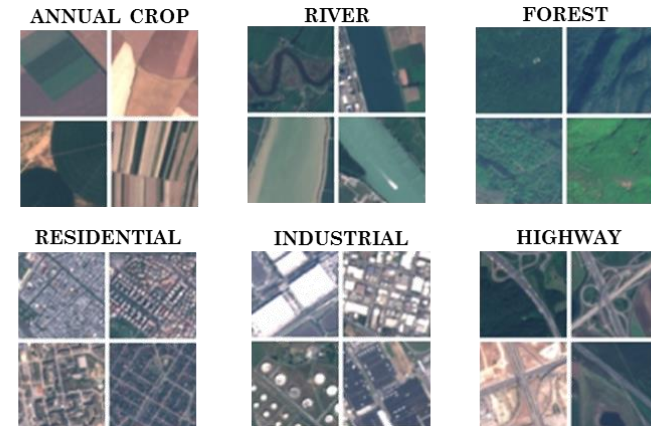
FULL FREE OPEN ACCESS

Automatic approach to land cover



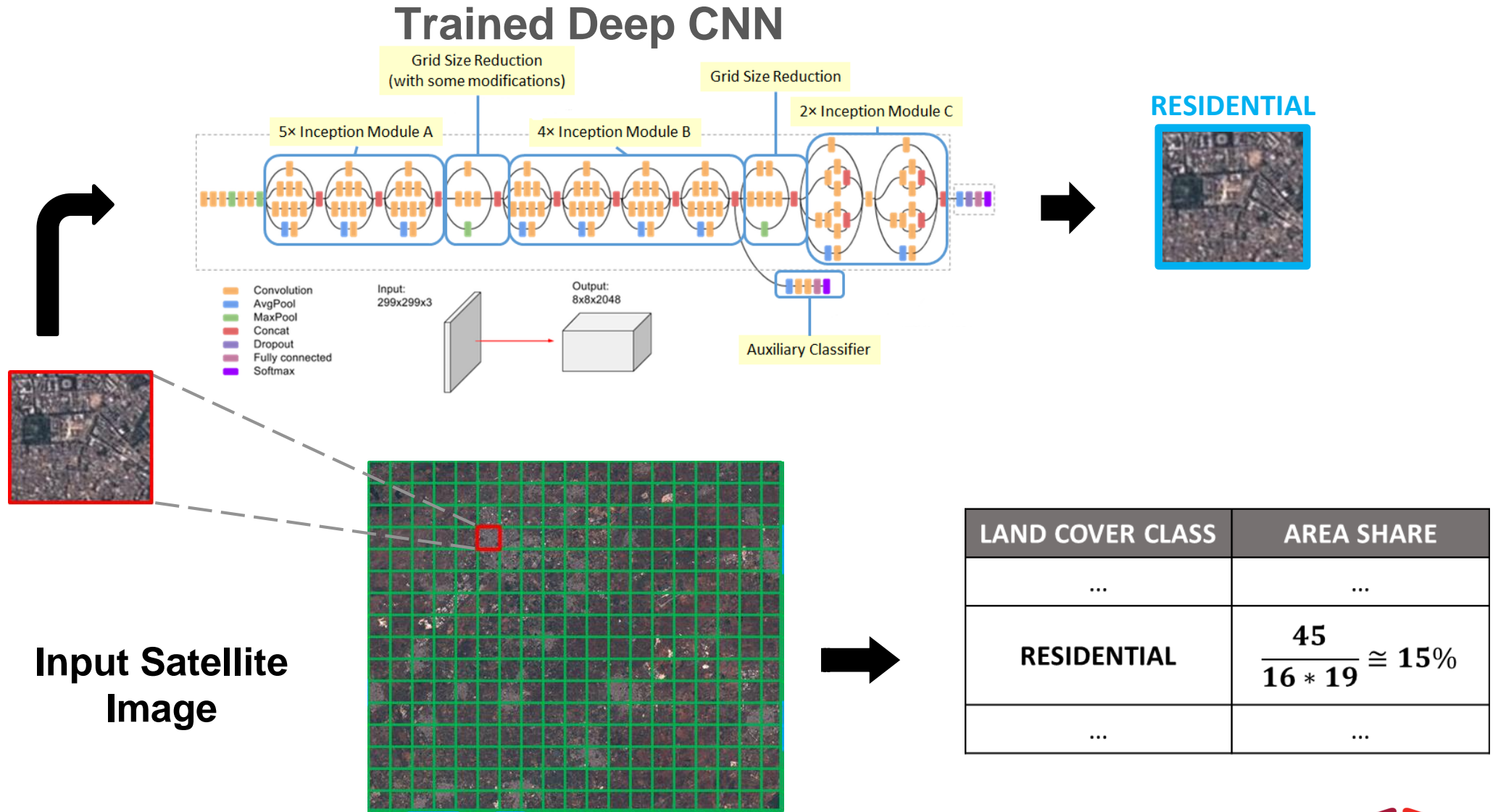
- Different LC classes have different reflectance spectra
- Variation of reflectance with EM frequency can be used to predict LC class
- Decision on each pixel does not depend on neighboring pixels

Computer Vision /Deep Learning



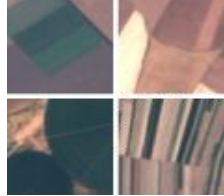
- Different LC classes have different visual/spatial patterns
- Trained ML algo predicts LC class of image pixels based on information from neighboring pixels
- Decision on each pixel depends on the whole sub-image (tile) the pixel belongs to

Automatic approach to land cover (classify and count)



Dataset EuroSAT

ANNUAL CROP



RIVER



FOREST



RESIDENTIAL



INDUSTRIAL



HIGHWAY



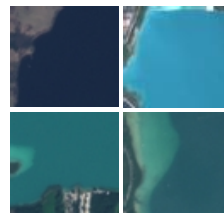
PASTURE



PERMANENT CROP



SEA LAKE

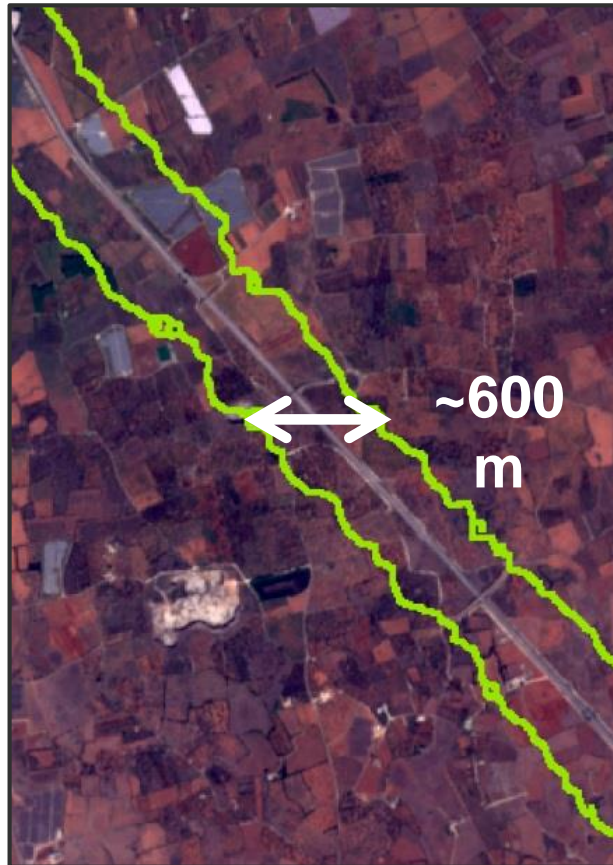


HERBACEOUS VEGETATION



- Available on <https://github.com/pfelber/eurosat>
- Data from Sentinel-2 Multi Spectral Instrument covering 34 European countries
- 27000 geo-referenced labeled images 640x640 meters
- 10 different Land Cover Classes
- 2000-3000 images per class
- Available versions: RGB (10x10 meter of resolution) and multi-spectral (13 spectral bands, 16-bit)

Overestimate of River e Highway classes



HIGHWAY

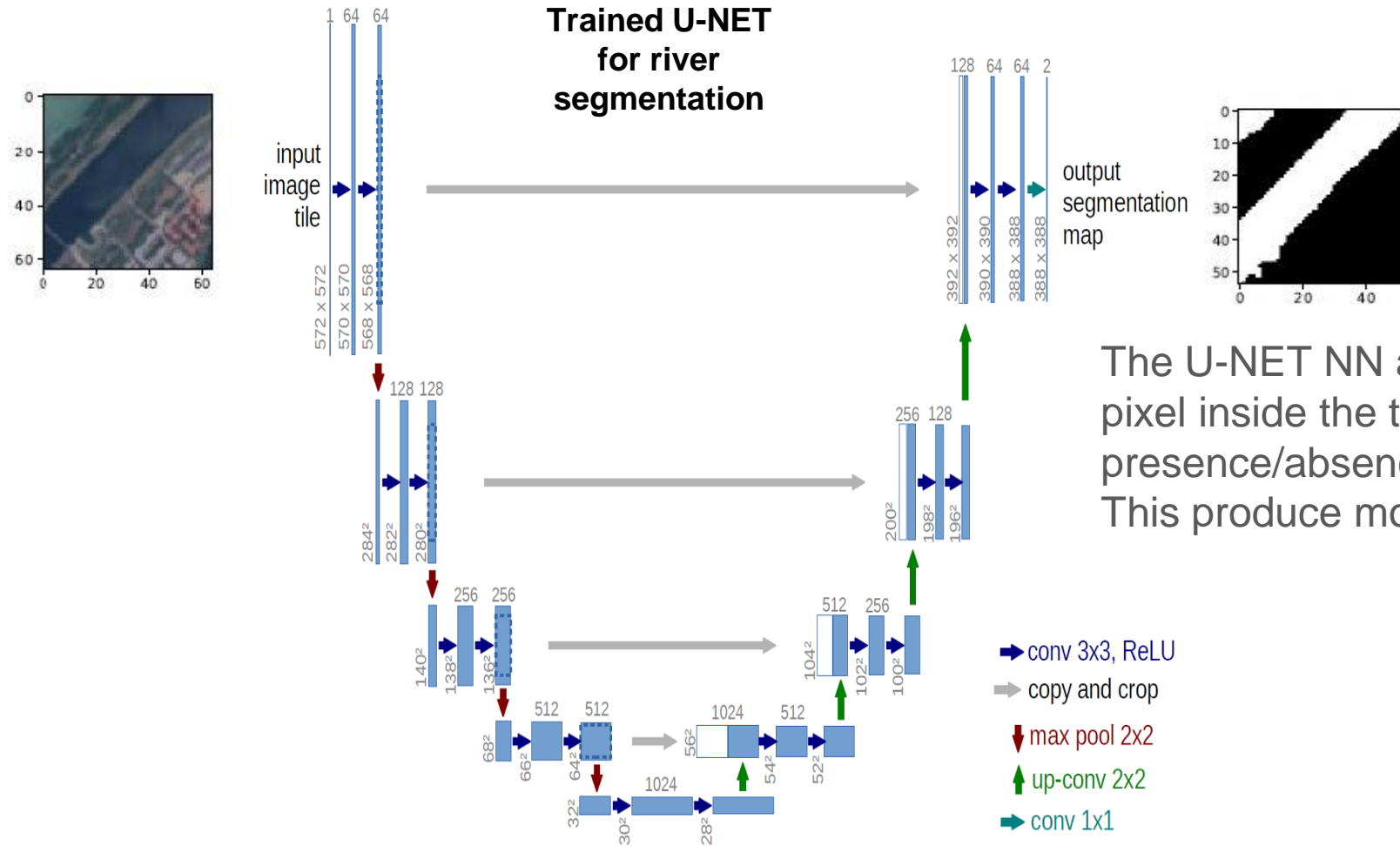


RIVER



- Overestimation of Eurosat *line-shape* classes as Rivers and Highways: giving a label e.g., highway to whole a tile 640m x 640m creates problems of overestimation.
- We apply two U-NET neural networks specialized for segmentation to river and highway classes

U-NET Architecture



The U-NET NN associates each pixel inside the tile to the presence/absence of the river class. This produce more detailed outputs

Dataset UNET

EuroSAT River

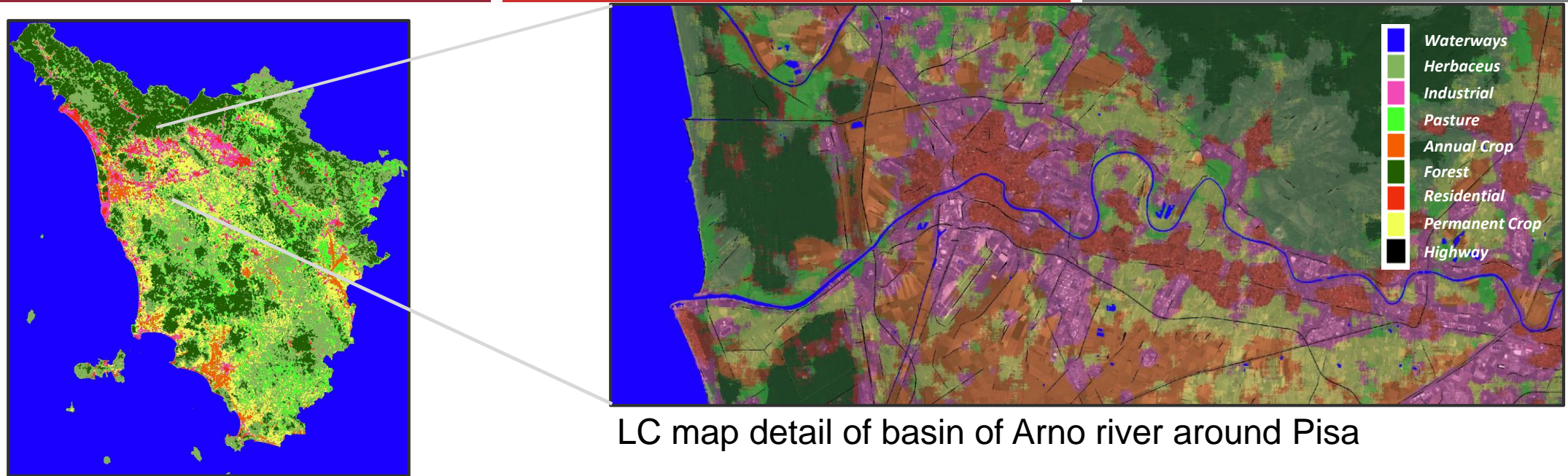


Mask / Label



- Creation of a dataset suitable for **UNET** training using satellite images of EuroSAT **River** class along with data from **Copernicus High Resolution Layer**.
- The label (mask image) must contain the information about the pixel
 - 0 does not belong to class River (black)
 - 1 belongs to the River class (white)
- Our dataset: 1500 validated segmentation masks.
- Building a similar training dataset for the **Highway** class using data from **Open Street Maps**.
- **We are ready to publish and share these datasets for river and highway classes segmentation ML tasks**

Merging of CNN + U-Net outputs



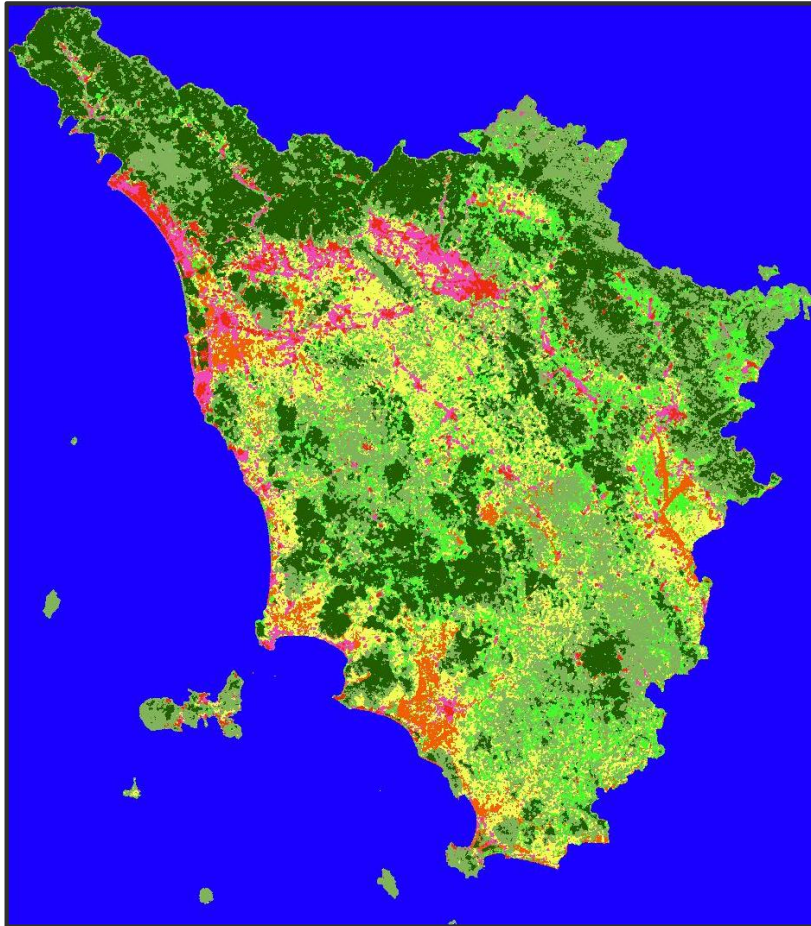
LC map detail of basin of Arno river around Pisa

Land Cover map of Tuscany Region

The final step assigns to each cell the class identified by the U-Net or the class identified by CNN, according to a hierarchy that depends on the degree of confidence we have on each algorithm.

1. River/Waterway
2. Highways
3. Other classes

Regional Estimate of Land Cover



LC CLASS	COVERAGE %
Annual Crop	4.1
Forest	34.1
Herb. Vegetation	27.2
Highway	0.7
Industrial	5.2
Pasture	10.6
Permanent Crop	14.6
Residential	2.4
River	1.6

Land Cover map of Tuscany Region

Conclusions

- LC statistics can benefit from machine learning methods and algorithms to achieve a high degree of automation in maps and statistics production.
- This would decrease the effort and costs
- Improve the frequency of the output.
- Classify-and-count generates good estimates for areal LC classes
- U-net approach improves estimates, but needs dataset with specific mask-label
- In the future, we could explore the possibility to assign labels using information from administrative sources, regional technical charts, cadastral maps, and agricultural census to extend U-Net approach to other LC classes
- Use further information such as SAR data from Sentinel 1 or satellite image obtained in different seasons e.g. Crop Classification

Thank you

Erika Cerasti
Fabrizio De Fausti,
Angela Pappagallo
Francesco Pugliese
Diego Zardetto