Data Quality and feature extraction at scale with RoboSat.pink

@o_courtin

@sotm 2019

Goals

Detect inconsistencies between two DataSets

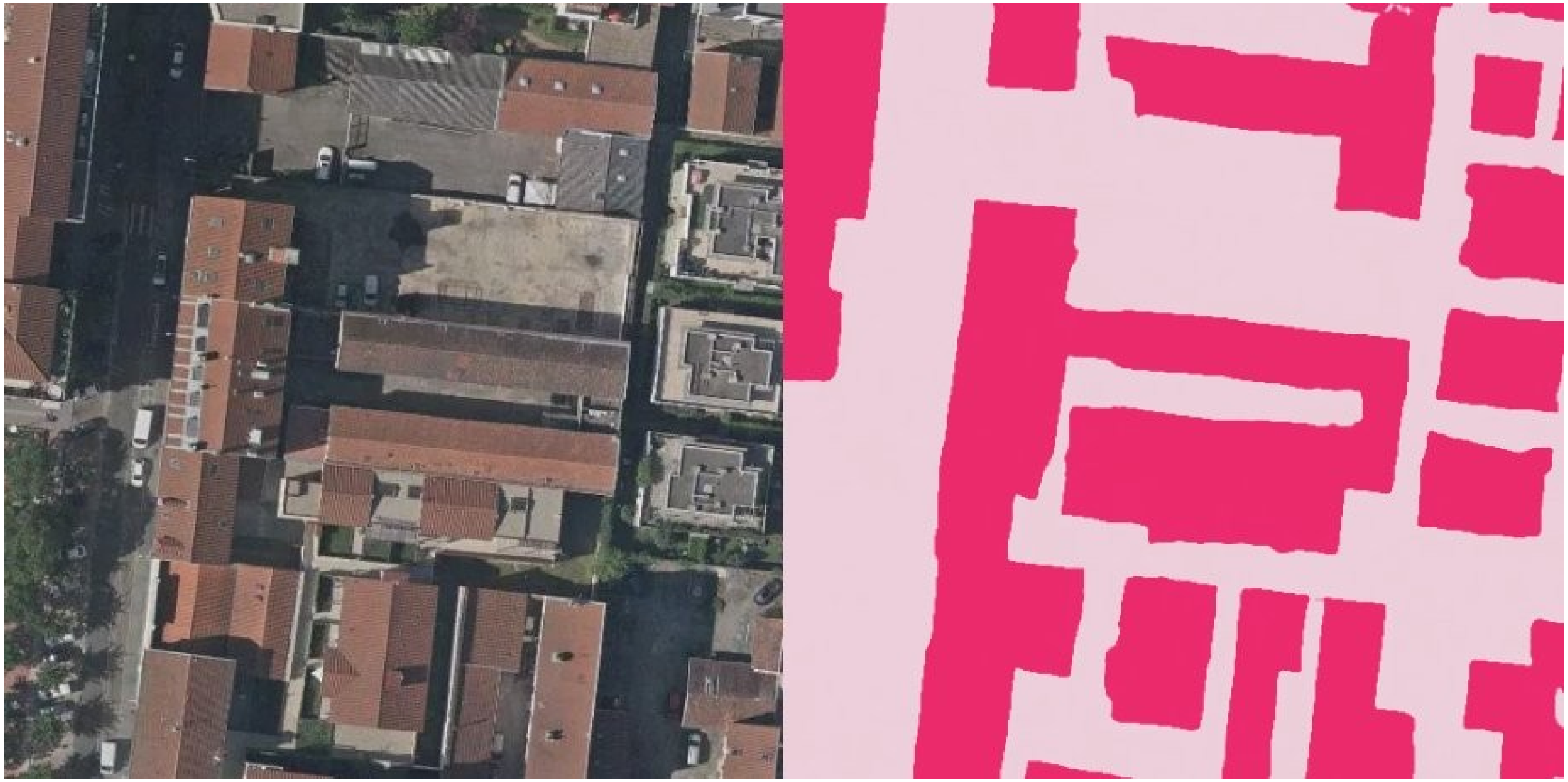
Train on a small area, predict on a larger one.

RoboSat.pink

DataSet Quality Analysis

Change Detection highlighter

Features extraction



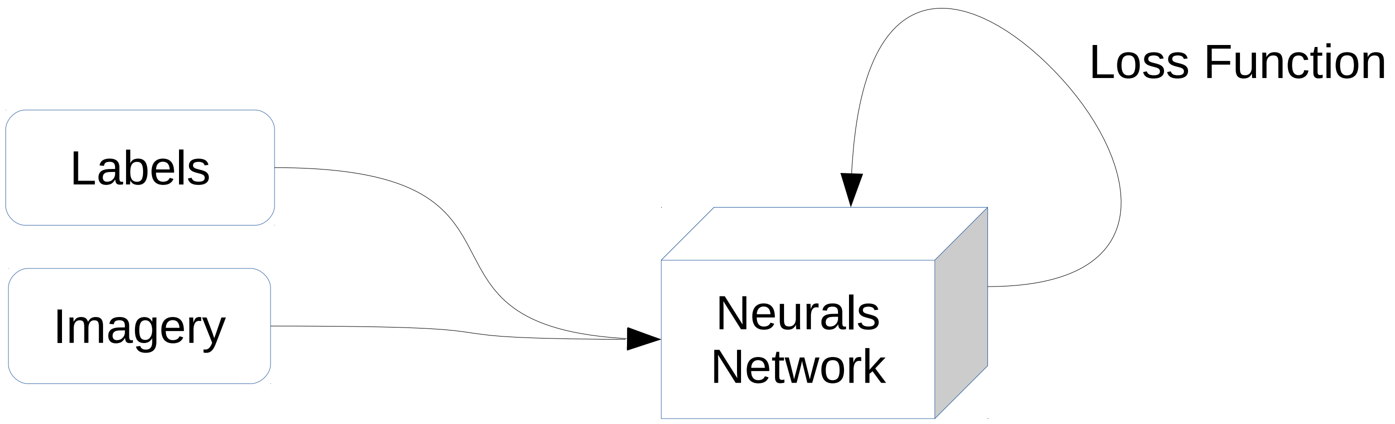
www.RoboSat.pink   Computer Vision framework for GeoSpatial Imagery

RoboSat != RoboSat.pink

https://github.com/mapbox/robosat

https://github.com/datapink/robosat.pink

https://github.com/mapbox/robosat/issues/184

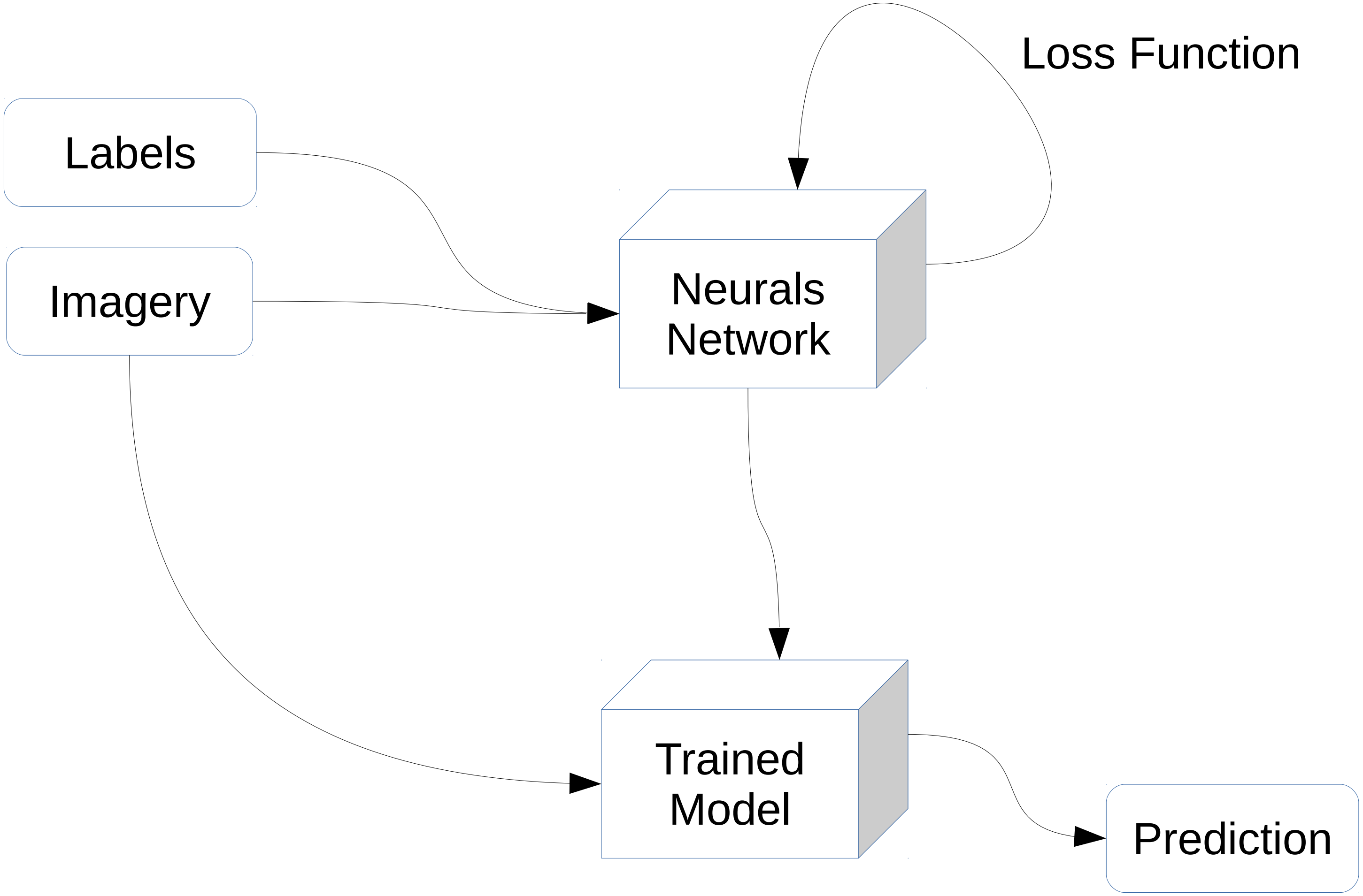


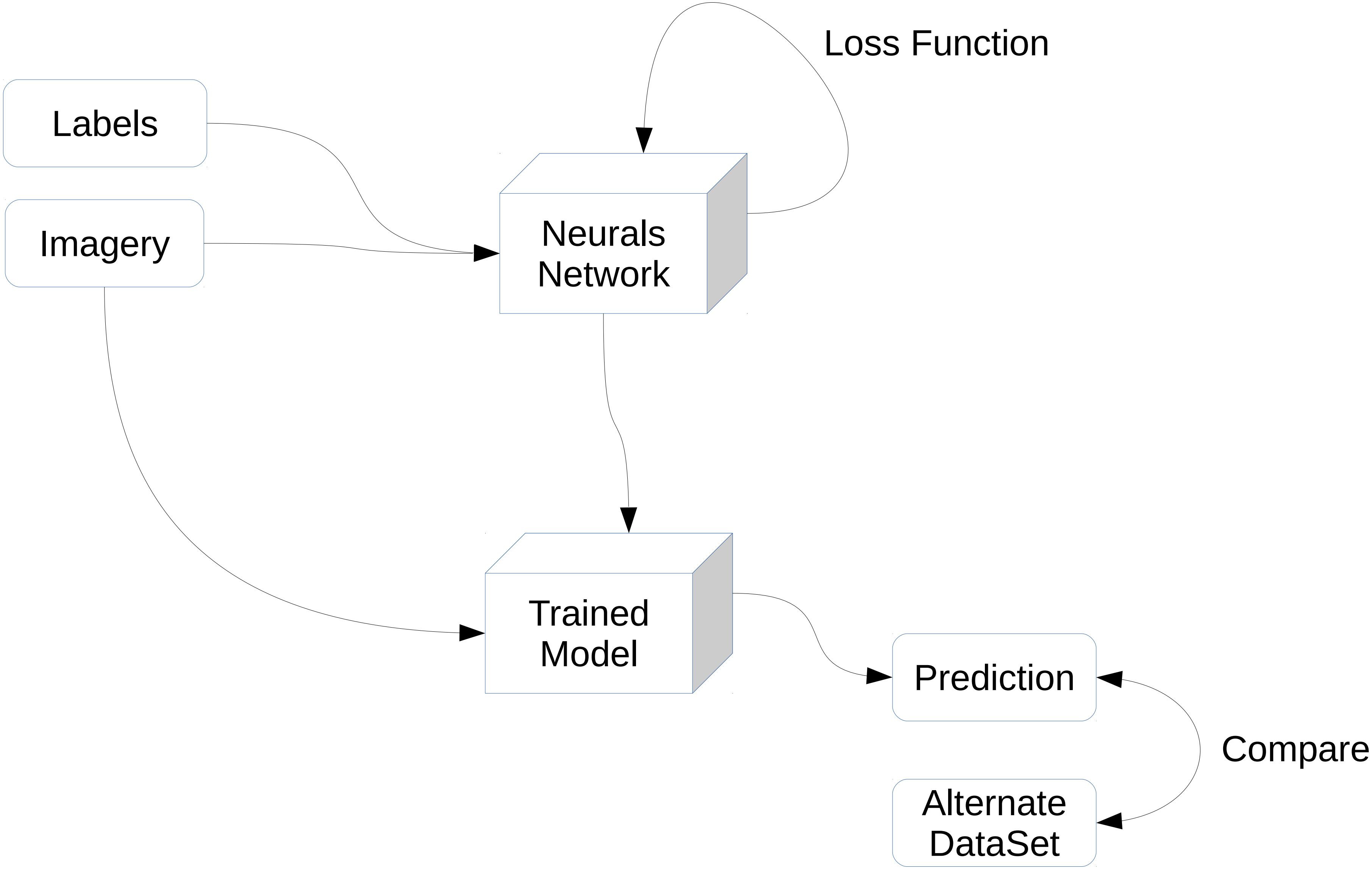
Labels

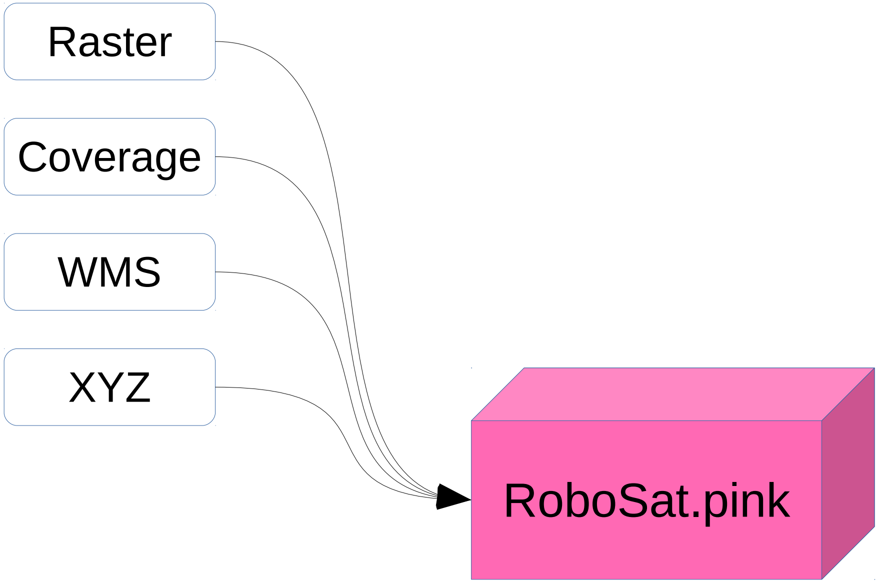
Imagery

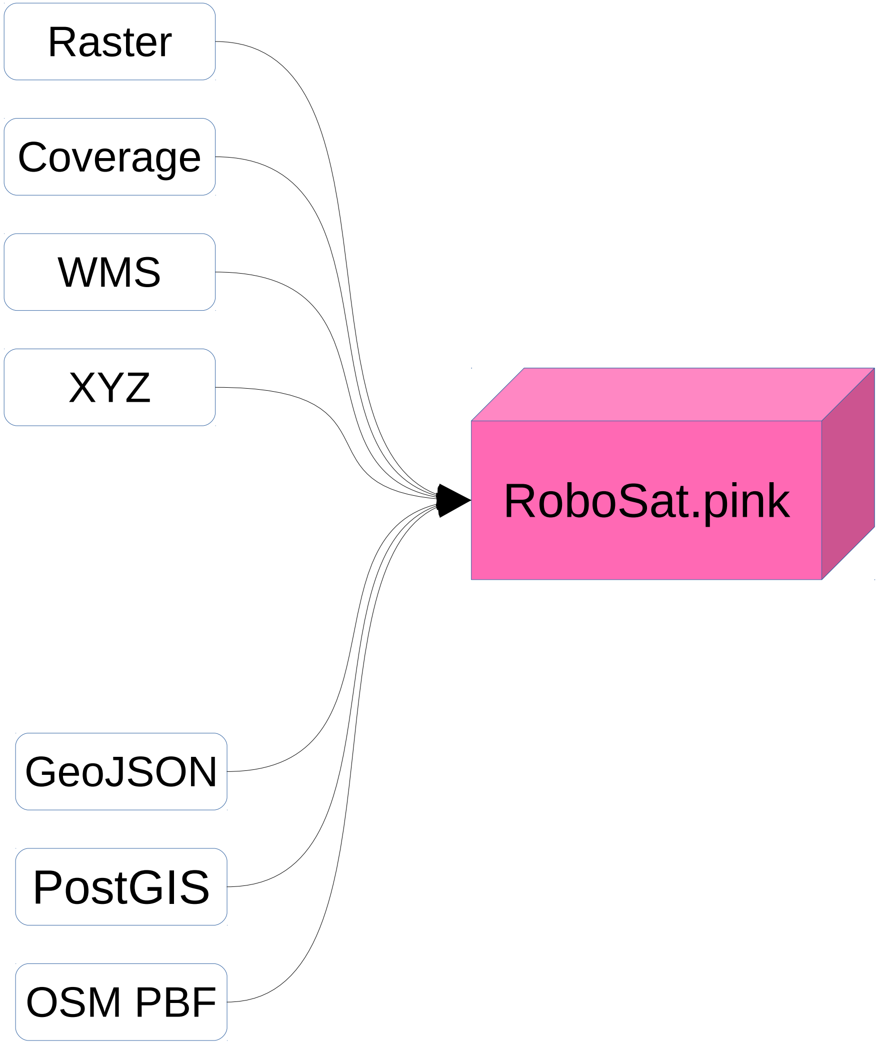
Neurals Network

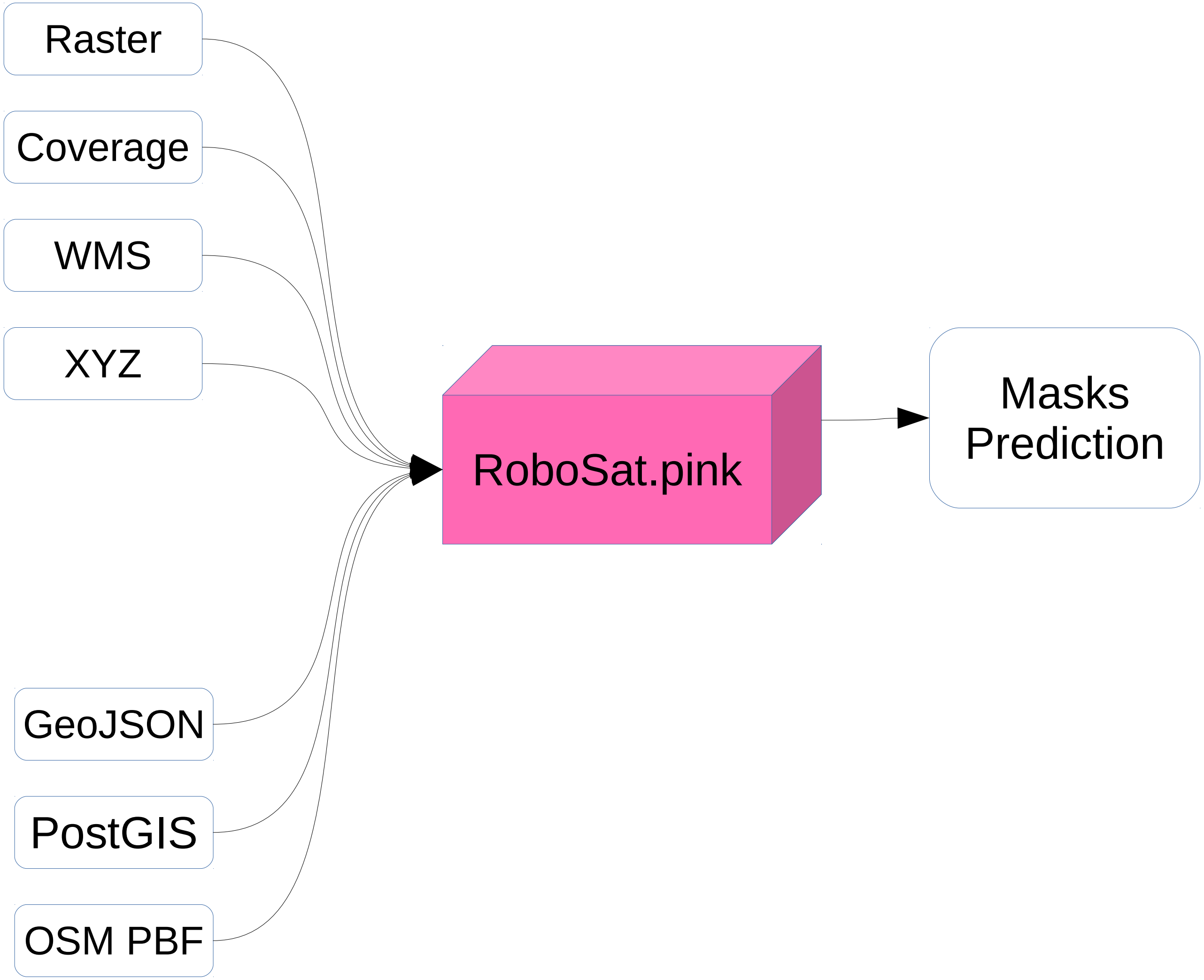
Loss Function

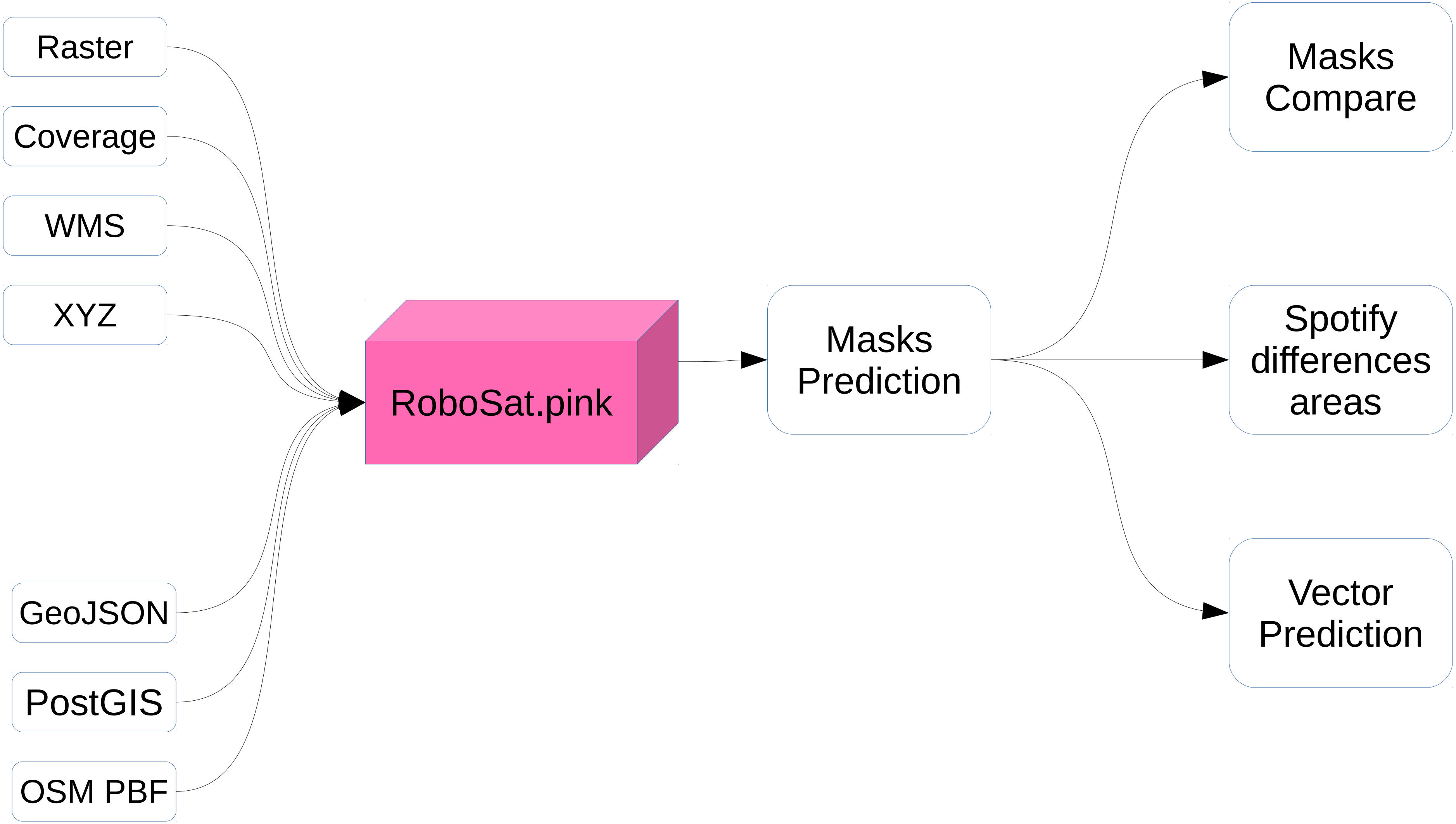

Loss Function
Labels
Imagery
Neurals Network
Trained Model
Prediction


Loss Function
Labels
Imagery
Neurals Network
Trained Model
Prediction
Compare
Alternate DataSet

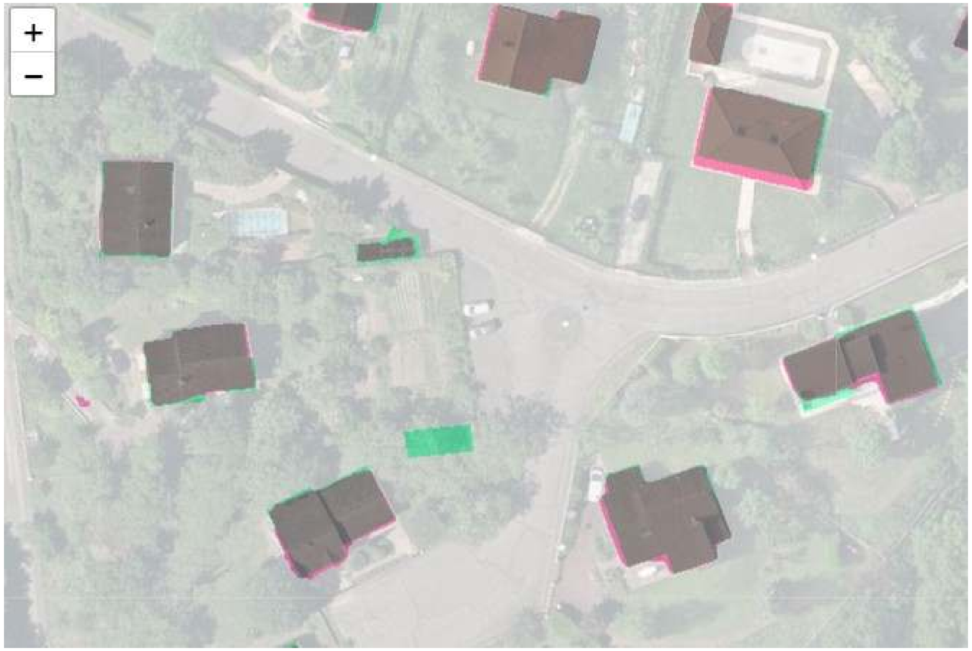

Raster
Coverage
WMS
XYZ
RoboSat.pink

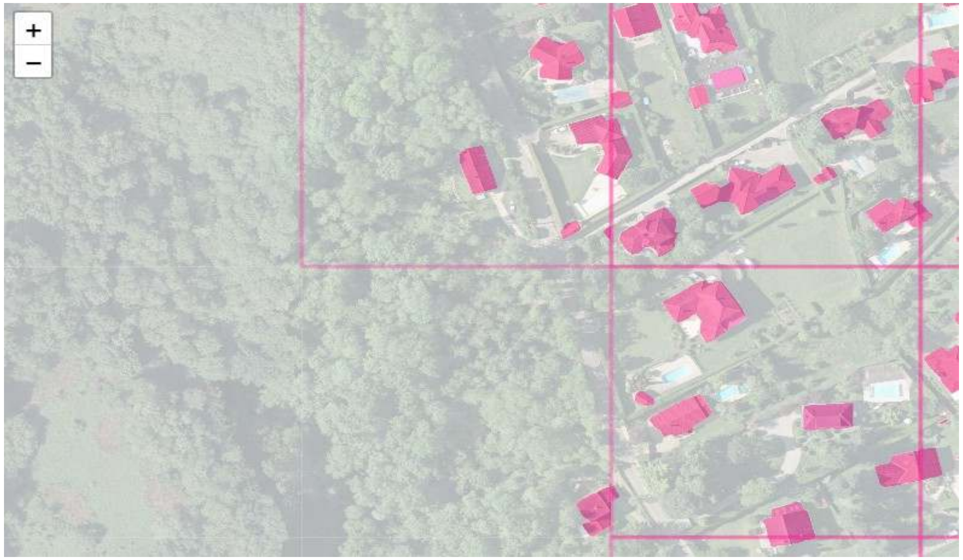

Raster
Coverage
WMS
XYZ
RoboSat.pink
GeoJSON
PostGIS
OSM PBF


Raster
Coverage
WMS
XYZ
GeoJSON
PostGIS
OSM PBF
RoboSat.pink
Masks
Prediction


Raster
Coverage
WMS
XYZ
GeoJSON
PostGIS
OSM PBF
RoboSat.pink
Masks Prediction
Masks Compare
Spotify differences areas
Vector Prediction

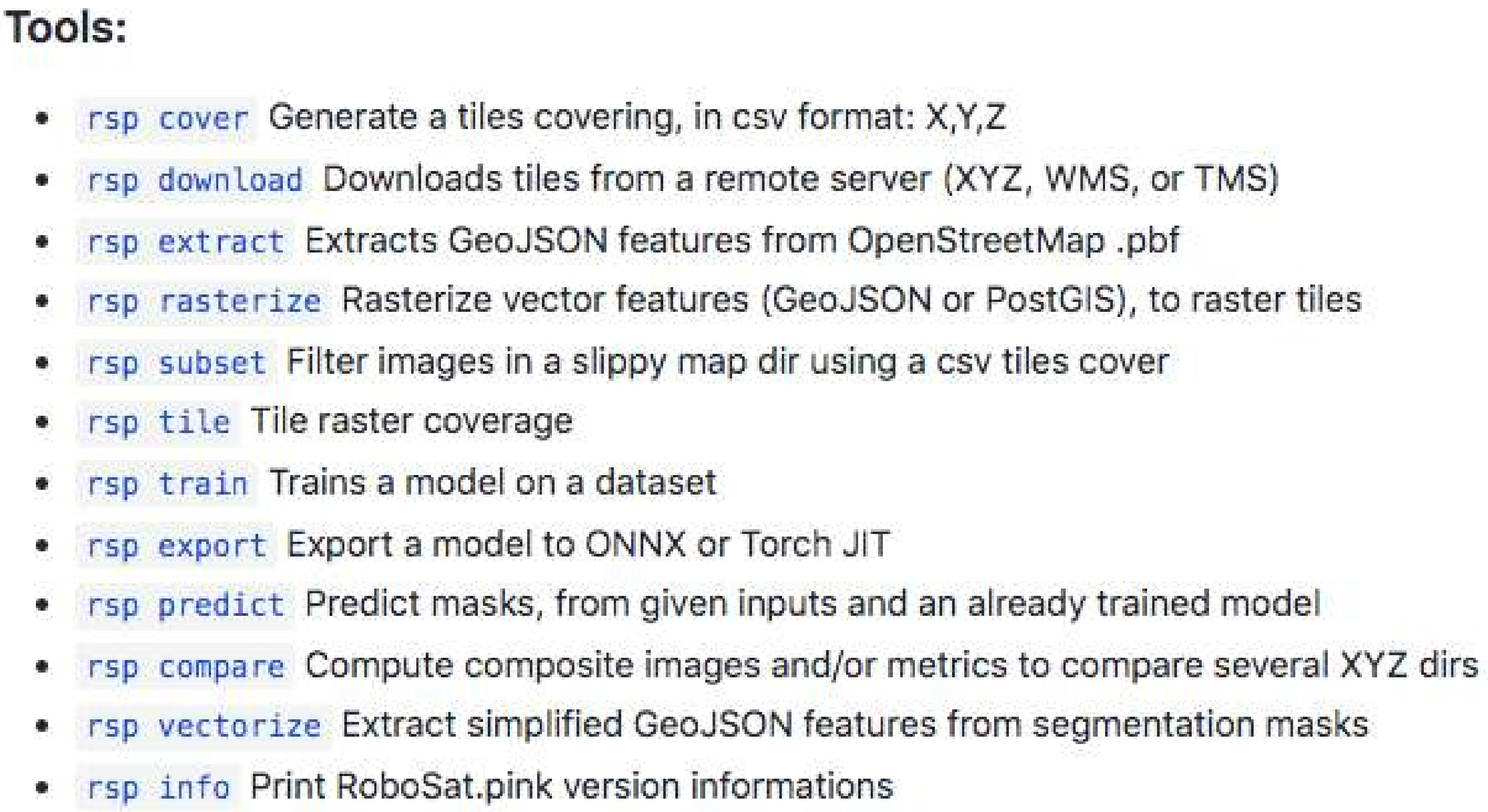
Compare Predicts against alternate datasets





Pink :	Predicted by trained model
Green :	Alternate dataset
Grey :	Both agree
Pink squares :	Significant differences

Command Line Interface



Tools:

- `rsp cover` Generate a tiles covering, in csv format: X,Y,Z
- `rsp download` Downloads tiles from a remote server (XYZ, WMS, or TMS)
- `rsp extract` Extracts GeoJSON features from OpenStreetMap .pbf
- `rsp rasterize` Rasterize vector features (GeoJSON or PostGIS), to raster tiles
- `rsp subset` Filter images in a slippy map dir using a csv tiles cover
- `rsp tile` Tile raster coverage
- `rsp train` Trains a model on a dataset
- `rsp export` Export a model to ONNX or Torch JIT
- `rsp predict` Predict masks, from given inputs and an already trained model
- `rsp compare` Compute composite images and/or metrics to compare several XYZ dirs
- `rsp vectorize` Extract simplified GeoJSON features from segmentation masks
- `rsp info` Print RoboSat.pink version informations

RoboSat.pink 101

https://github.com/datapink/robosat.pink/blob/master/docs/101.md

Easy to deploy

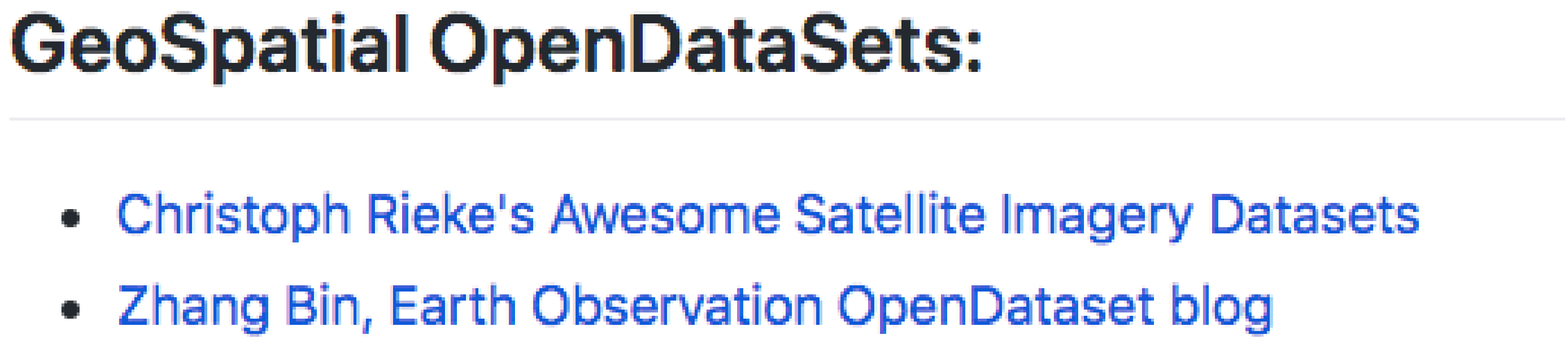
```
pip3 install RoboSat.pink
```

So all you need is :

- Imagery
- GPU
- Initial skills
- Labels

So all you need is :

- Imagery → *any file format readable by GDAL*
- GPU → *NVIDIA >= 8Go RAM*
- Initial skills → *GeoSpatial Data and CLI fluency*
- Labels → *that’s often the key point*

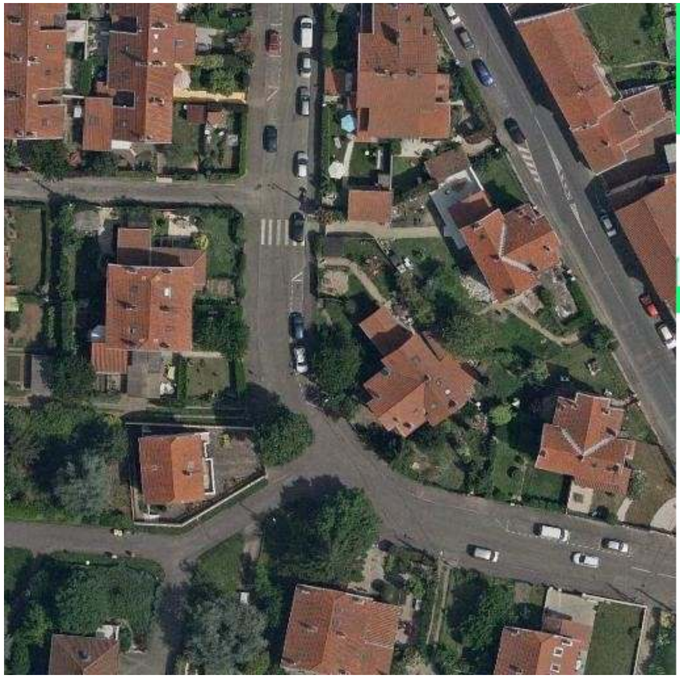


GeoSpatial OpenDataSets:

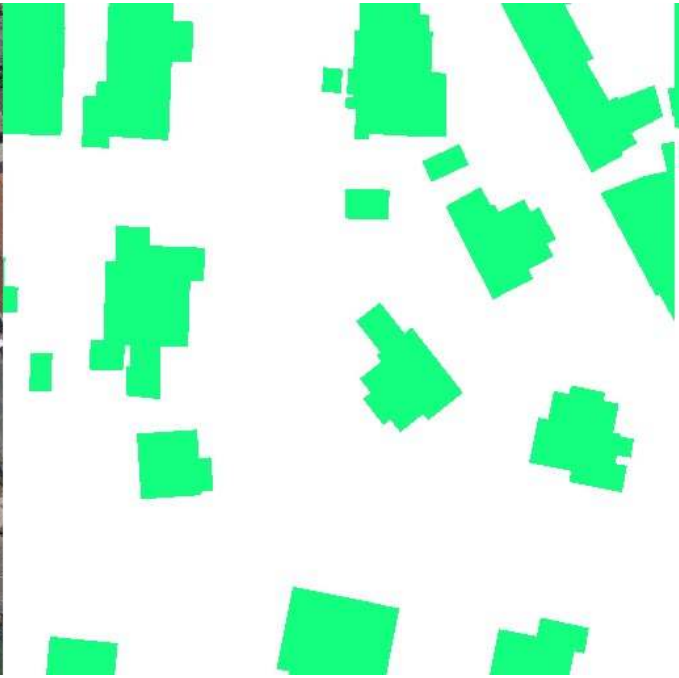
- Christoph Rieke's Awesome Satellite Imagery Datasets
- Zhang Bin, Earth Observation OpenDataset blog

https://github.com/datapink/robosat.pink#geospatial-opendatasets

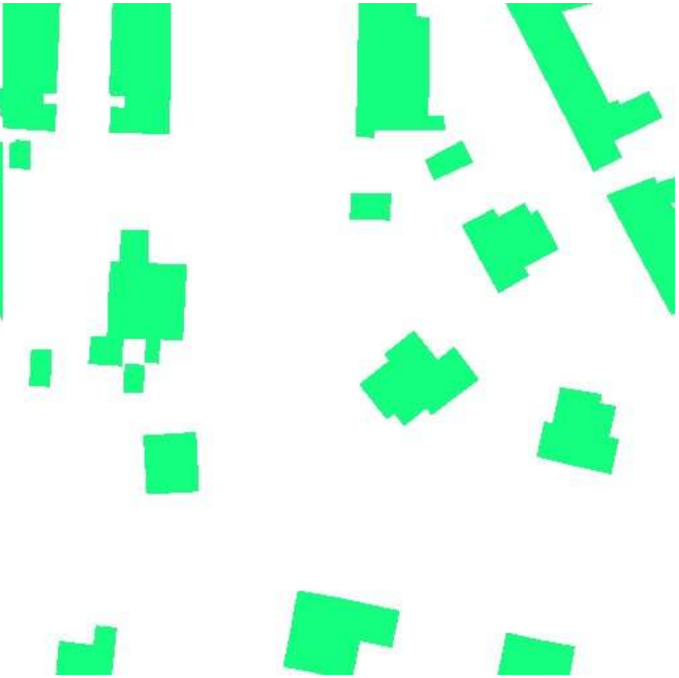
OSM not accurate enough for a good Training DataSet



Imagery



OpenData roofprints



OSM Building footprints

GIGO

An Ideal OpenDataSet

OpenData Licence compliant

Worldwide landscapes representative

Retina TileSize 512px

Mixed resolutions, bands, and sensors imagery

Labels (building, roads, vegetation…) with pixel accuracy

An Ideal OpenDataSet

OpenData Licence compliant

Worldwide landscapes representative

Retina TileSize 512px

Mixed resolutions, bands, and sensors imagery

Labels (building, roads, vegetation…) with pixel accuracy

No significant progress on this since @SOTM 2018 !

An Ideal OpenDataSet

OpenData Licence compliant

Worldwide landscapes representative

Retina TileSize 512px

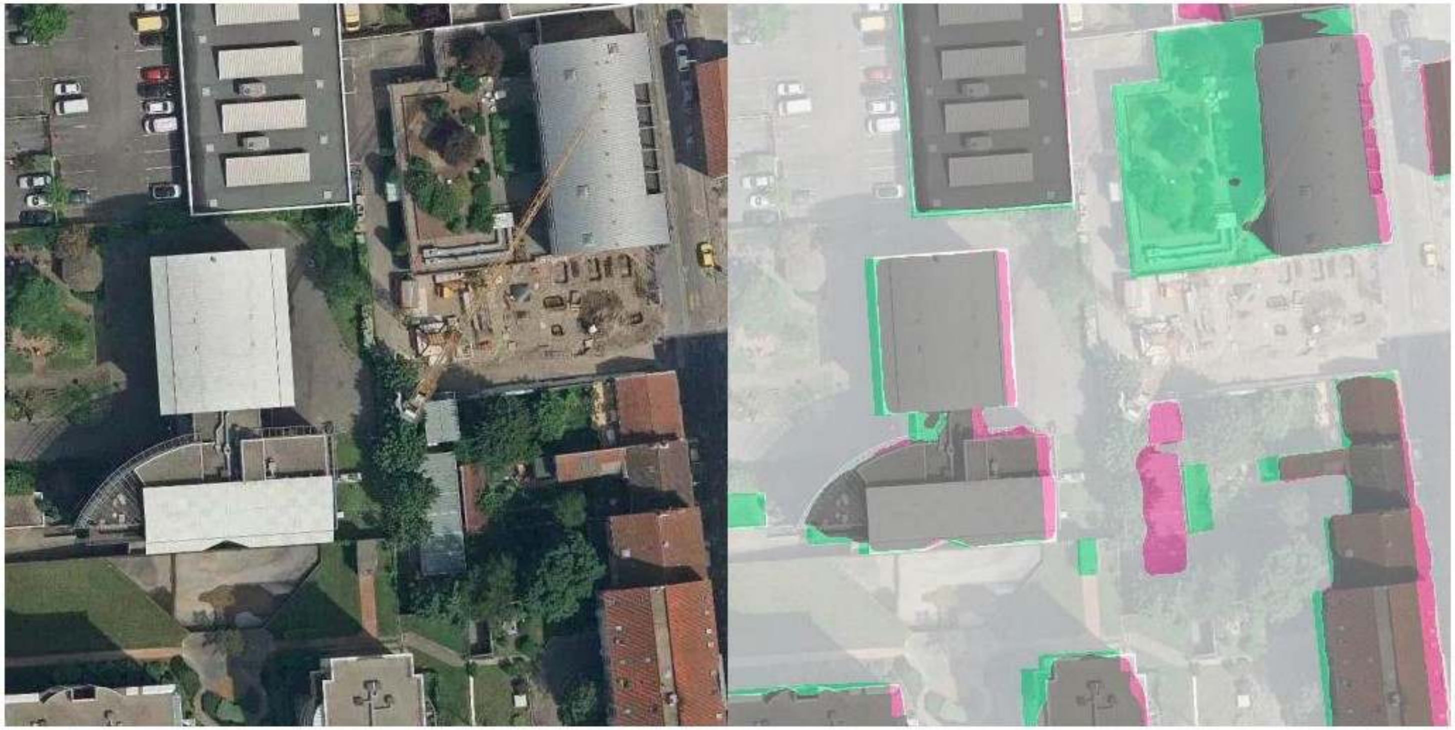
Mixed resolutions, bands, and sensors imagery

Labels (building, roads, vegetation…) with pixel accuracy

No significant progress on this since @SOTM 2018 !

TODAY 15 :00 : Machine Learning DataSet BoF in Kleiner Horsaal

From OpenData to OpenDataSet



https://github.com/datapink/robosat.pink/blob/master/docs/from_opendata_to_opendataset.md

Surface based semantic Loss

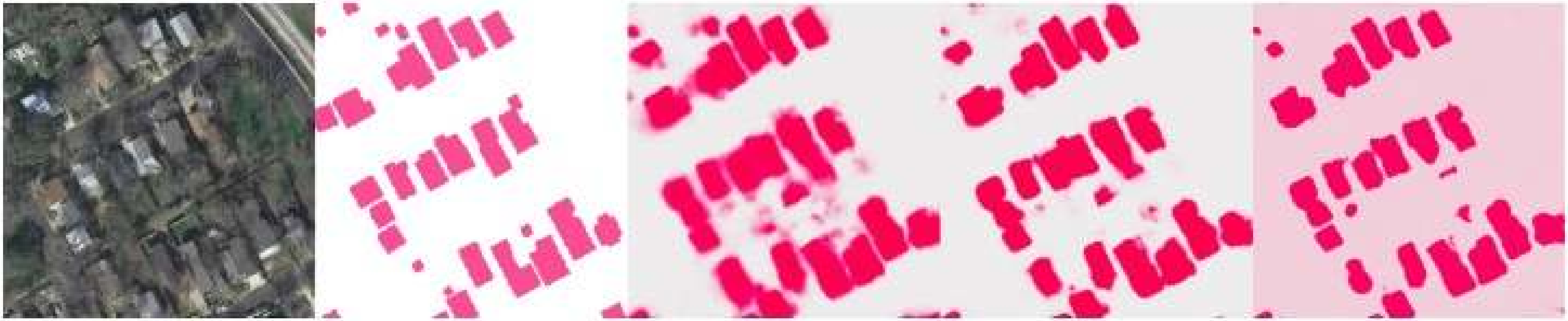
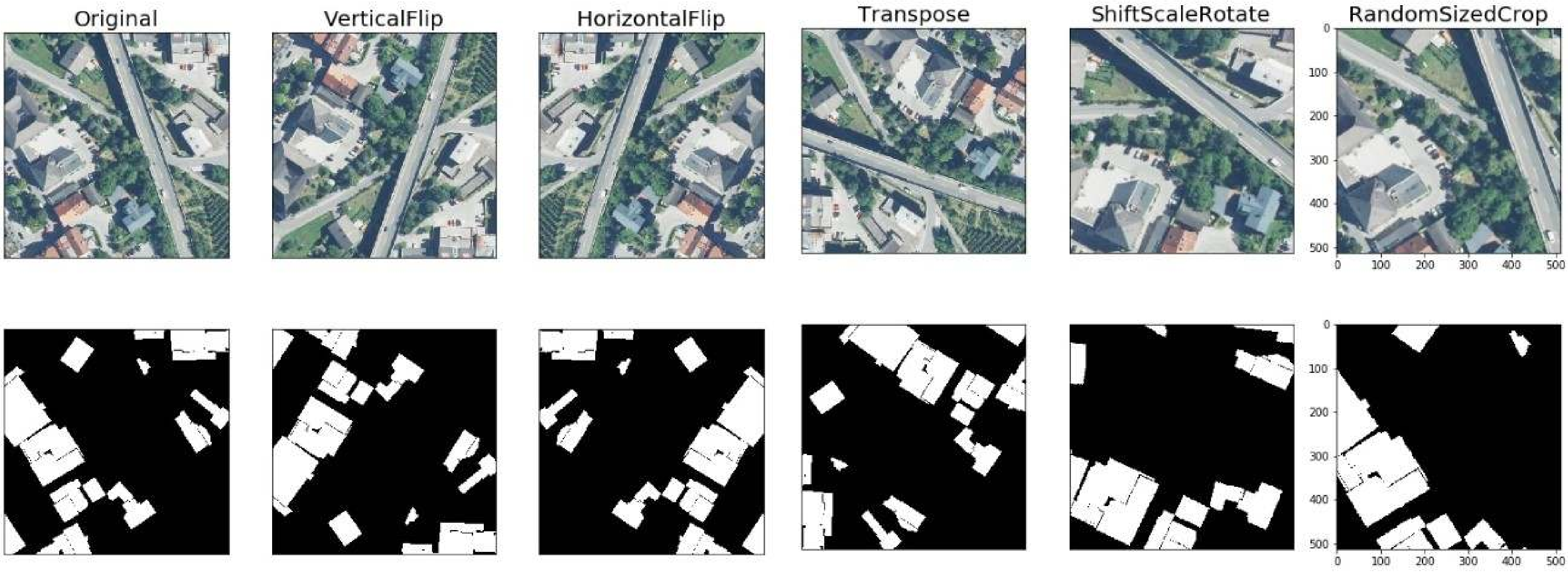


Image Label Cross Entropy mIoU Lovasz

http://www.cs.toronto.edu/~wenjie/papers/iccv17/mattyus_etal_iccv17.pdf
http://www.cs.umanitoba.ca/~ywang/papers/isvc16.pdf
https://arxiv.org/abs/1705.08790

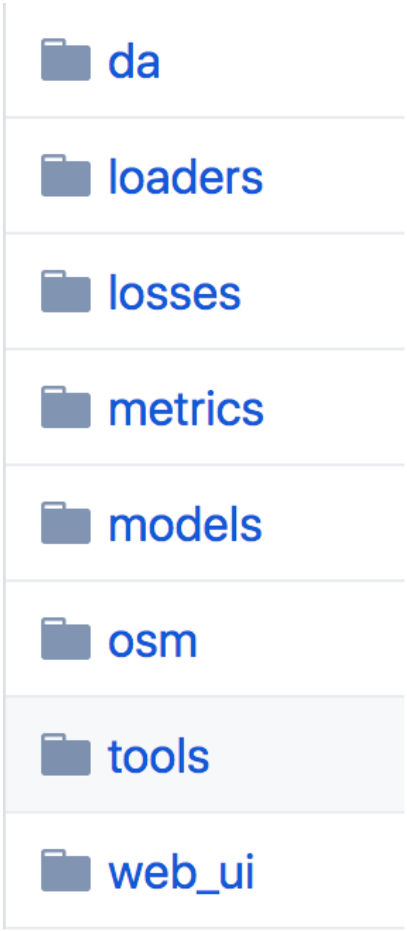
Data Augmentations



Original VerticalFlip HorizontalFlip Transpose ShiftScaleRotate RandomSizedCrop

https://arxiv.org/abs/1809.06839
https://github.com/albu/albumentations

More than an application, an easy to extent framework



da

loaders

losses

metrics

models

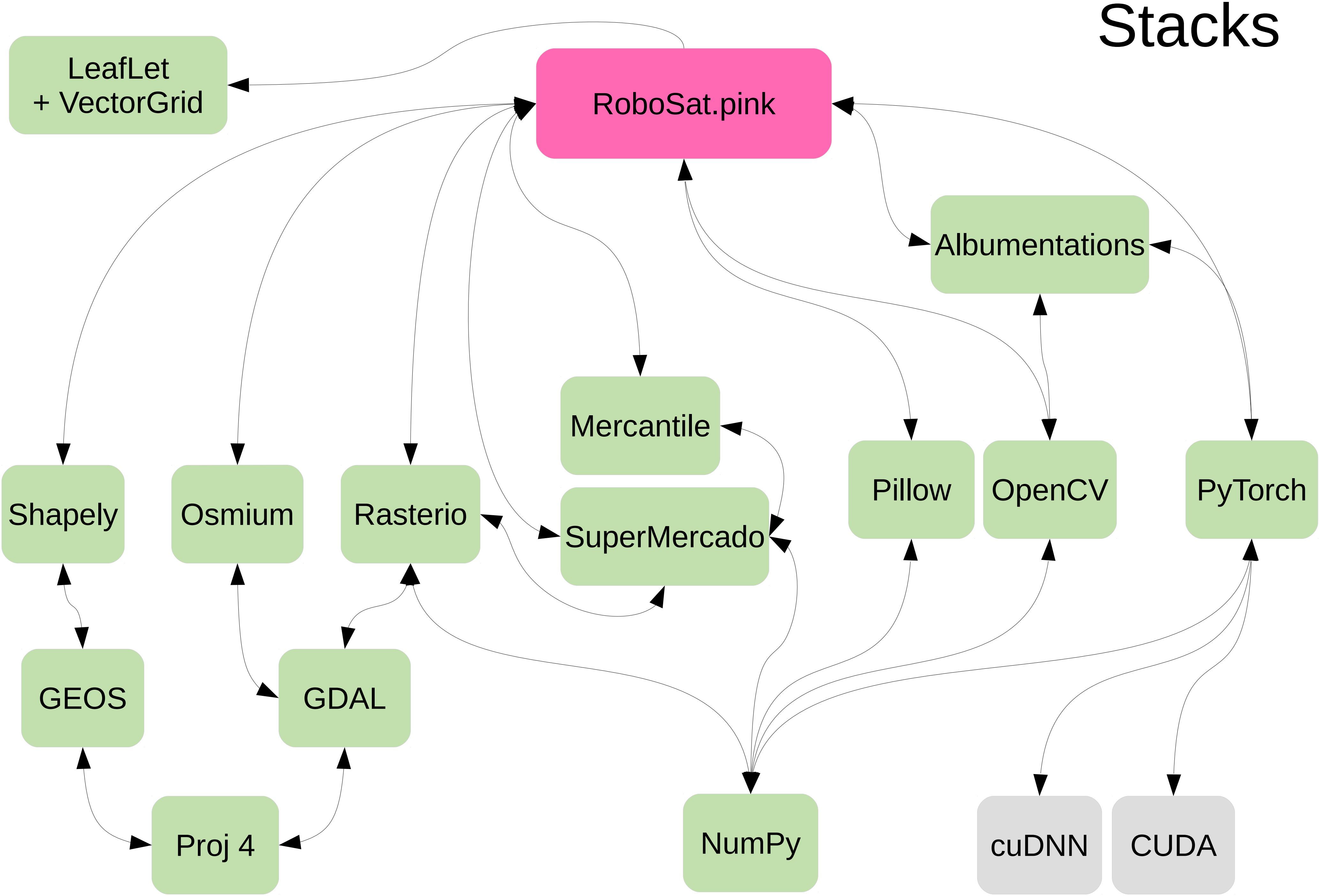
osm

tools

web_ui

https://github.com/datapink/robosat.pink/blob/master/docs/extensibility_by_design.md

Stacks



LeafLet
+ VectorGrid

RoboSat.pink

Albumentations

Mercantile

SuperMercado

Shapely

Osmium

Rasterio

Pillow

OpenCV

PyTorch

GEOS

GDAL

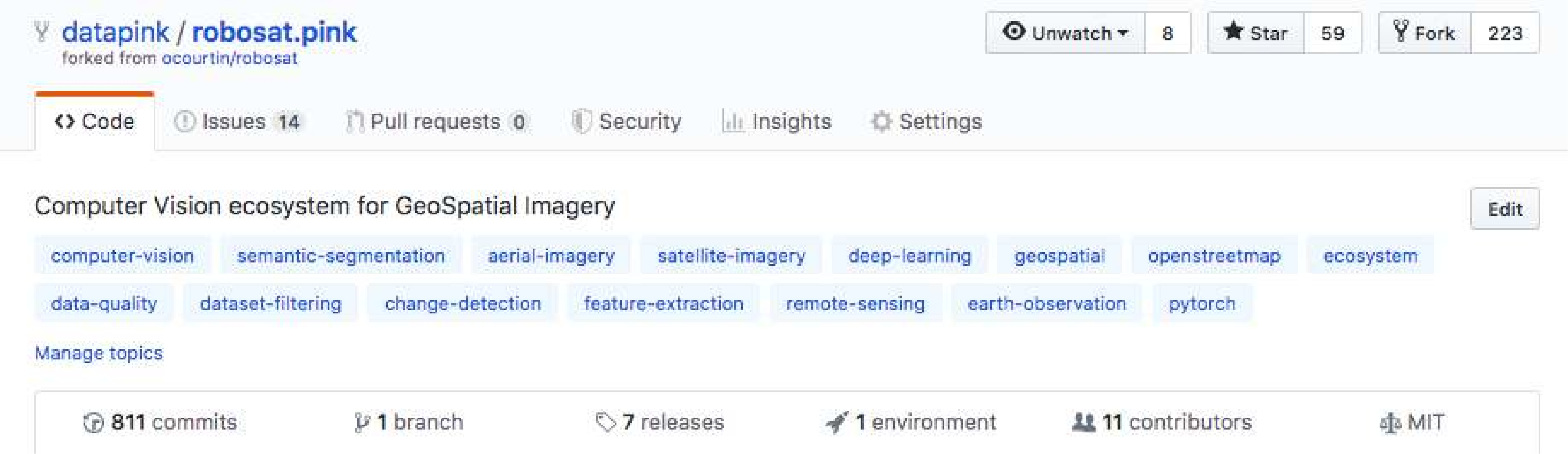
Proj 4

NumPy

cuDNN

CUDA

Open Source



datapink / robosat.pink

forked from ocourtin/robosat

Computer Vision ecosystem for GeoSpatial Imagery

computer-vision semantic-segmentation aerial-imagery satellite-imagery deep-learning geospatial openstreetmap ecosystem data-quality dataset-filtering change-detection feature-extraction remote-sensing earth-observation pytorch

Manage topics

811 commits · 1 branch · 7 releases · 1 environment · 11 contributors · MIT

Request For Funding

- Increase (again) prediction accuracy :
    - on low resolution imagery
    - even with few labels
    - feature extraction when they are (really) close
    - with multibands and Data Fusion

- Add support for :
    - MultiClass
    - Linear features extraction
    - PointCloud data support
    - StreetView Imagery
    - Time Series Imagery
    - OSM editor integration

- Improve (again) performances

https://github.com/datapink/robosat.pink#requests-for-funding

Few performances Metrics

16 tiles = 4 Retina Tiles = 1Mp

rsp train	~5 Mp/s, -per epoch-
rsp tile	~5 Mp/s
rsp predict	~10 Mp/s
rsp compare	~50 Mp/s
rsp rasterize	~50 Mp/s
rsp vectorize	~50 Mp/s

8 cores CPU, single GPU (either RTX or V100), SSD

How to scale it, or improve it again ?

rsp train	add more GPU, reduce dataset redundancy, improve model, loss or optimizer
rsp tile	add more CPU use raster compression
rsp predict	export model to ONNX or JIT, then use an high performance inference solution.

Cost Effective GPU Servers

RTX

ThreadRipper

Own server

OVH
Innovation for Freedom

« Someone else servers »

Why performances matters ?

- Playful and Human Learning

- Time and money saver

- No Planet B

Why using DeepLearning for Mapping ?

Easy to spotify at scale inconsistencies beetwen two datasets

If you provide good labels on a imagery,
infere at scale on similar new imageries

So a time and money saver

Why using RoboSat.pink ?

OSM easy integration (XYZ, Osmium, Leaflet...)

Build-in WebUI

High Performances

Easy to deploy

GIS Standards compliancy

Handle MultiBands Imagery and DataFusion

Accurate (state of art Computer Vision)

Extensible by design

Open Source

DATA PINK

@data_pink

www.datapink.com

Take Away

- Industrial FOSS4G state of art GeoSpatial Imagery framework available

- Performances already OK to use it for region / small country ,
  even on cheap GPU server, scale if you provide better hardware…

- No need anymore to be a Computer Vision expert to use it

- Plain OpenData can be use to train accurate model

- Pixel accurate labelled training OpenDataSet will be a game changer
  => BoF 15:00 Kleiner Horsall

RoboSat.pink powered by @data_pink