

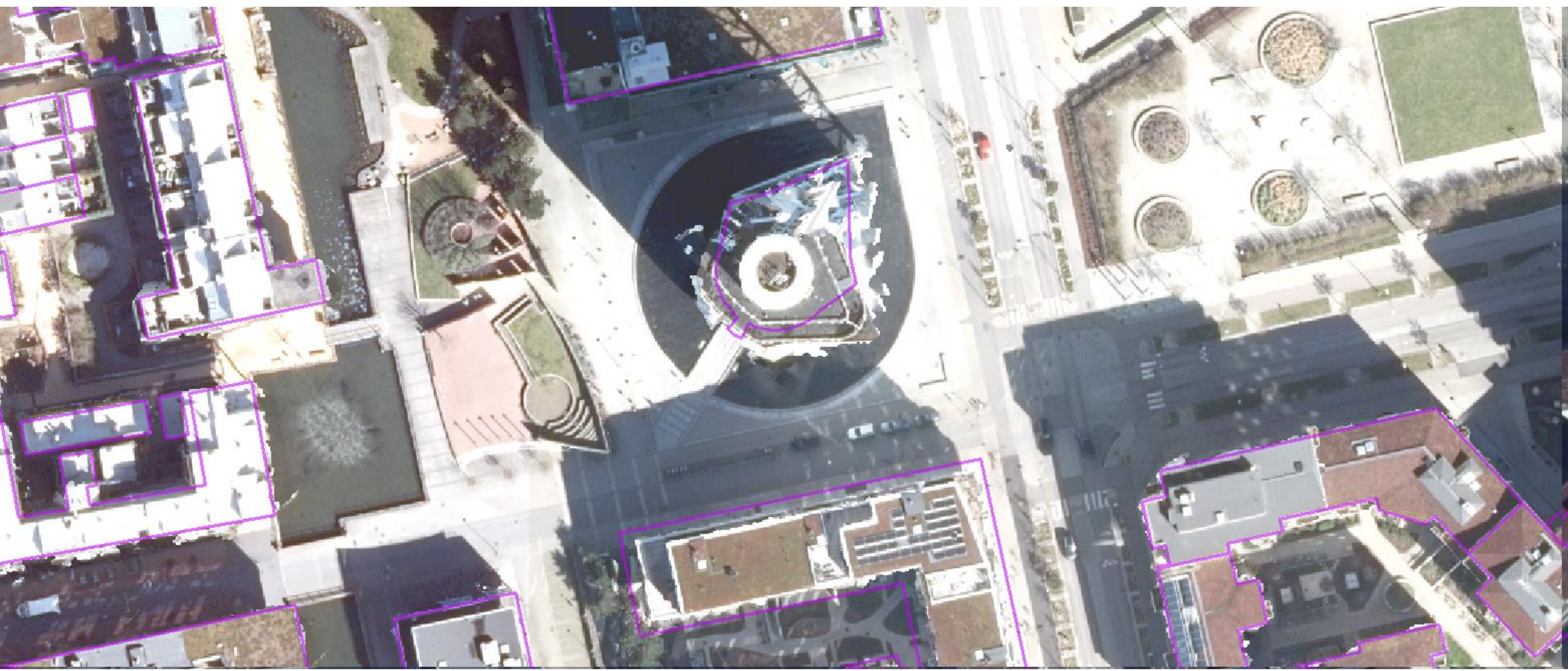
MACHINE LEARNING AND AI AT LANTMÄTERIET

ERIK NILSSON 2024-11-14

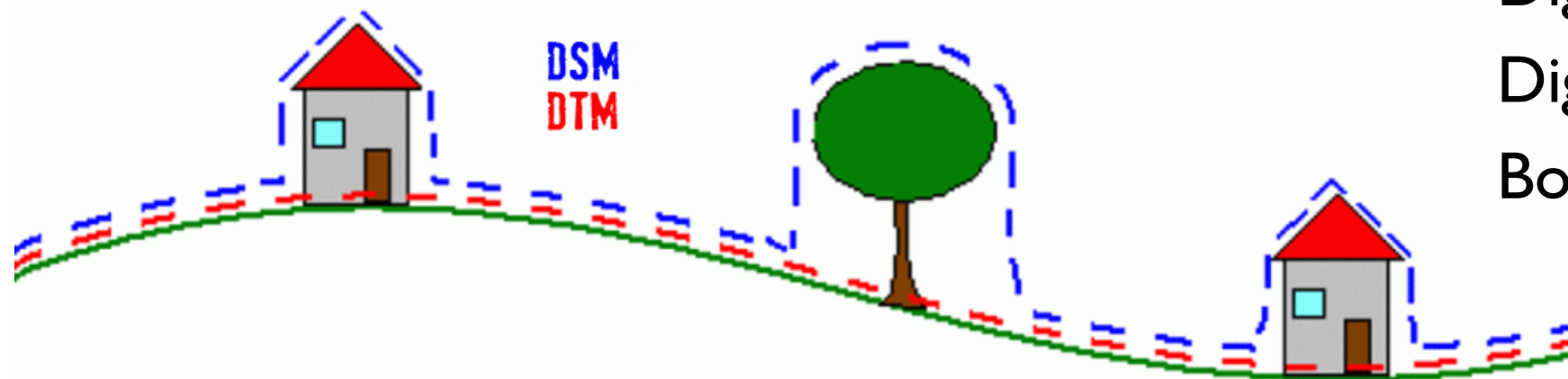
BACKGROUND

- Lantmäteriet started testing ML for mapping purposes in 2018
- In 2022 we started a project to use ML to find buildings to support our mappers
- In march 2024 we started using ML generated buildings as a background layer to help our mappers
- During 2024 we have started a project to detect forest roads. Hopefully we will start using it in production in the beginning of 2025.
- We have a team consisting of 6 persons working with ML but only one is doing it (almost) full time and the rest about 15-25% each.

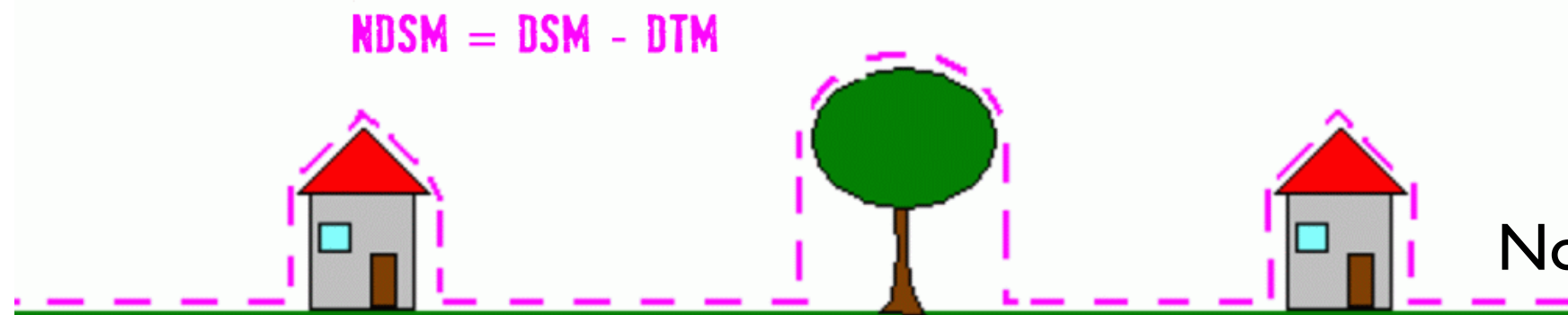
ORTOPHOTO VS TRUE ORTOPHOTO



HEIGHT MODELS



Digital Surface Model (DSM)
Digital Terrain Model (DTM)
Both are "height over sea"

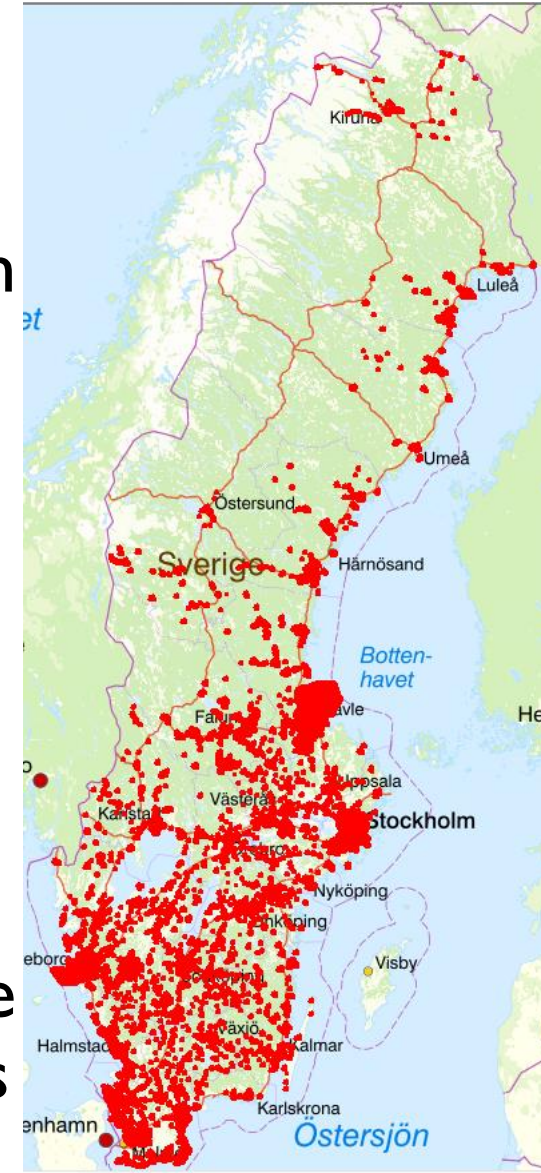


Normalized surface model =
nDSM = DSM - DTM
"height over ground"

BUILDINGS

DATA FOR TRAINING

- The building data we can use is, in general, collected by the municipalities in the city centers and by Lantmäteriet in all other areas.
- The quality and actuality varies a lot. Some municipalities report very small buildings and Lantmäteriet does not collect or make updates on buildings or changes less than (around) 15-20 m²
- Lantmäteriets building data is, in general, not good enough to use for training.
- Sweden is 1600 km from north to south covering 6 climate zones and aerial photos can be both with or without leafes



TRAINING DATA

- We have randomly obtained tiles with data. These have been manually controlled and approved/nor approved.
- Tiles were also manually created to get training data for different types of cases like bridges, ports, camping areas etc. But also if our mappers have found strange results when evaluating our results.
- We have also mapped 10 areas 2,5*2,5 km very accurate and randomly created a lot of tiles from that areas.
- We have tried rotation and mirroring images to create more training data

TRAINING

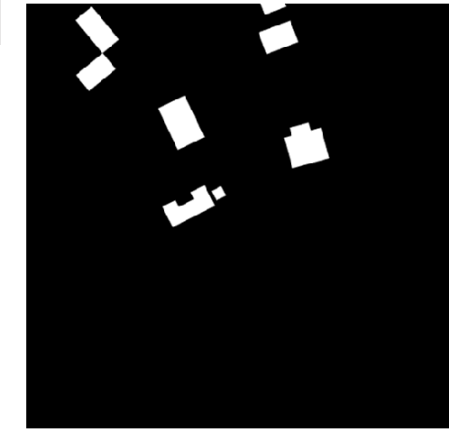
- We use tiles of size 512* 512 pixels. The resolution of the images we use is 0.25m så every tile covers an area of 128*128 m on the ground.
- We generate tif-images of all training data. Then we can use it for training even if new orthophotos are made or the buildings change in our databases.
- We have about 30 000 tiles.
- The base data in all our training has been "True orthos" with R, G, B and IR- channels and rasterized buildings.
- We have tried several different height models like DSM, DSM and DTM and nDSM. We found that nDSM was the best in our tests.
- Our trainings takes about 4-5 days each
- When training takes such long time you need to be smart and patient when planning the trainings. Just change one thing at a time to be able to track what has effect on the results.



+



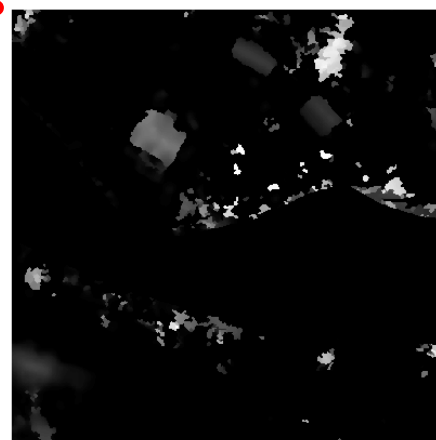
=



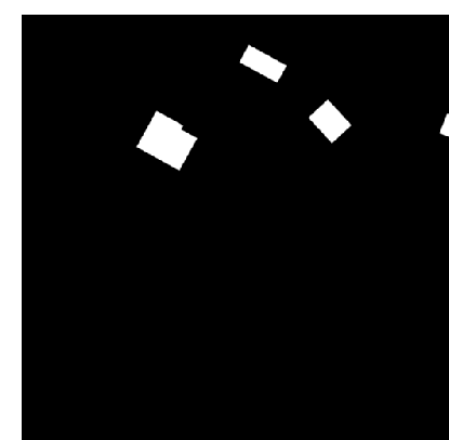
Training



+



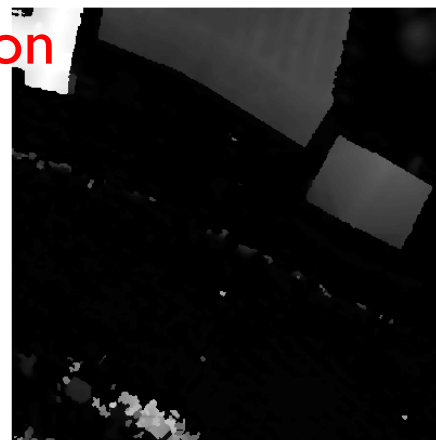
=



Detection






+




+ML =


?


DETECT

-  Object found by ML
-  Ground truth
-  Area where ML and GT match



 Object found by ML

 Ground truth

 Area where ML and GT match



EVALUATION

Blad	Precision	Recall	F1_Score	AP	True_Pos	False_Pos	False_Neg
T624_42_2500_Kullerod	0,8	0,78	0,79	1,02	117	30	33
T648_28_2525_Hunnebostrand	0,91	0,8	0,85	1,14	1925	182	473
T672_53_0025_Falun	0,88	0,72	0,79	1,23	2349	315	915
T684_57_0000_Ivarsmyran	0,61	0,66	0,64	0,93	35	22	18
T728_84_7500_Lovskar	0,83	0,6	0,7	1,38	464	96	311
Summa					4890	645	1750
Totalt	0,88	0,74	0,80				
Max	0,91	0,8	0,85	1,38			
Min	0,61	0,6	0,64	0,93			
Medel	0,81	0,71	0,75	1,14			
	0,12	0,08	0,08	0,18			

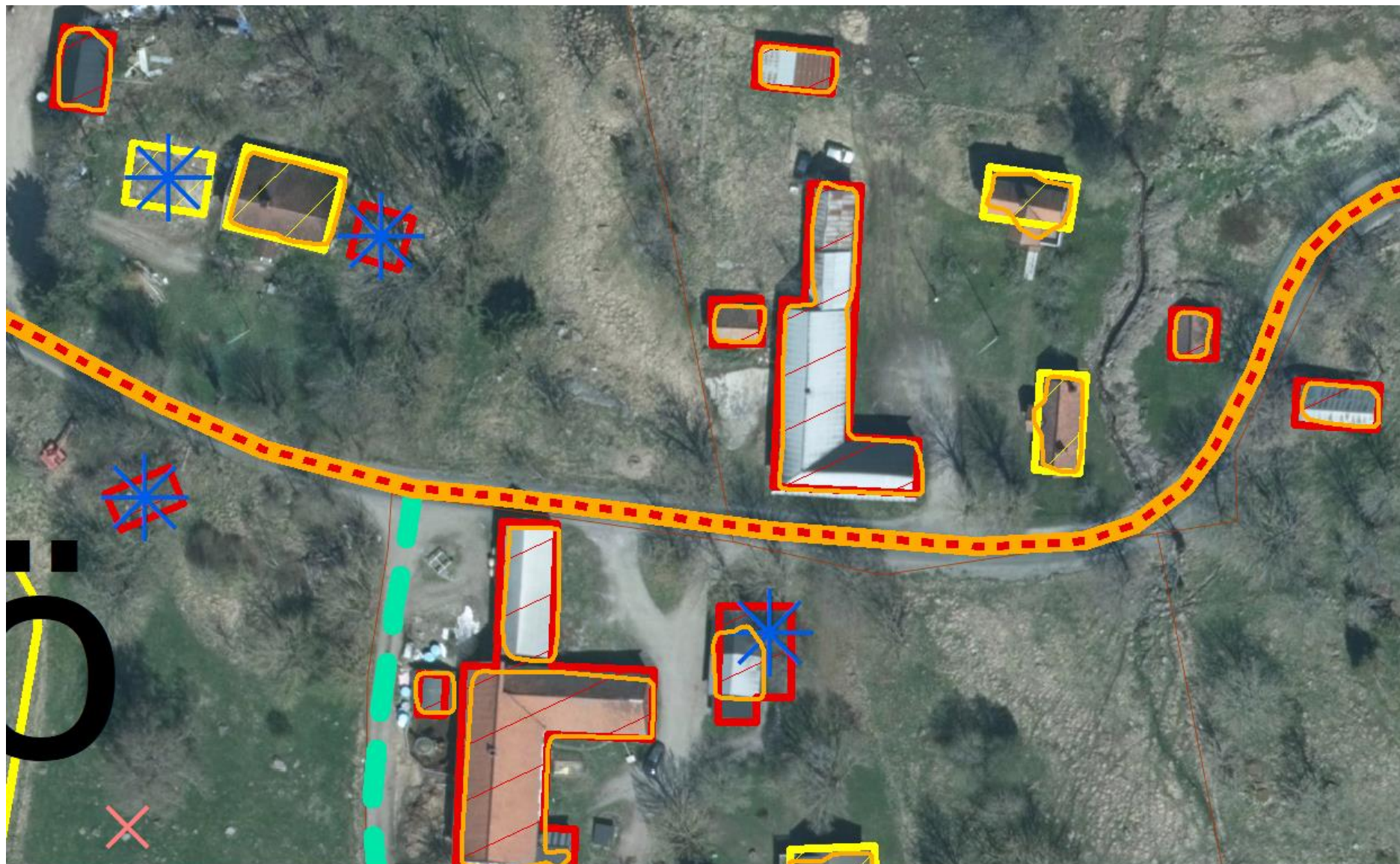
This means that 88% of the buildings found are real buildings and that we found 74% of the buildings that actually exists.

USAGE IN MAPPING

- Since the beginning of march 2024 we use buildings found by ML as a background layer for our topographic mappers
- The mapper decides if a change will be made in the database
- Tests indicates that this helps us find 1-2% more buildings than without support from ML buildings

ARCMAP – POLYGONS AND HIGHLIGHTING DIFFERENCES

- ML_byggnad
- ML_byggnad_polygon
 -
- ML_byggnad_punkt
- ML_byggnad_diff



Forest roads

FOREST ROADS

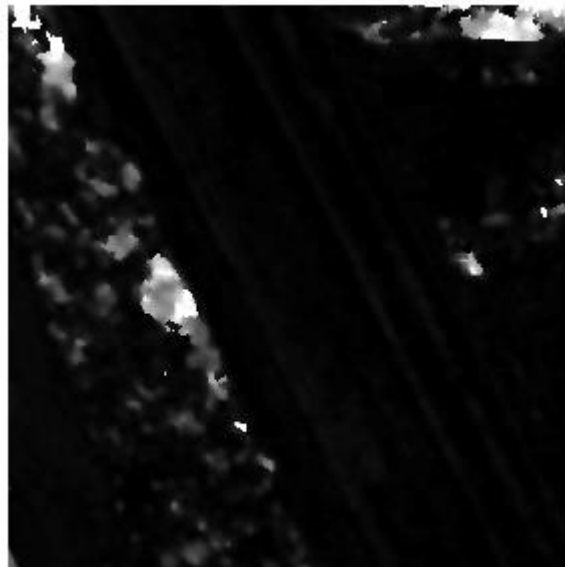
- This is ongoing work
- The aim is to create road lines using ML that are so good that the geometry can be copied by the mappers into our road database
- 2 focus areas - Areas with wind turbines and newly built roads
- We use roads from National Road Database as training data. This data is good to use as it is.

TRAINING DATA ROADS

True orthophoto



nDSM
(DSM- DTM)



Rasterized roads

From National Road Database



RESULTS

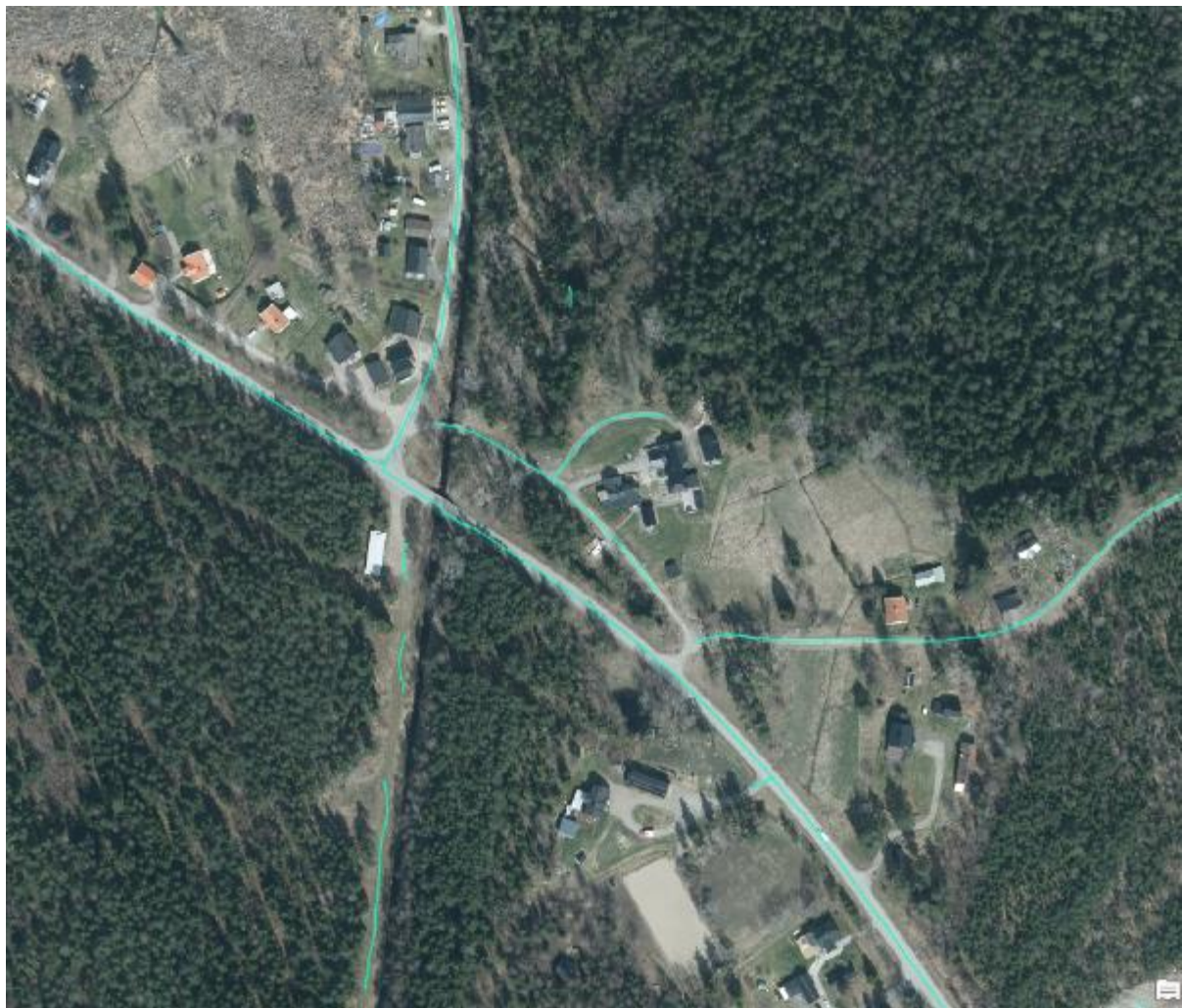


Figure 1: Detected roads

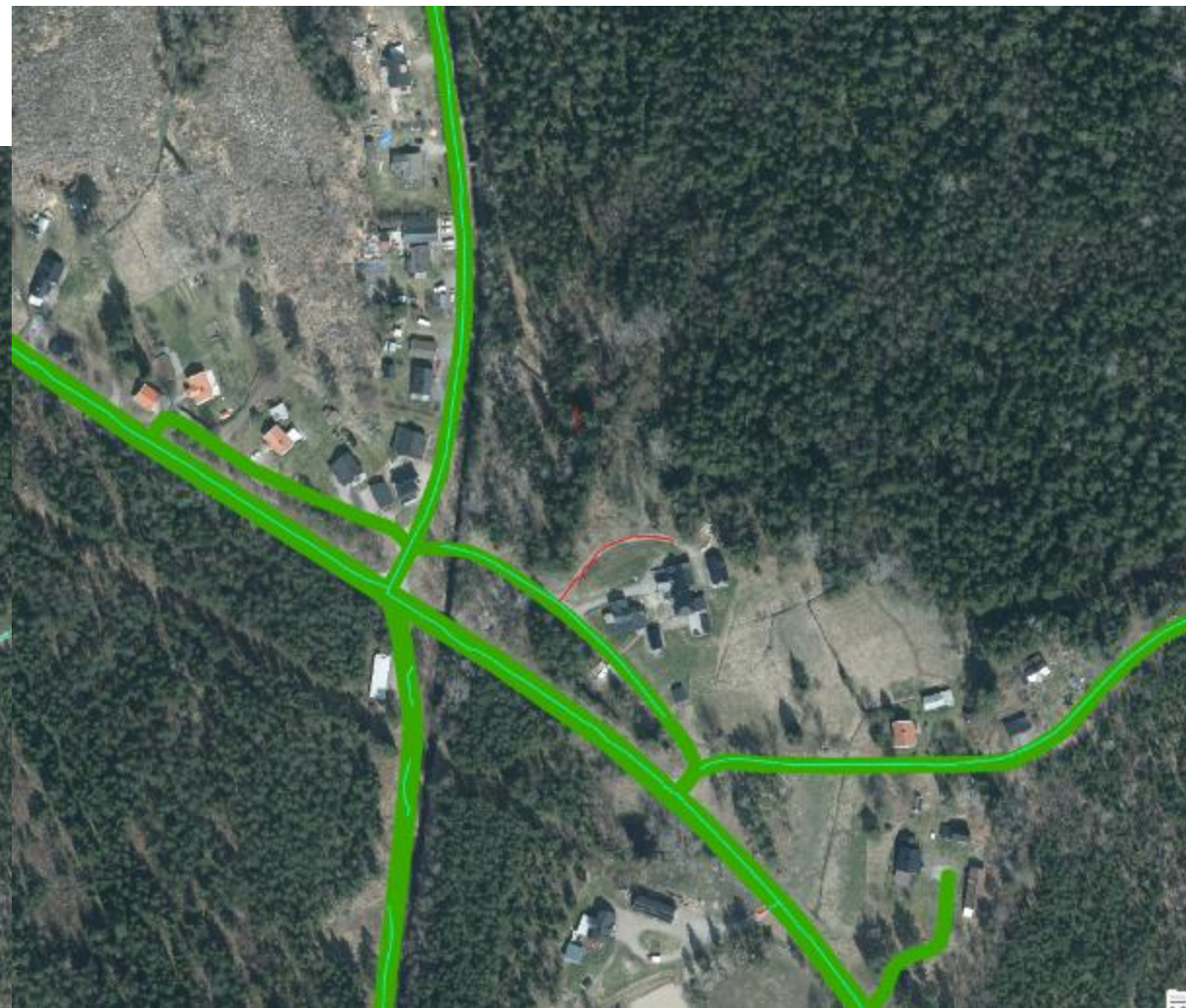


Figure 2: Roads from National Road Database are green. Road from figure 1 is still visible. Roads not existing in National Road Database is Red.

POST PROCESSING

- Based on a method from ESRI
- Buffers and connects segments that are close to each other
- Red line is the result after post processing



RESULTS

- **Red lines** are detected roads that coincides with National Road Database
- **Blue lines** are detected roads that not coincides with National Road Database
- **Green lines** are roads from National Road Database which has not been found by ML



TECHNICAL ENVIRONMENT

- NVIDIA A10 Tensor Core GPU on Vmware-platform
- Linux
- Python and ML framework PyTorch.
- UNET
- Tversky loss / Topological loss
- FME for data management and ArcGIS Pro for post processing of roads
- The technical environment is placed in a virtual "container" which makes it easy to move to another computers.

OTHER AI INITIATIVES

- We have a chat bot for customer service
- A lot of work was made with text-recognition on old documents from our archive. The work was stopped after complains from our legal department.

WHAT WE HAVE LEARNED

- You cant create a good model from bad data
- We have used photogrametric surface models but think we would have better results with surface models from lidar-data. Lidar-data and ortho photos collected at the same time are on top of our wish list.
- Takes a lot of time to train and evaluate models
- Be patient!
- Needs good performace on the hardware
- Make efforts to optimize the whole training process to make optimal use of the hardware
- The legal department can have different opinions than the business on how to use data and AI

TACK! VI FINNS PÅ...

WEBBPLATS

www.lantmateriet.se

LINKEDIN

www.linkedin.com/company/lantmateriet

FACEBOOK

www.facebook.com/lantmateriet

INSTAGRAM

www.instagram.com/lantmateriet

KONTAKT

kundcenter@lm.se

TELEFON

0771-63 63 63

LANTMÄTERIET

