ETH zürich

Mapping Vulnerable Populations with AI

Dr. Rodrigo Caye Daudt

03 May 2022, Brussels

Motivation

The International Committee of the Red Cross (ICRC) runs numerous humanitarian action and disaster relief operations, e.g. vaccine distribution, helping war refugees, natural disaster response, etc.

The organisation of such operations requires accurate, up to date data regarding various regions of interest. Such information is often inaccurate or out of date.

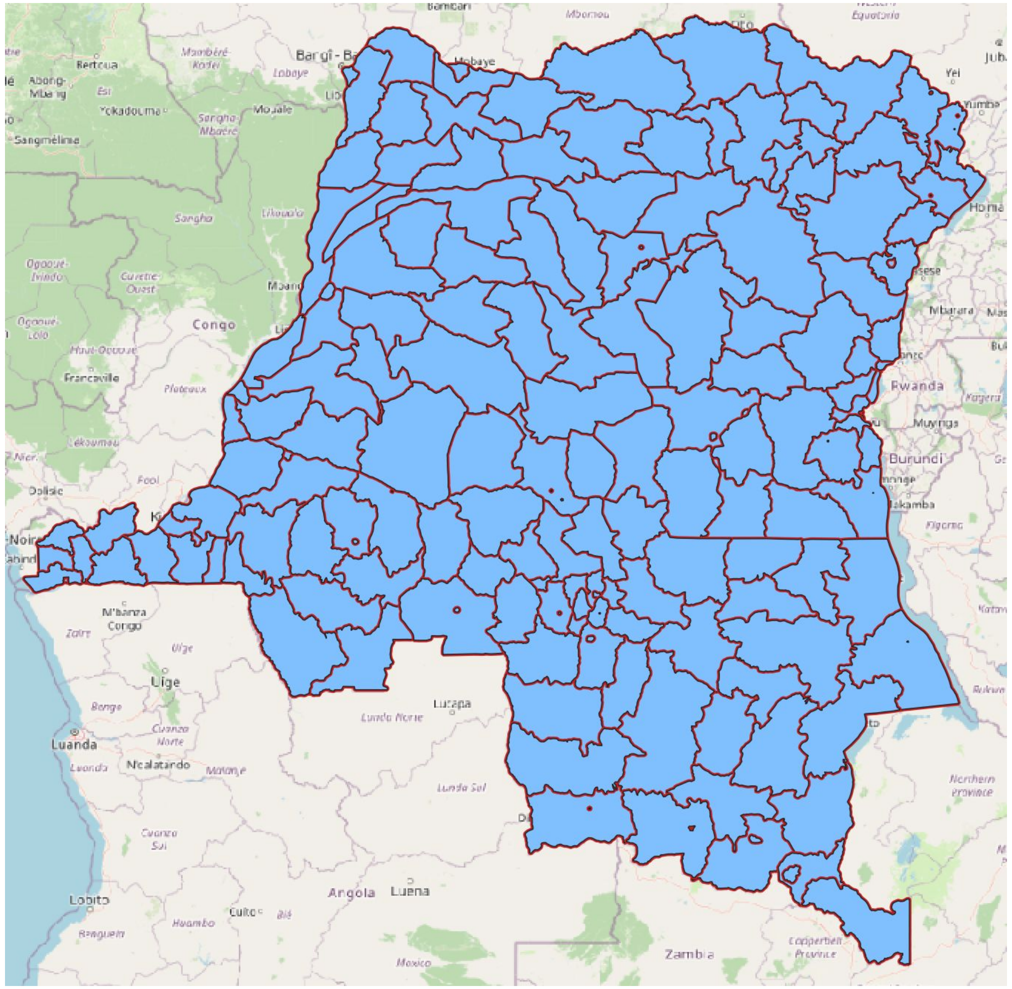
Motivation

Data in developing countries are also often too coarse to be used.

Motivation

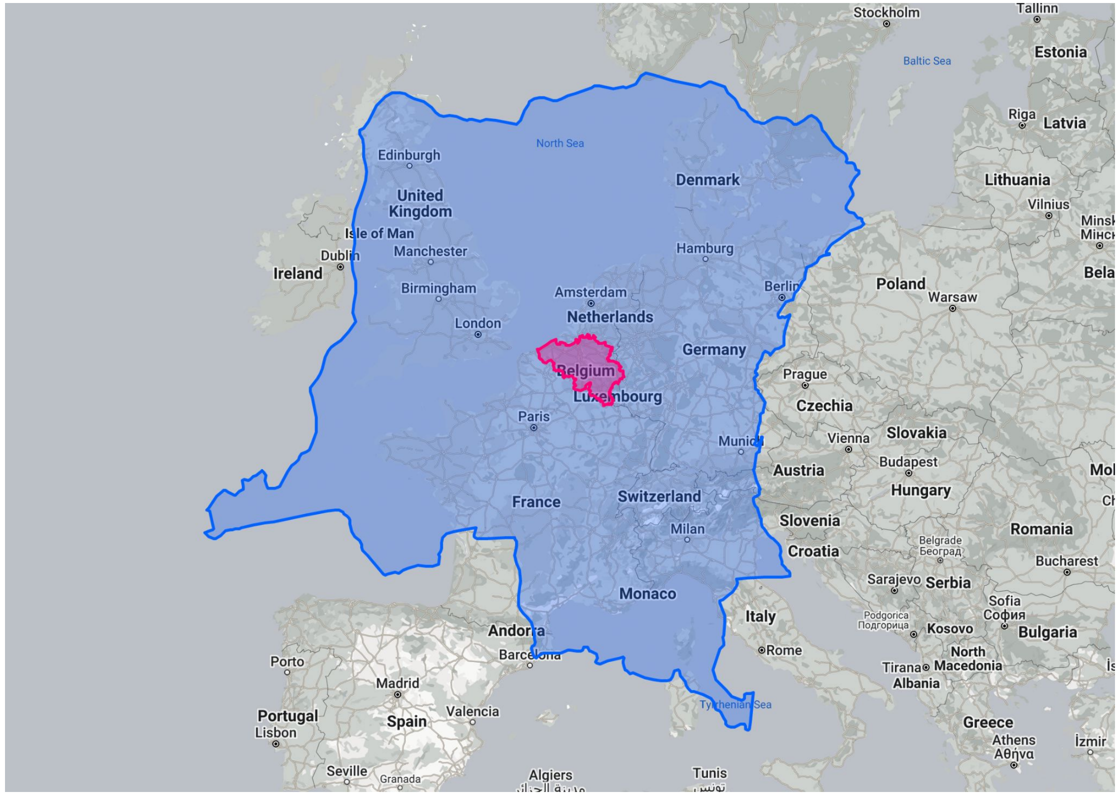
Data in developing countries are also often too coarse to be used.

Example: census data for one of the countries of interest, the Democratic Republic of the Congo (DRC), is available at the finest scale for only 189 administrative regions.



Country of Interest: Democratic Republic of the Congo (DRC)

Country of Interest: Democratic Republic of the Congo (DRC)

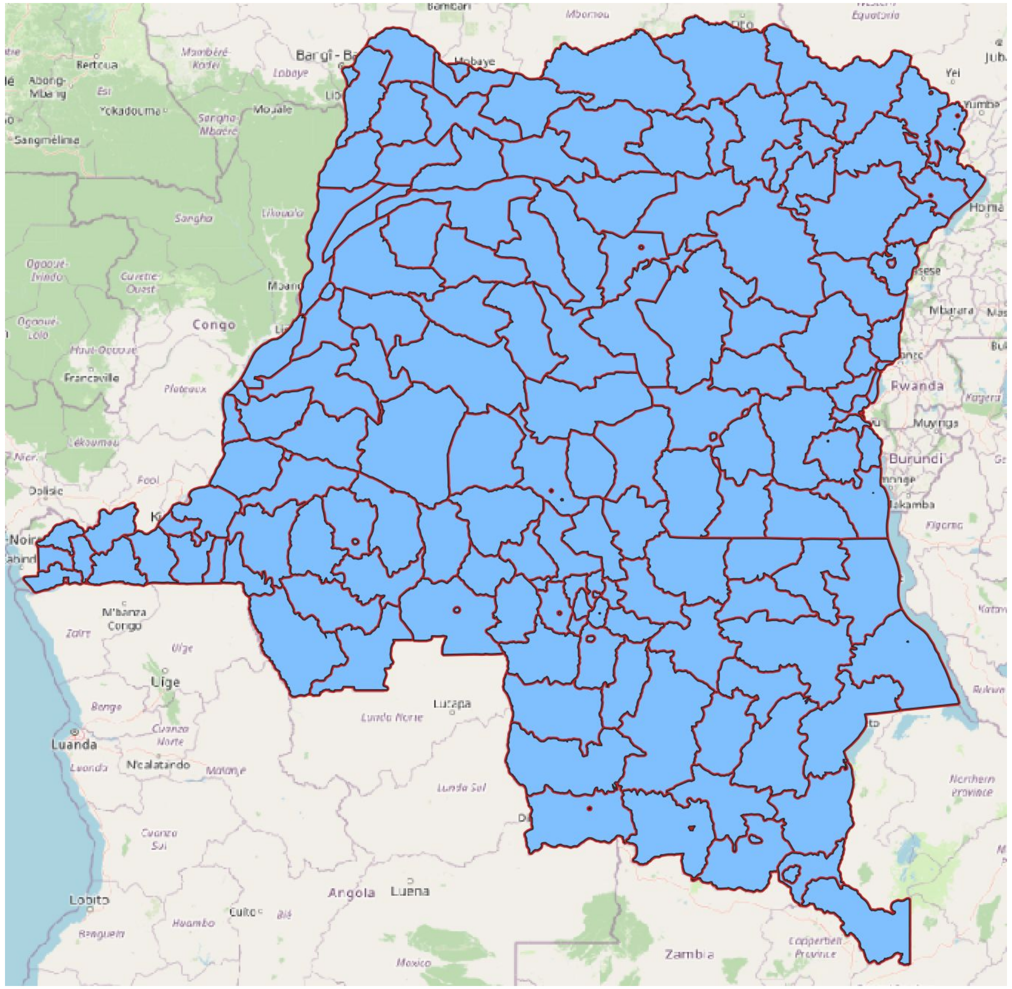


Motivation

Data in developing countries are also often too coarse to be used.

Example: census data for one of the countries of interest, the Democratic Republic of the Congo (DRC), is available at the finest scale for only 189 administrative regions.

This means that the average region size is approximately 12.5 thousand km^2.



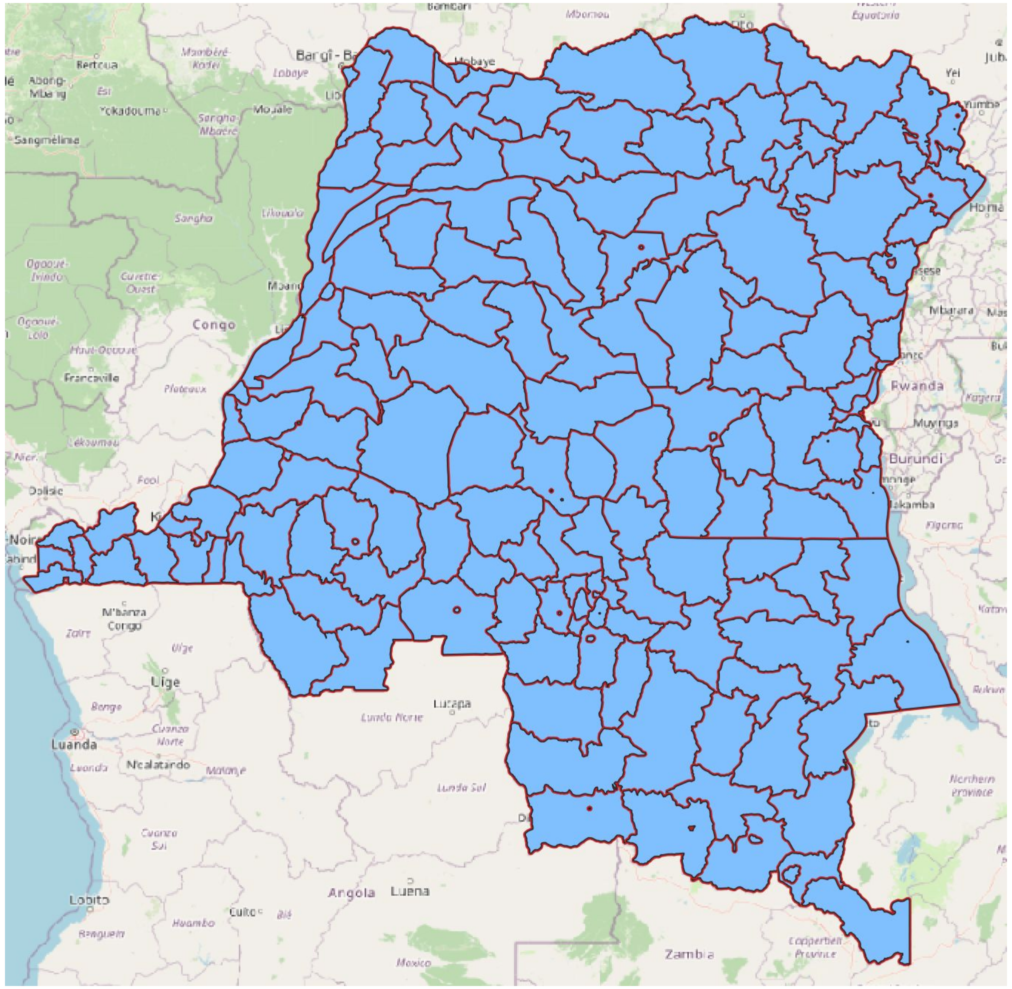
Motivation

Data in developing countries are also often too coarse to be used.

Example: census data for one of the countries of interest, the Democratic Republic of the Congo (DRC), is available at the finest scale for only 189 administrative regions.

This means that the average region size is approximately 12.5 thousand km^2.

This means that each administrative region is roughly **half the size of Belgium**.



Project Partners

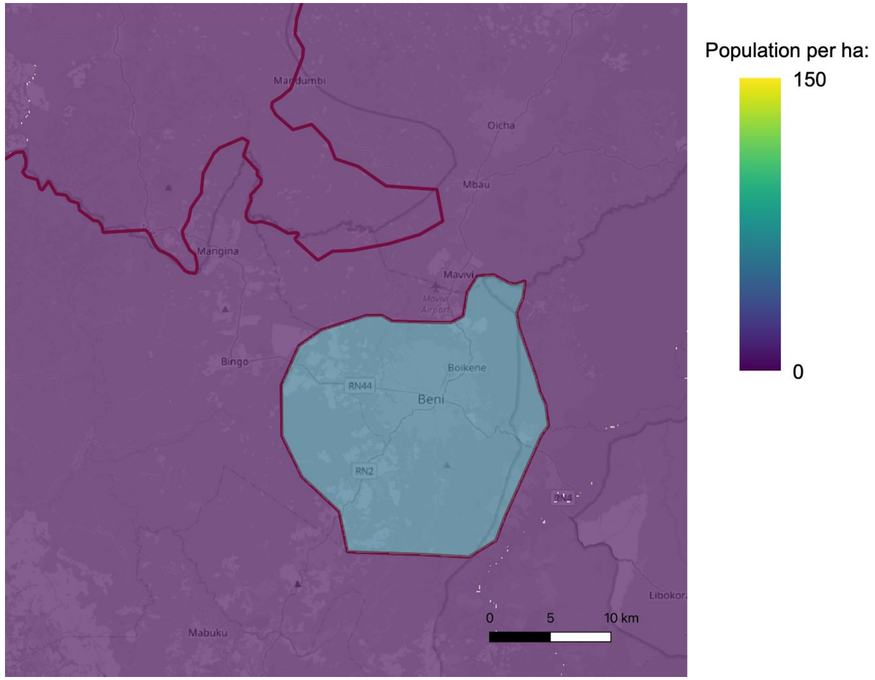
ICRC: Application and ground operations

ETH Zurich and **EPFL**: expertise in remote sensing, geodata, and deep learning.

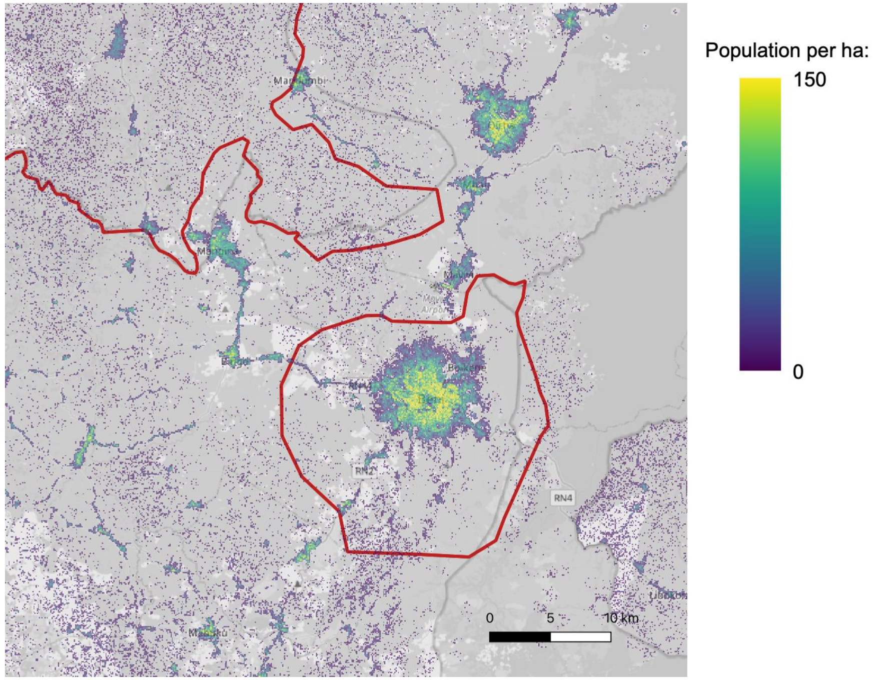
QCRI: expertise in extracting information from social media data

Objective

We aim to map population at a a fine-grained scale to support humanitarian operations by the ICRC (and others who are interest in such maps).



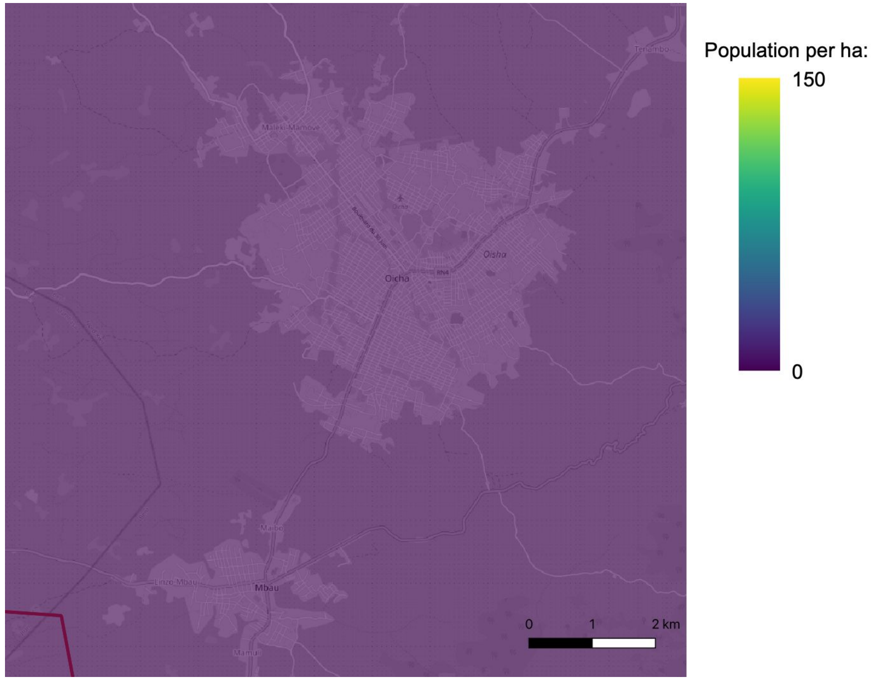
Census population data.



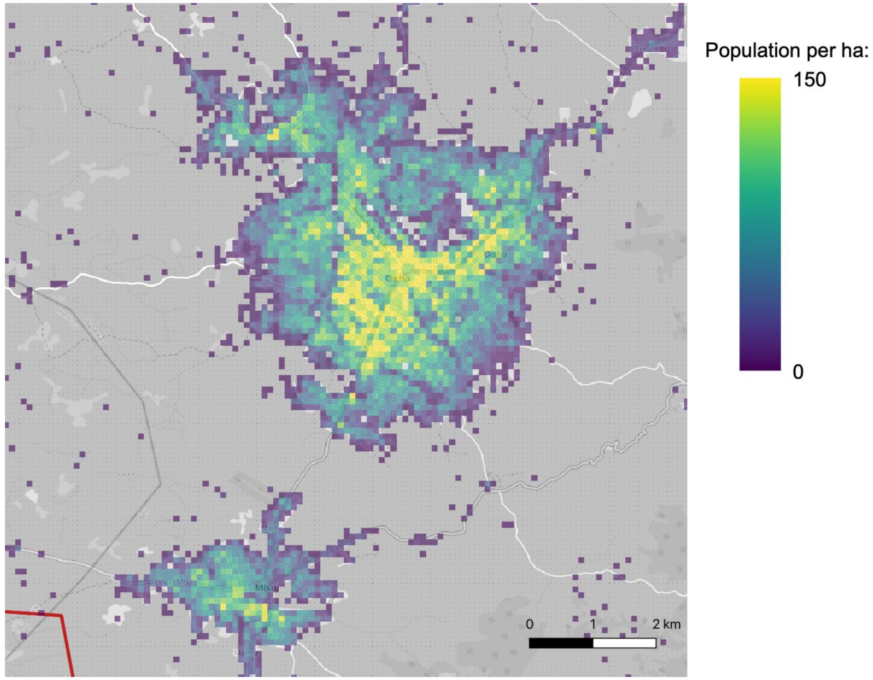
Fine-grained population estimates.

Objective

The importance is even more apparent in small and medium towns.



Census population data.



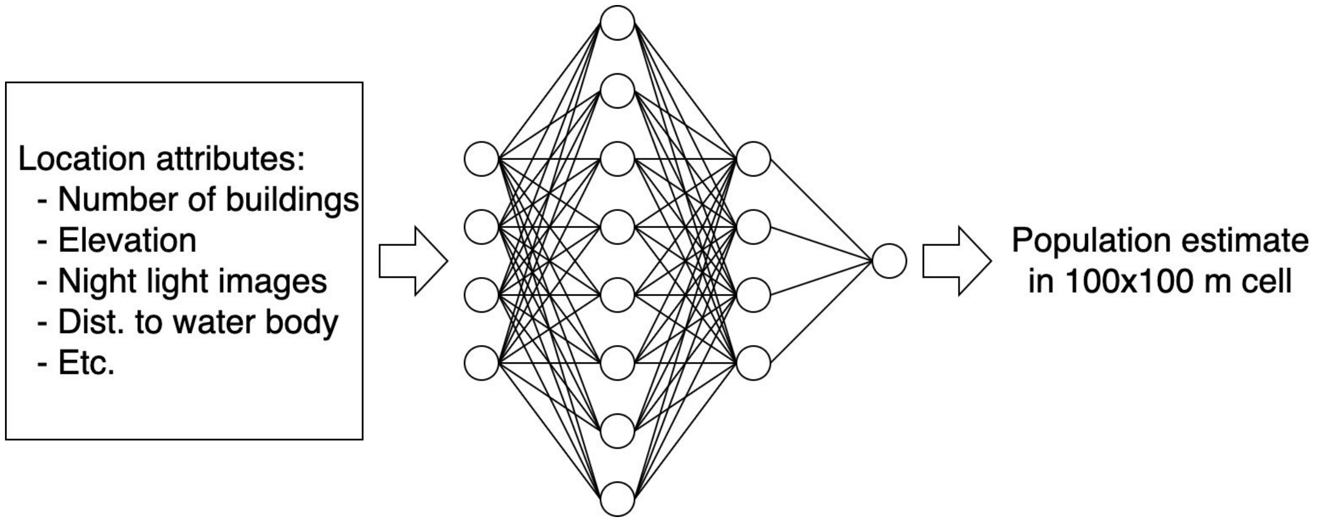
Fine-grained population estimates.

Our Approach

We use a weakly supervised **neural network** to estimate population densities in a regular grid of 100x100 meter cells.

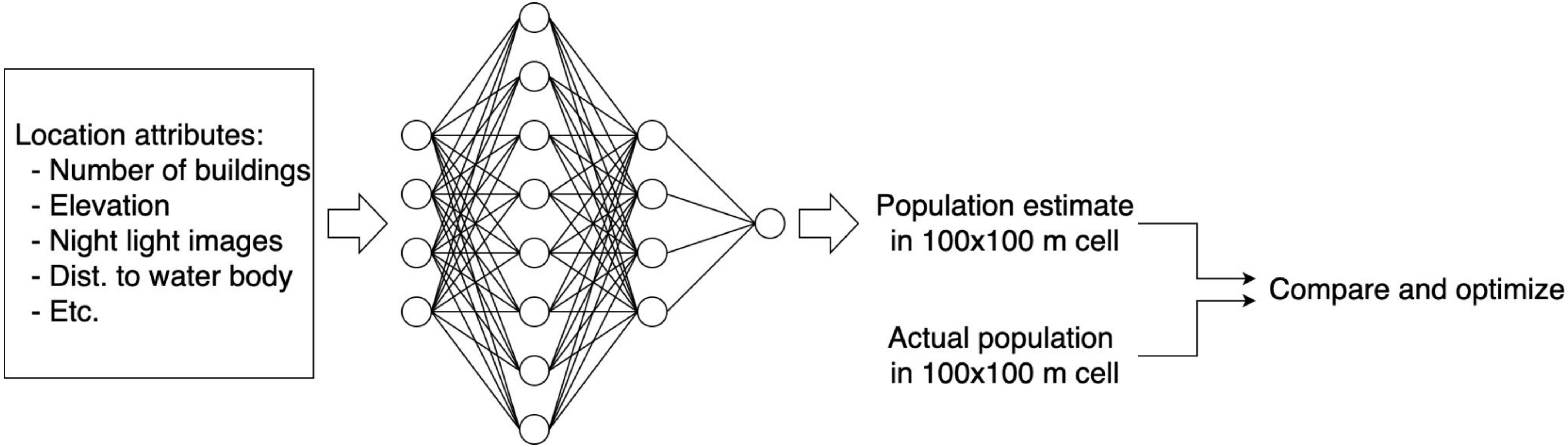
The networks uses several attributes of a given location (covariates) to perform disaggregation of census data.

Training and evaluating the neural network is tricky since **ground truth data are only available at a coarse level**, which impedes a direct naive application of machine learning methods.



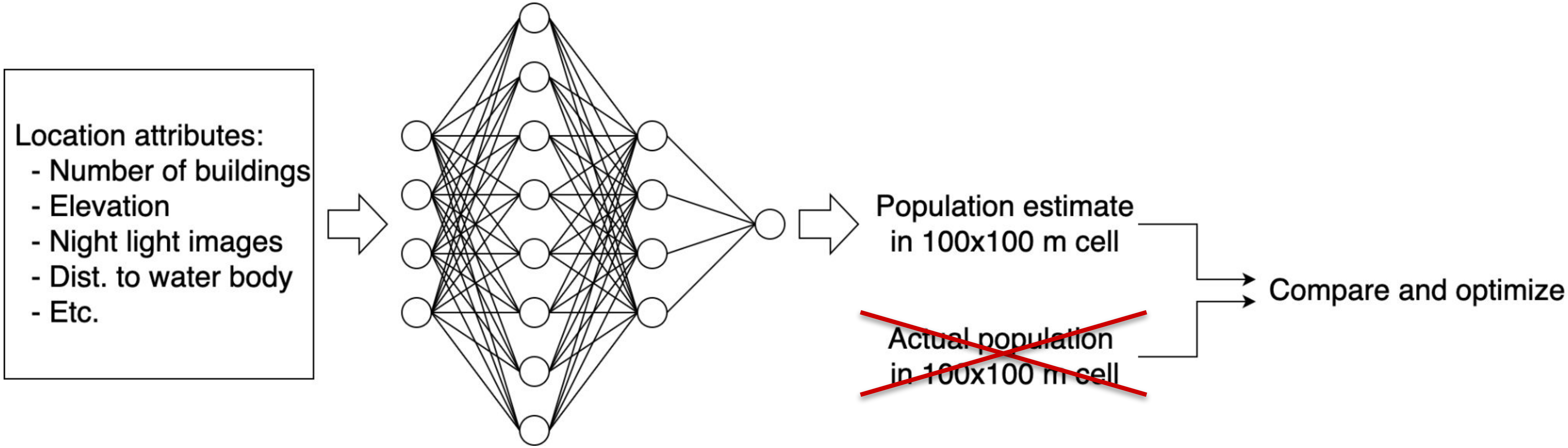
Weakly Supervised Learning

Most machine learning systems learn by comparing predictions with known **ground truth data**.



Weakly Supervised Learning

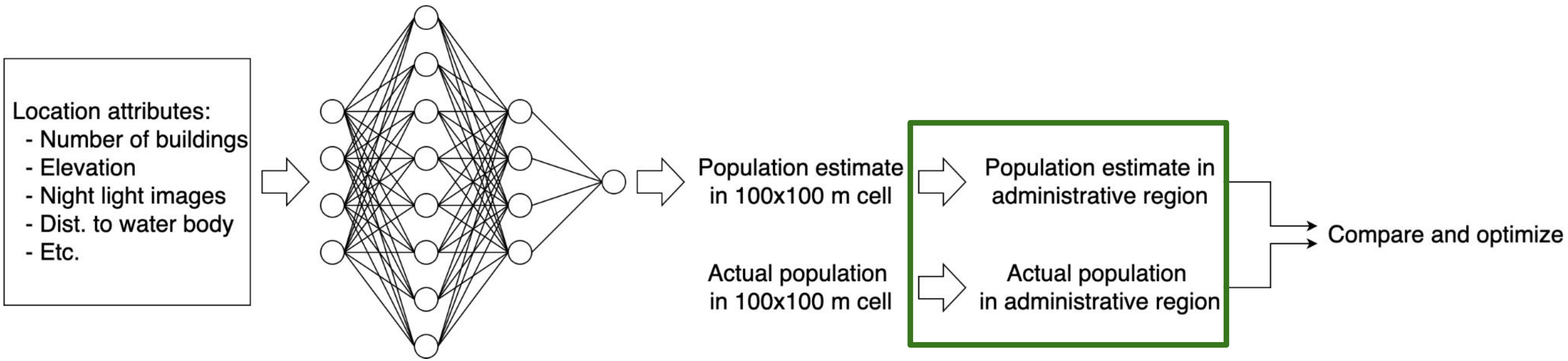
Most machine learning systems learn by comparing predictions with known **ground truth data**.



Unfortunately, such ground truth data are **unavailable**. We only have reference data at the level of **administrative regions**.

Weakly Supervised Learning

Predictions are then **aggregated** inside administrative regions for comparison with ground truth.



This is challenging, but most importantly, this is **possible**.

In the DRC, for example, regions contain on average 1.25 million cells of 100x100 meters. Such large regions are tricky to handle in this manner.

Covariates

The selection of covariates that help in predicting population densities is important.

These are extracted from various remote sensing and geodata products.

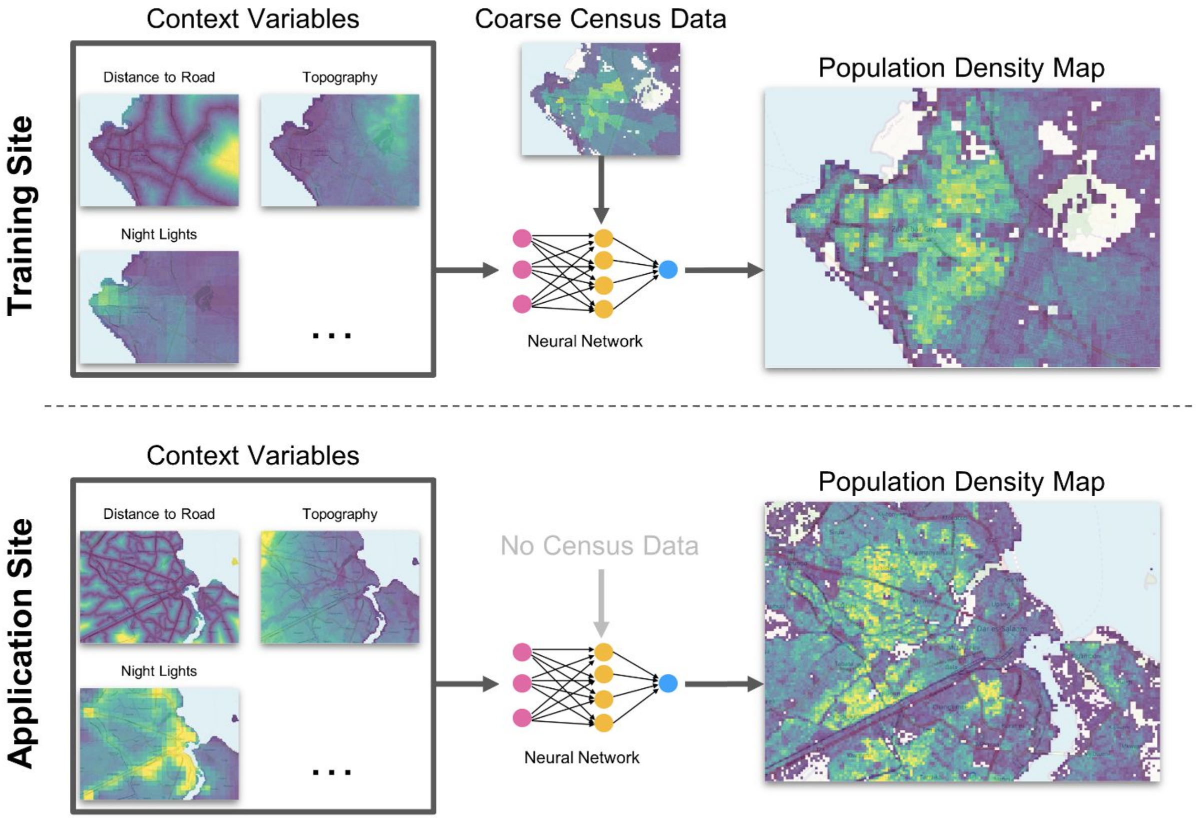
Building count is, unsurprisingly, the most important covariate.

Some of the covariates that are currently used are:

- Building count
- Distance to coastline
- Distance to water
- Buildings mean area
- Protected areas
- Travel time to closest city
- Night light images
- Terrain slope
- Elevation
- Distance to roads
- Distance to road intersections
- Distance to waterways

Beyond Available Census Data

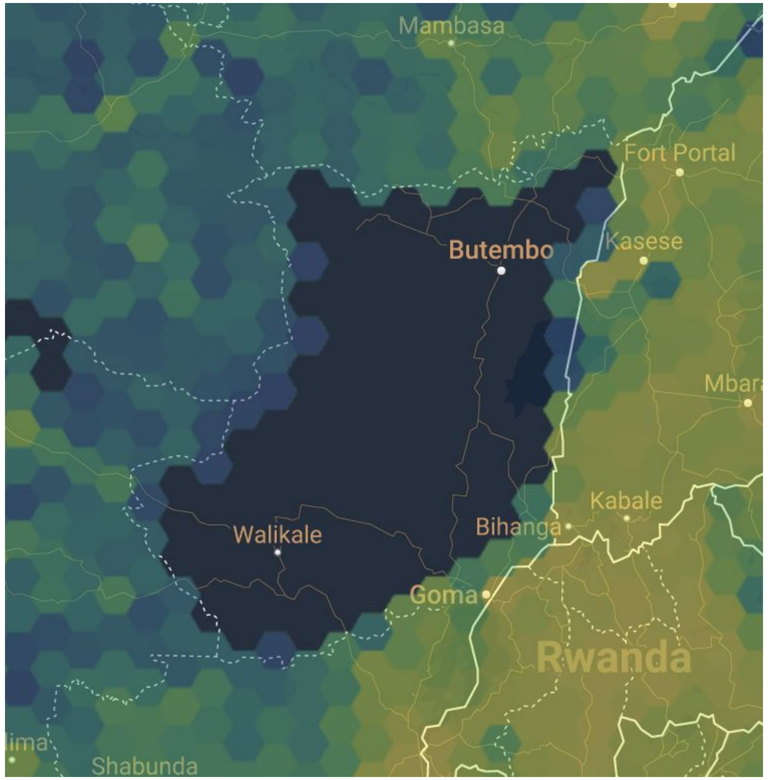
We can also apply the neural network to a country where census data are unavailable or unreliable, as long as we have access to data relative to a **similar** (neighbouring) country.



Challenges

There are many challenges that we are still working on:

- Building maps are often not complete due to geopolitical reasons
- The massive aggregation from 100x100 meter blocks to administrative region is makes it difficult to train the neural network
- Accurately generalizing from one country to another is more challenging than it sounds
- Verifying the accuracy of the produced maps independently from census data is tricky
    - Luckily the ICRC can help us with ground surveys!
- Conflicts often cause migrations that are hard to model, but are also correlated with humanitarian applications
- Census data are not always accurate or up to date



Missing building data in the DRC in the Open Buildings dataset.

ETH zürich

What Next?

Estimate population demographics: age distribution, sex distribution, etc.

Acquire additional information from satellite images and social media, eg. neighborhood characteristics (residential, commercial, industrial, etc.).

In progress: completing building maps using Sentinel-2 images.

Estimate uncertainties associated to population estimates.

Increase the scope of the current work to more countries, ideally covering at least the entire African continent.

Share the produced maps with the ICRC and other humanitarian action organisations.

Dr. Rodrigo Caye Daudt
Scientific assistant
rcayedaudt@ethz.ch

ETH Zürich
Photogrammetry and Remote Sensing
Stefano- Franscini-Platz 5
8049 Zürich

https://rcdaudt.github.io/

Project partners:

Nando Metzger (ETH Zurich)
Rodrigo Caye Daudt (ETH Zurich)
Konrad Schindler (ETH Zurich)

John E. Vargas-Muñoz (EPFL)
Benjamin Kellenberger (EPFL)
Devis Tuia (EPFL)

Thao Ton-That Whelan (ICRC)

Ferda Ofli (QCRI)
Muhammad Imran (QCRI)