Doc. ref: ESA AO/1-10552/21/I-DT

Self-supervised and Zero-shot learning for extracting feature representations in Earth Observation based forest monitoring

Ridvan Salih Kuzu (DLR)

Oleg Antropov (VTT)

Use-Case Overview

- **Severe forest disturbances significantly influence forest ecosystem services and biodiversity**. Accurate delineation and early detection of areas affected by disturbances are critical for estimating extent of damage, assessing economical influence and guiding forest management activities.
- In this Use-case we have concentrated on **natural and human-induced disturbances over boreal forests.** Such disturbances include typical forest management operations such as forest clearcutting and thinning, as well as various kinds of naturally occurring damages, such as damage because of windstorm or heavy snow load.

Timely monitoring of such damages requires the use of not only optical but preferably imaging radar imagery (or rather both image sources for improved delineation) and includes image time series approaches

In this use-case, we were relying on optical Sentinel-2 and C-band SAR Sentinel-1 images

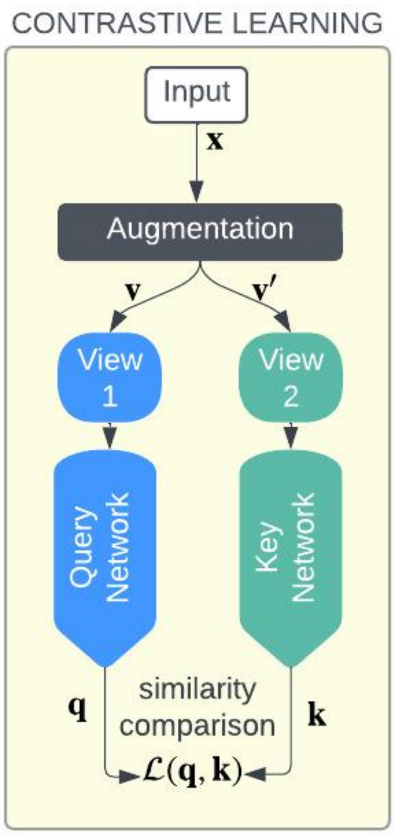
Use-Case Overview

- Social / economic / scientific impact:
  - Timely detection of forest windthrow helps to guide forest removal work and assess economic losses, and timely detection and delineation of snow damaged forest allows to plan forest removals to avoid insect breaks.
  - Combining accurate and timely statistics on forest state and forest removal with forest growing models enables improved forest monitoring systems utilizing satellite images
  - Forest management activities such as forest clearcutting and thinning are performed on a regular basis over practically all managed forest areas. Effective wide area monitoring is needed for quality control and effective accounting of forest resources on local, regional and national levels.

- Data sources and existing ground-truth data:
  - Satellite image data will be represented by Sentinel-1 and Sentinel-2 images.
  - Reference data on forest disturbances are represented by forest owner reports on forest management operations, as well as on analysis of MS-NFI and ALS datasets to derive information on intact forest stands after disturbance events.
- EO users:
  - Developed approaches will be demonstrated using ESA Forestry TEP to a wide community of academic and forestry users including forest owners, as well as using conference presentations and journal articles.
  - Dedicated workshop for forest company representatives and forest remote sensing academic community will be arranged by the end of the project in connection with UC1 and UC5 to communicate developed methods.

DLR

Background and Motivation



Advantages of unsupervised learning for forest change:

- Capability to analyze large amounts of data quickly and efficiently
- Minimization of manual labor and time required for training data preparation

Limitations of the existing pretrained models:

- General purpose
- Prioritizing different problems other than focusing forest change
  - **BigEarthNet**: land cover classification, land use analysis, and vegetation monitoring.

Proposed solution

- Self-supervised Learning (SSL) in a contrastive manner on Forest-Dense regions
- Transferring the weights to different forest monitoring use cases using zero-shot learning.

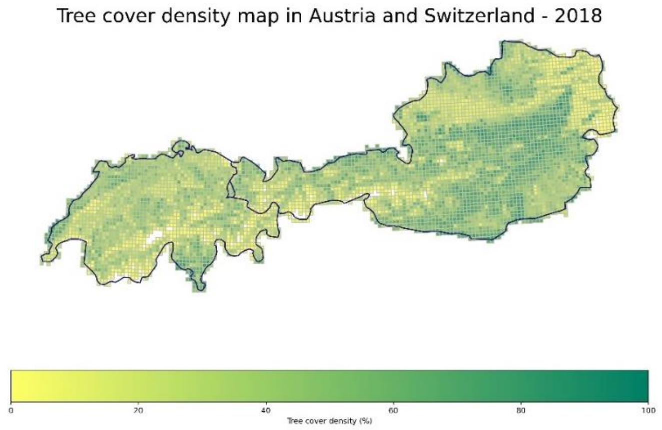
Datasets

Dataset for Self-supervised Learning:

Forest Dense Area Dataset

- Austria: **3417 patches**, 256x256
- Switzerland: **1722 patches**, 256x256

~10 times smaller than BigEarthNet Dataset



Dataset for Zero-shot Anomaly Detection:

Forest Anomaly Dataset

- Snow Damage
- Windstorm Damage
- Clearcut Damage

Boreal forest snow damage

- For snow damage, the dataset consists of Sentinel-1 images taken between November 2017 and March 2018. The dataset is challenging as there may be other types of change in addition to snow damage in the same region over the same period of time. The dataset also includes a forest mask and reference maps indicating the areas affected by snow damage and those that remain intact.

Windstorm damage

- A severe windstorm occurred on June 22, 2021. The dataset includes one Sentinel-2 image taken before and one taken shortly after the event, enabling Sentinel-2 change detection to be performed. Additionally, several Sentinel-1 images taken both before and after the windstorm event are provided for analysis.

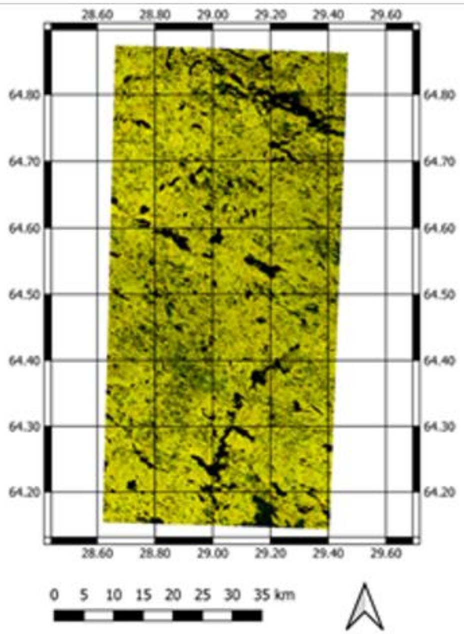
Clear cutting damage

- Three Sentinel-2 and three Sentinel-1 images from 2015, 2016, and 2017, respectively. The dataset is aimed to detect clearcutting/forest thinning change detection.

Datasets

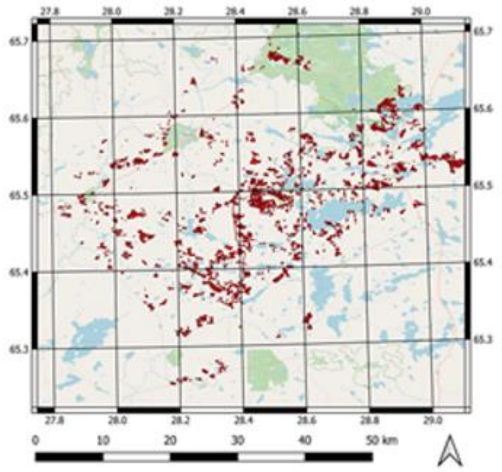
Boreal forest snow damage using Sentinel-1 images

- Severe snow-load damage took place in North-Eastern Finland during winter 2017-2018 season. The damages were fairly well documented.
- EO dataset is represented by time series of Sentinel-1 images taken between November 2017 and March 2018.
- Detecting damage caused by snow can be a challenging use case, as there may be other observable changes that take place in the same region during the same time period
- Reference data included forest mask and ground reference indicating logging reports after snow-load damage, as well as sample of intact forest stands.

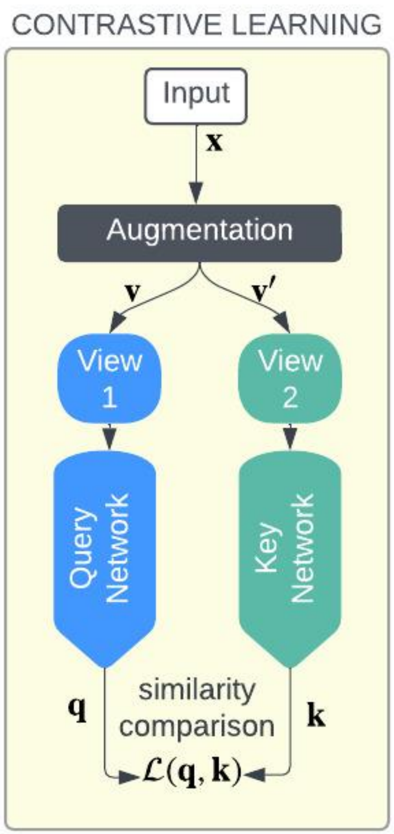


Windstorm damage in boreal forest

- A severe windstorm occurred on June 22, 2021, over Northern Finland, lead to extensive forest damages.
- The EO dataset includes one Sentinel-2 image taken before and one taken shortly after the event, enabling Sentinel-2 change detection to be performed. Additionally, several Sentinel-1 images taken both before and after the windstorm event are provided for analysis.
- Reference data included forest mask and ground reference indicating logging reports after snow-load damage, as well as sample of intact forest stands.



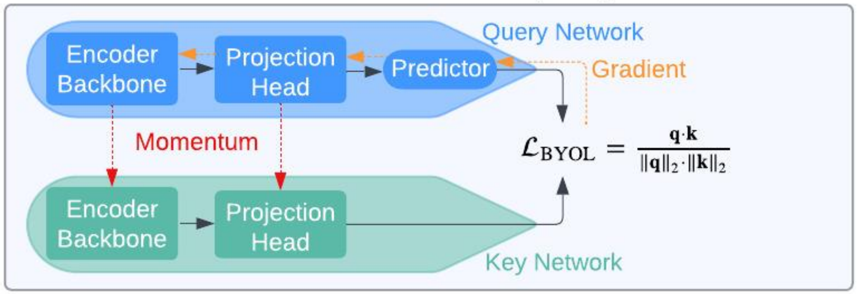
Conceptual Idea of Self-supervised Learning



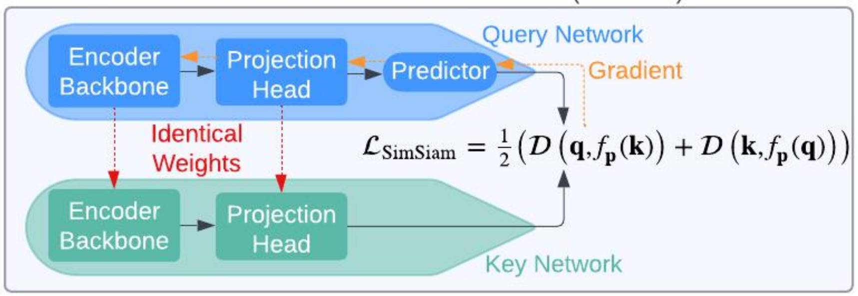
- Self-supervised learning uses unlabeled data to train models without manual annotations.
- Query and key networks are a type of neural network architecture for self-supervised learning.
- Data augmentation creates different "views" of the input data by applying different transformations.
- The query and key is trained to produce similar representations for corresponding views and dissimilar representations for non-corresponding views.
- By learning to associate and disassociate views of data, the query and key networks learn meaningful representations for downstream tasks.

Conceptual Idea of Self-supervised Learning

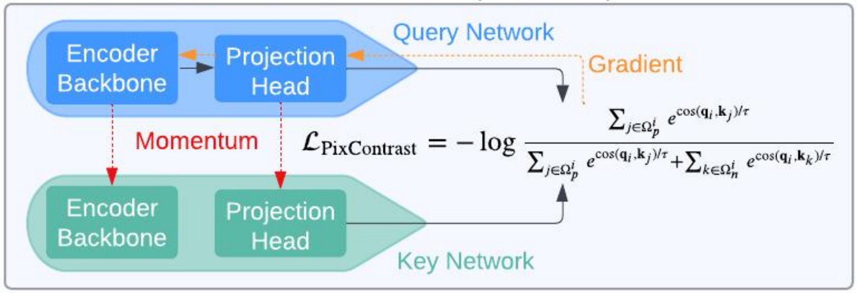
BOOTSTRAP YOUR OWN LATENT LEARNING (BYOL)



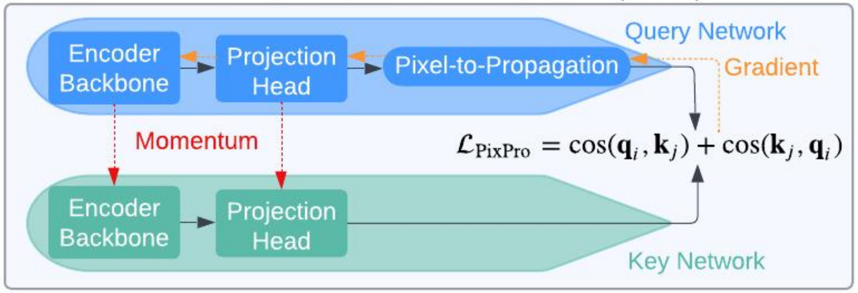
SIMPLE SIAMESE REPRESENTATION LEARNING (SimSiam)



PIXEL-LEVEL CONTRASTIVE LEARNING (PixContrast)

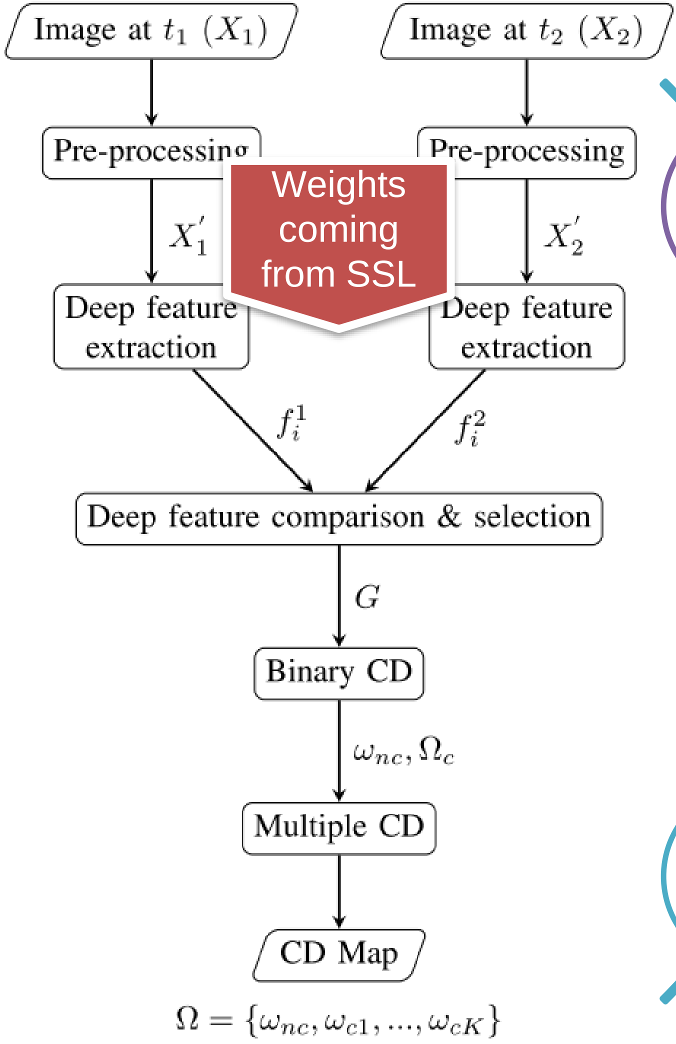


PIXEL-TO-PROPAGATION CONSISTENCY LEARNING (PixPro)



1. Grill, Jean-Bastien, et al. "Bootstrap your own latent-a new approach to self-supervised learning." Advances in neural information processing systems 33 (2020): 21271-21284.
2. Chen, Xinlei, and Kaiming He. "Exploring simple siamese representation learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
3. Xie, Zhenda, et al. "Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Zero-shot Learning based on Deep Change Vector



Given co-registered pre-change and post-change images I_1 and I_2 that represent a scene consisting of a set of pixels Ω, **the goal of DCVA is to group the Ω into two Ω_c** (set of changed pixels) **and Ω_{nc}** (set of unchanged pixels)

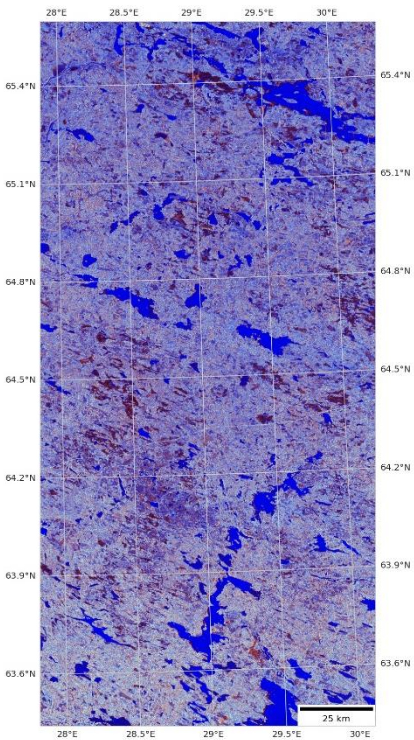
Changed pixels are expected to show higher magnitude of G (change vector) while the unchanged pixels are expected to show lower magnitudes.

Following this, a threshold on the magnitude to group pixels into changed (Ω_c) and unchanged (Ω_{nc}) classes

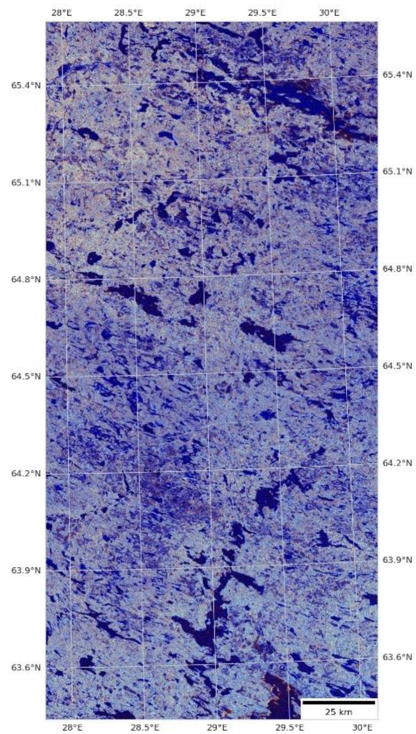
Optionally, changed pixels can be further sub-grouped using any clustering method.

Experiments: Snow Damage

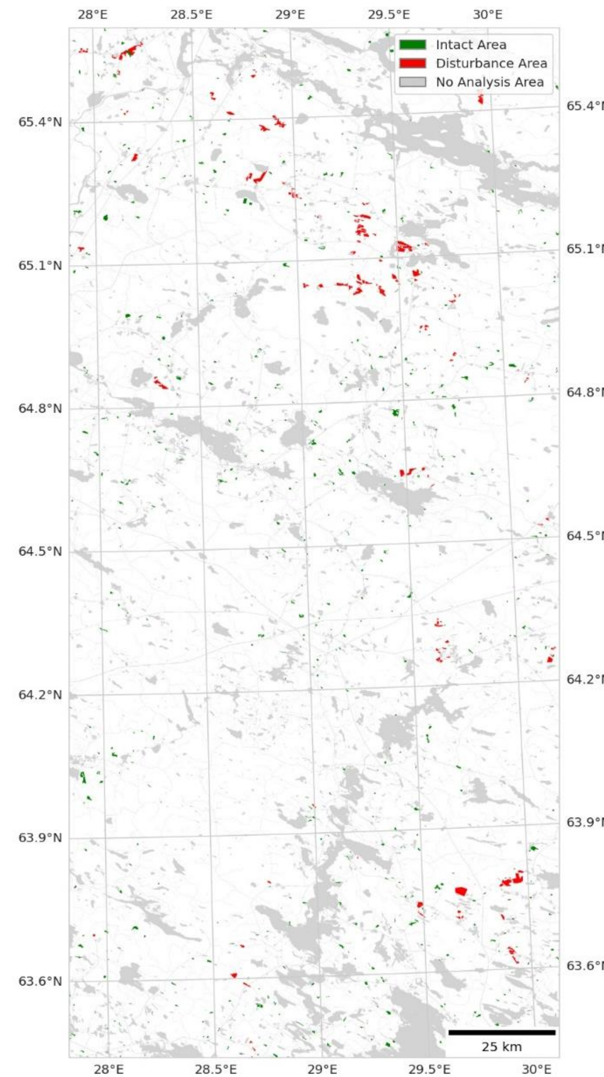
While continuous monitoring is necessary for assessing snow damage, DCVA can only be applied to two images - the before and after event images. As a result, deep change vectors are generated using these two images.



Before the event

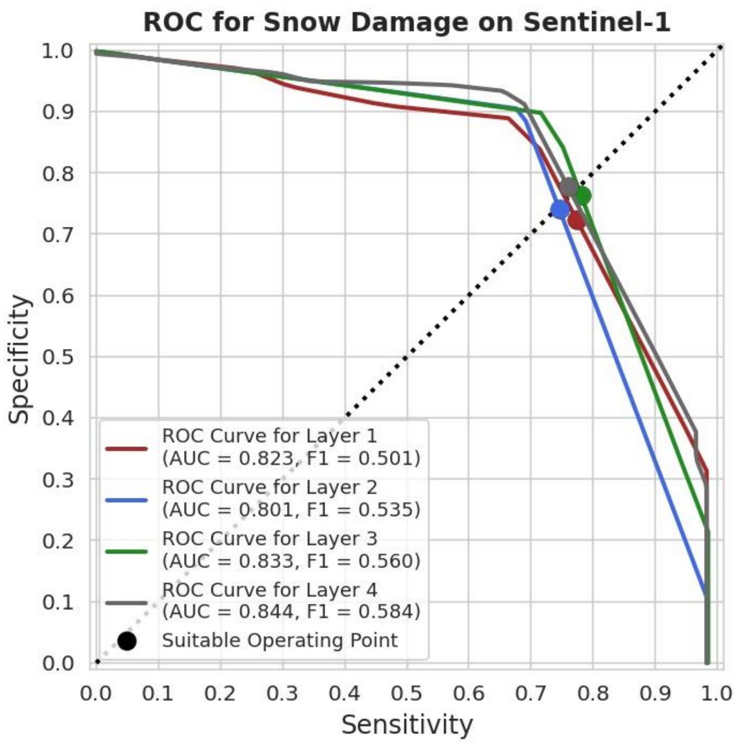


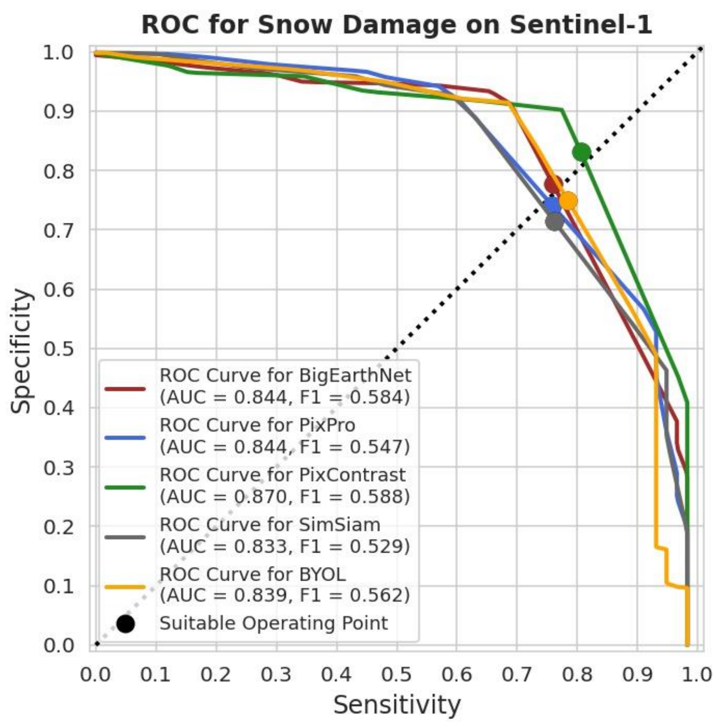
After the event



Ground-truth

Experiments: Snow Damage

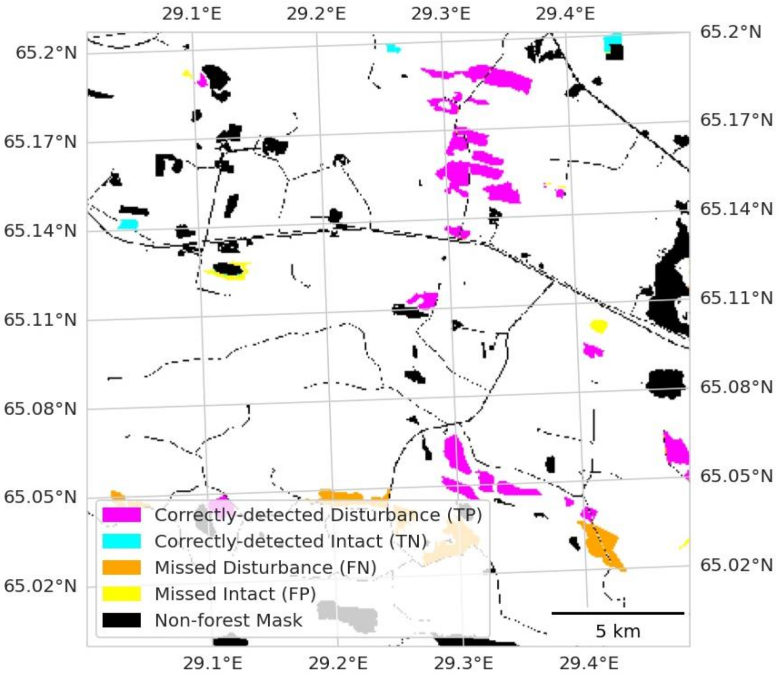


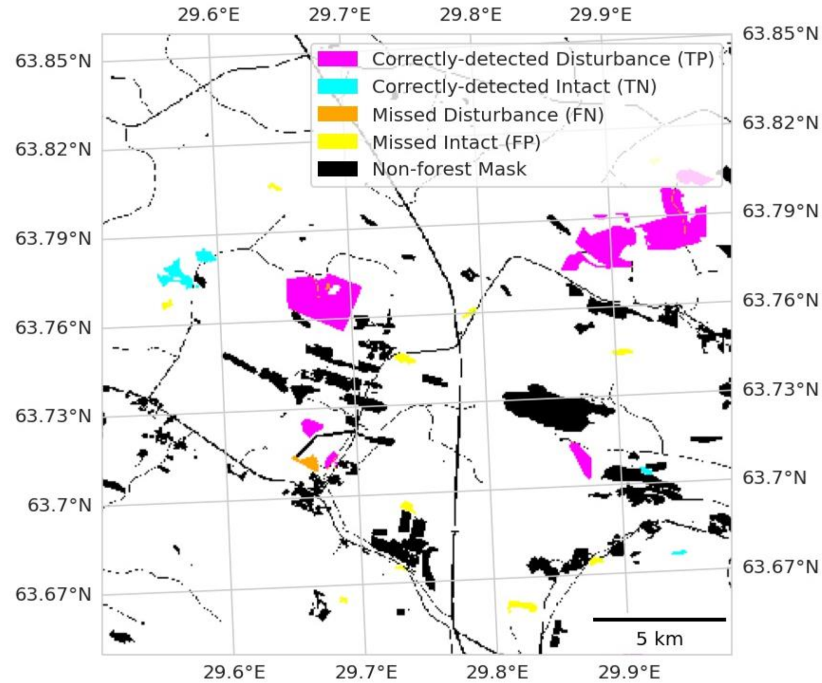


PixContrast weights outperforms the baseline **BigEarthNet** weights
in Snow damage assessment.

Experiments: Snow Damage

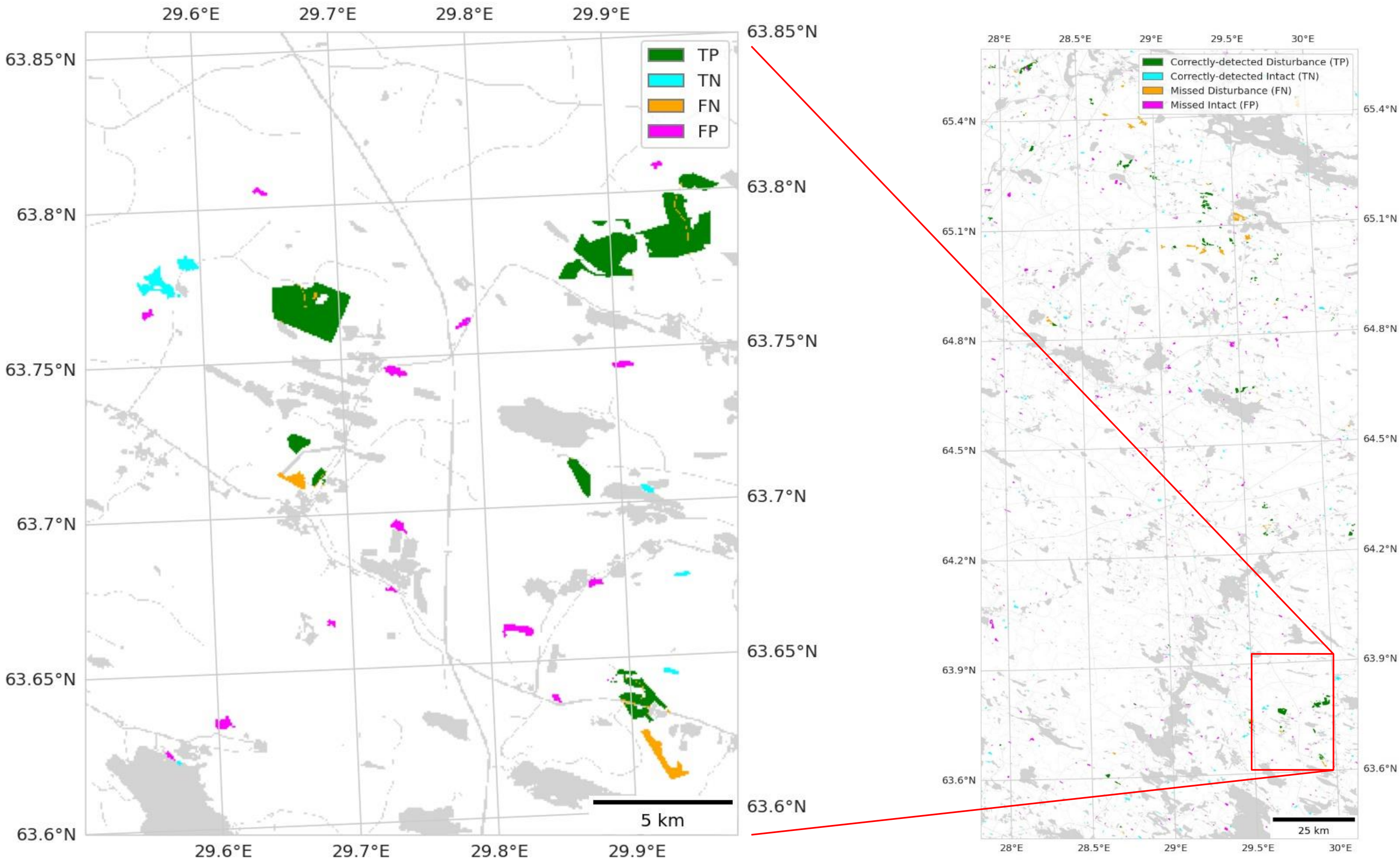
Prediction results when the pretrained Resnet-18 has been utilized as feature extractor:





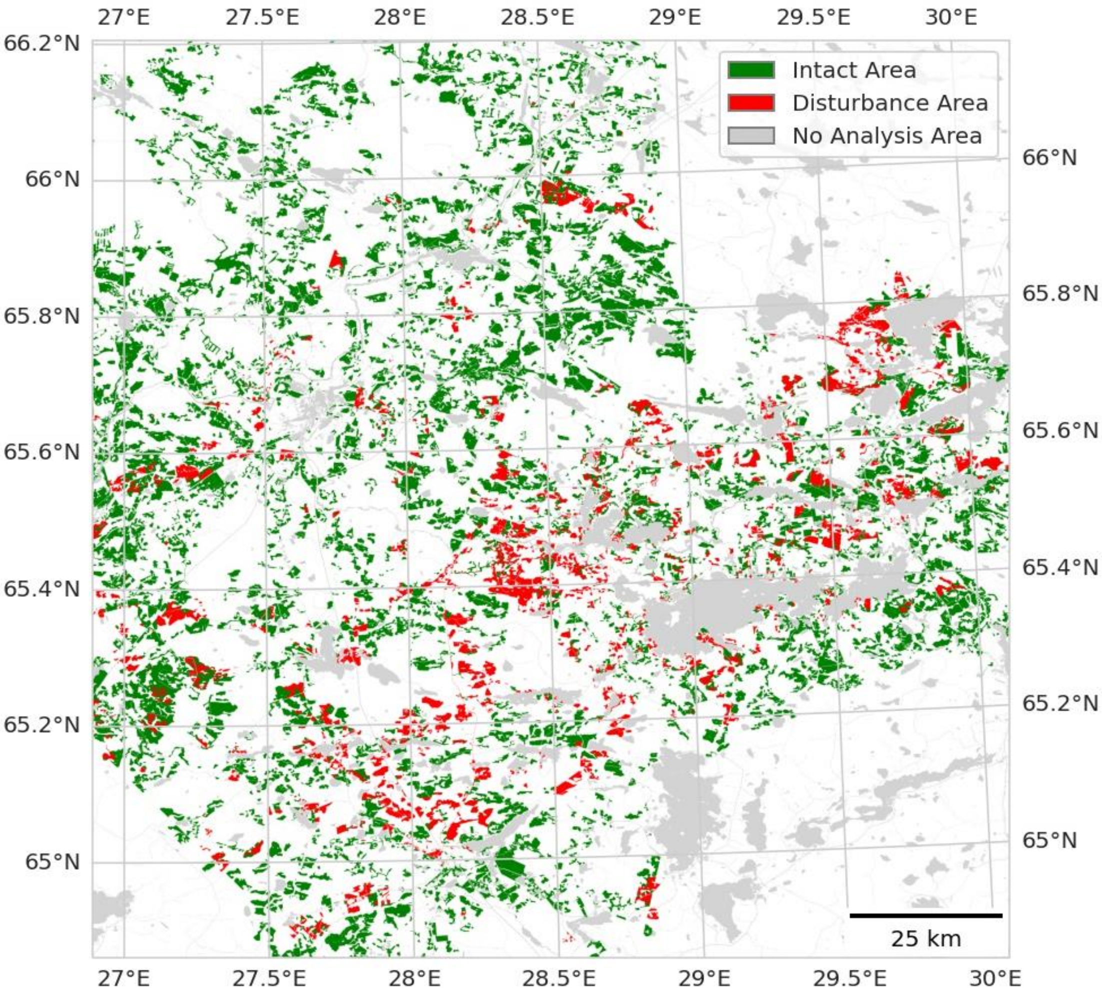
- *Data:* 2-channel Sentinel-1 imagery (VV-VH)
- *Pre-processing*: image normalization, cropping saturated values.
- *Deep feature extraction:* 3rd layer of the model with BigEarthNet for weights.
- *Post-processing:* eliminating small objects using morphological operations

Experiments: Snow Damage

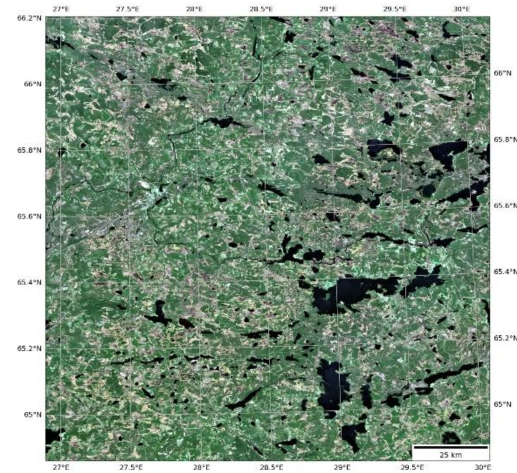


DLR

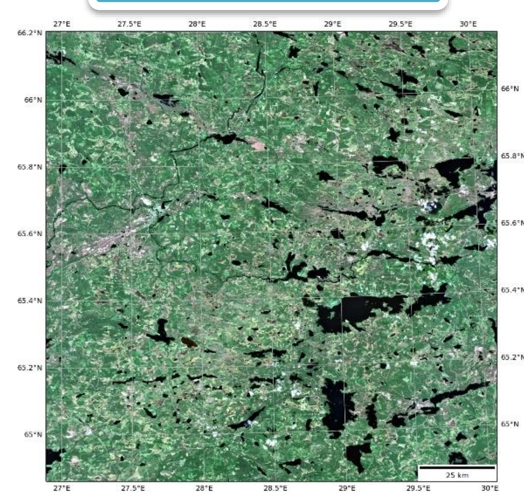
Experiments: Windstorm Damage



Ground-truth



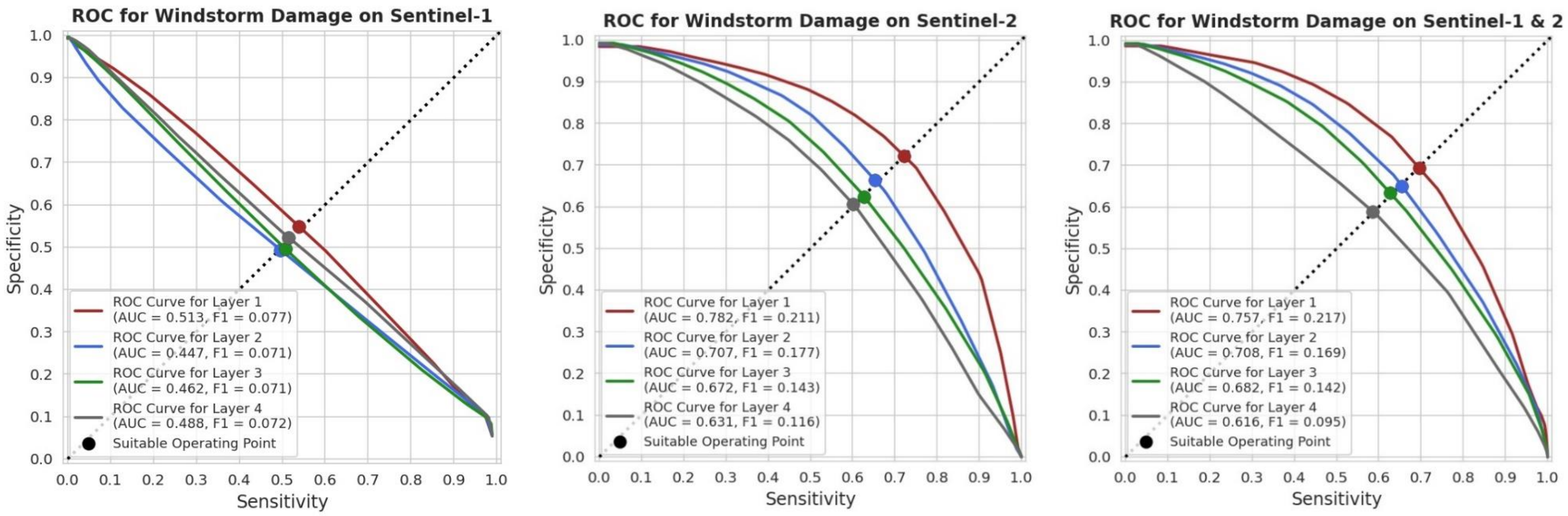
Before the event



After the event

Experiments: Windstorm Damage

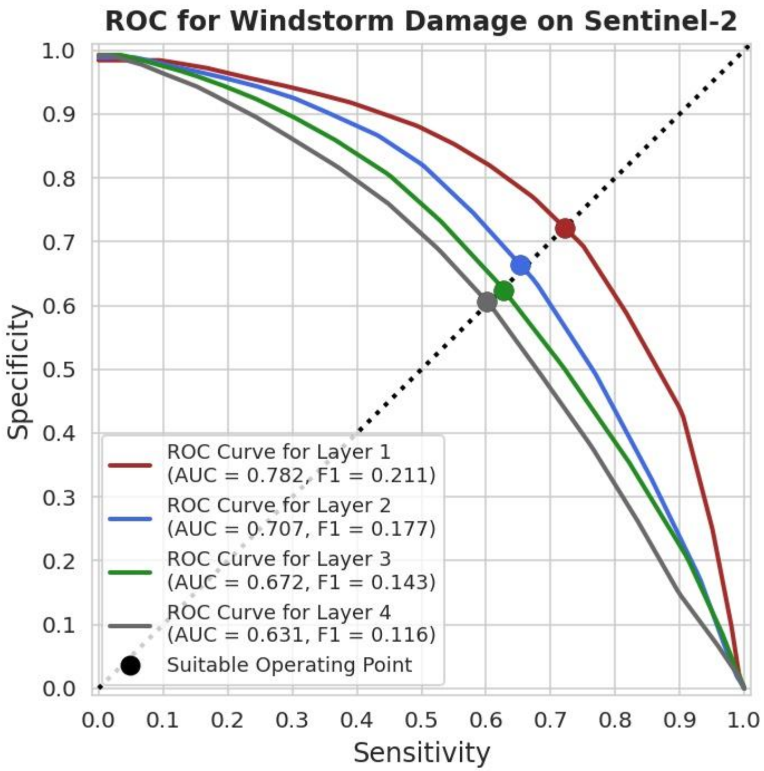
Comparisons of input types for detecting windstorm damage on BigEarthNet based model:

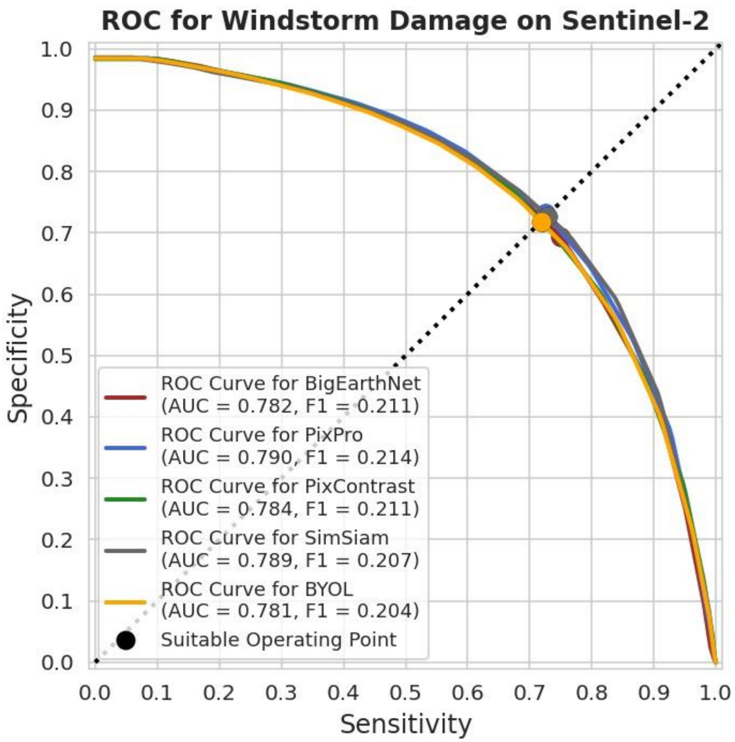


Sentinel-2 is better performing for windstorm damage assessment

Low-level layers of Resnet-18 result in better prediction

Experiments: Windstorm Damage

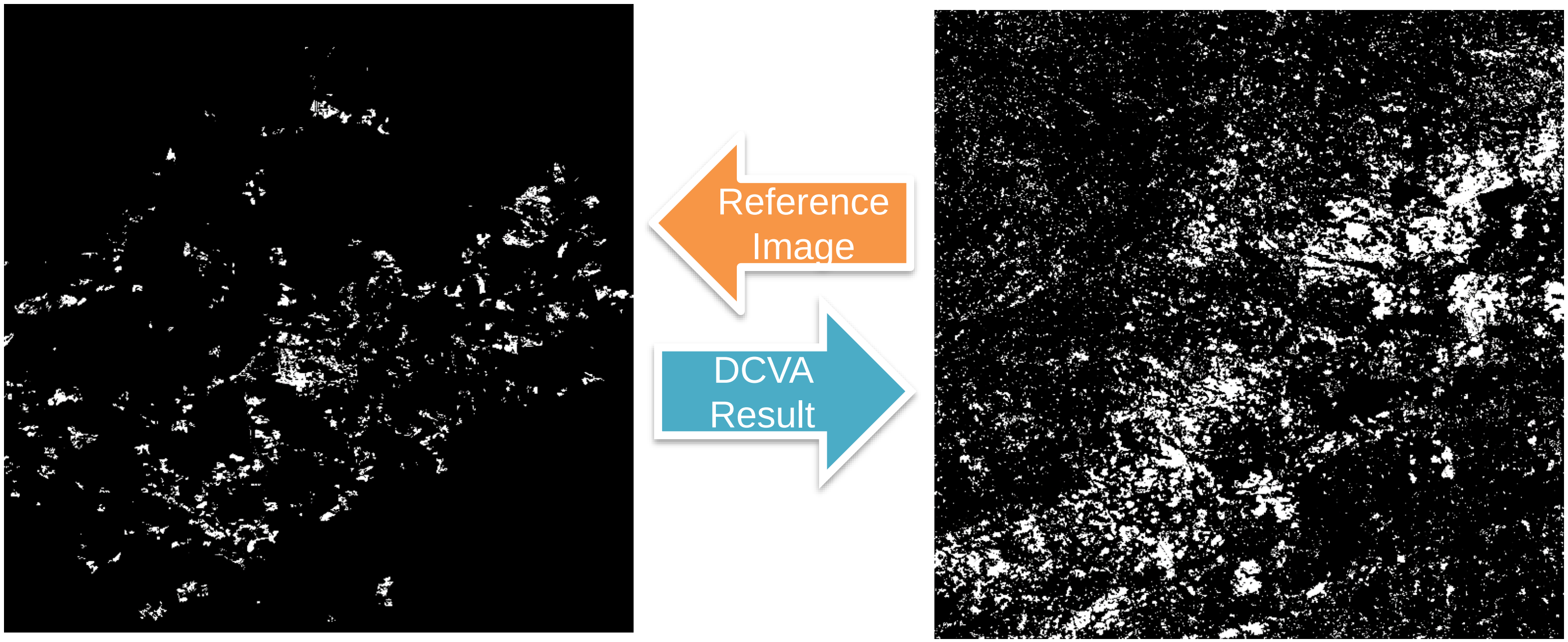




PixPro weights outperforms the baseline **BigEarthNet** weights in Windstorm damage assessment.

Experiments: Windstorm Damage

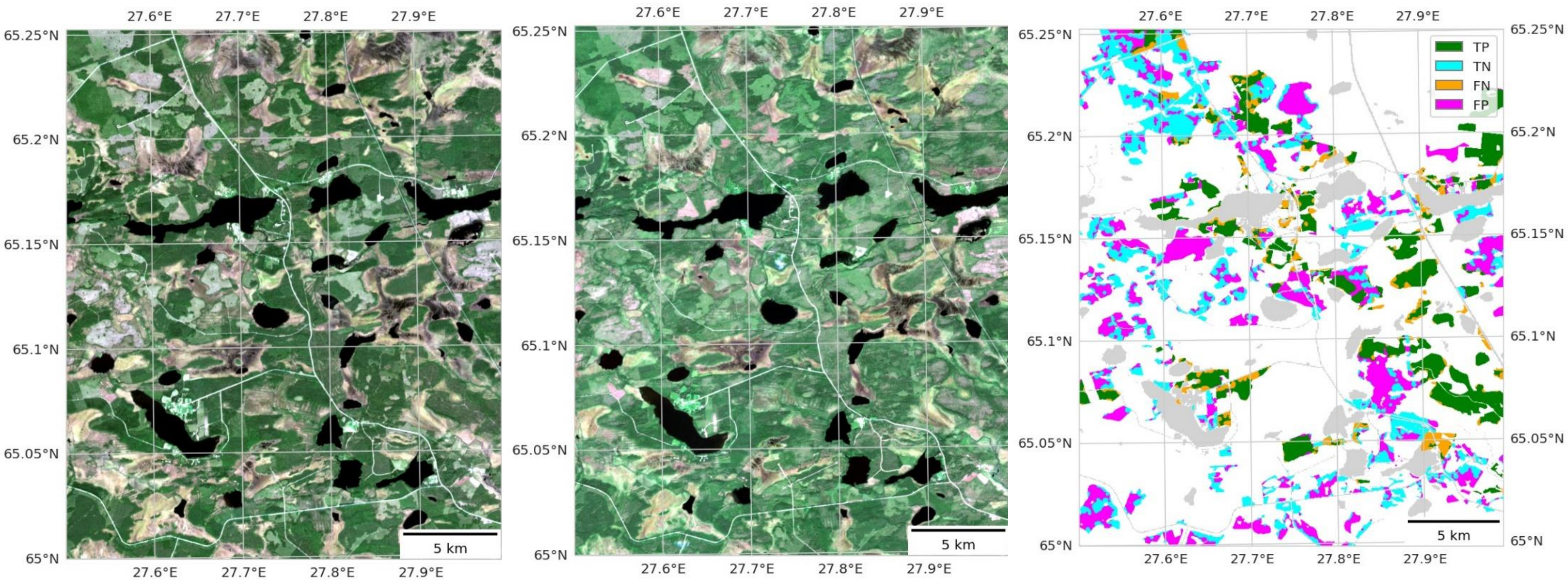
Method	Sensitivity	Specificity	Accuracy
DCVA – Sentinel 2	52.22	86.22	85.19



- *Data:* 4-channel Sentinel-2 imagery (RGB + NIR)
- *Pre-processing*: image normalization, cropping saturated values.
- *Deep feature extraction: 1st* layer of the model with BigEarthNet for weights.
- *Post-processing:* eliminating small objects using morphological operations

Experiments: Windstorm Damage

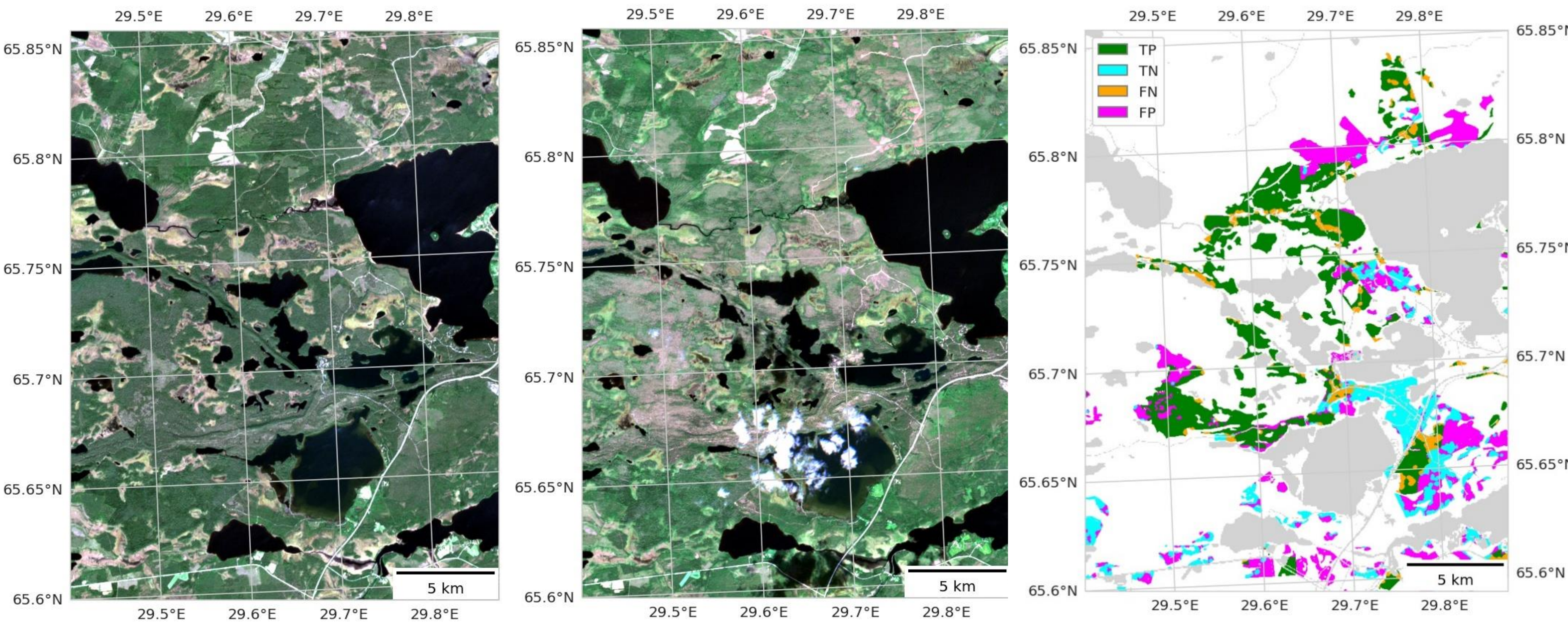
Prediction results when the pretrained Resnet-18 has been utilized as feature extractor:



FALSE POSITIVES are real changes even observed by eye in S2 images, but not labelled as Windstorm.

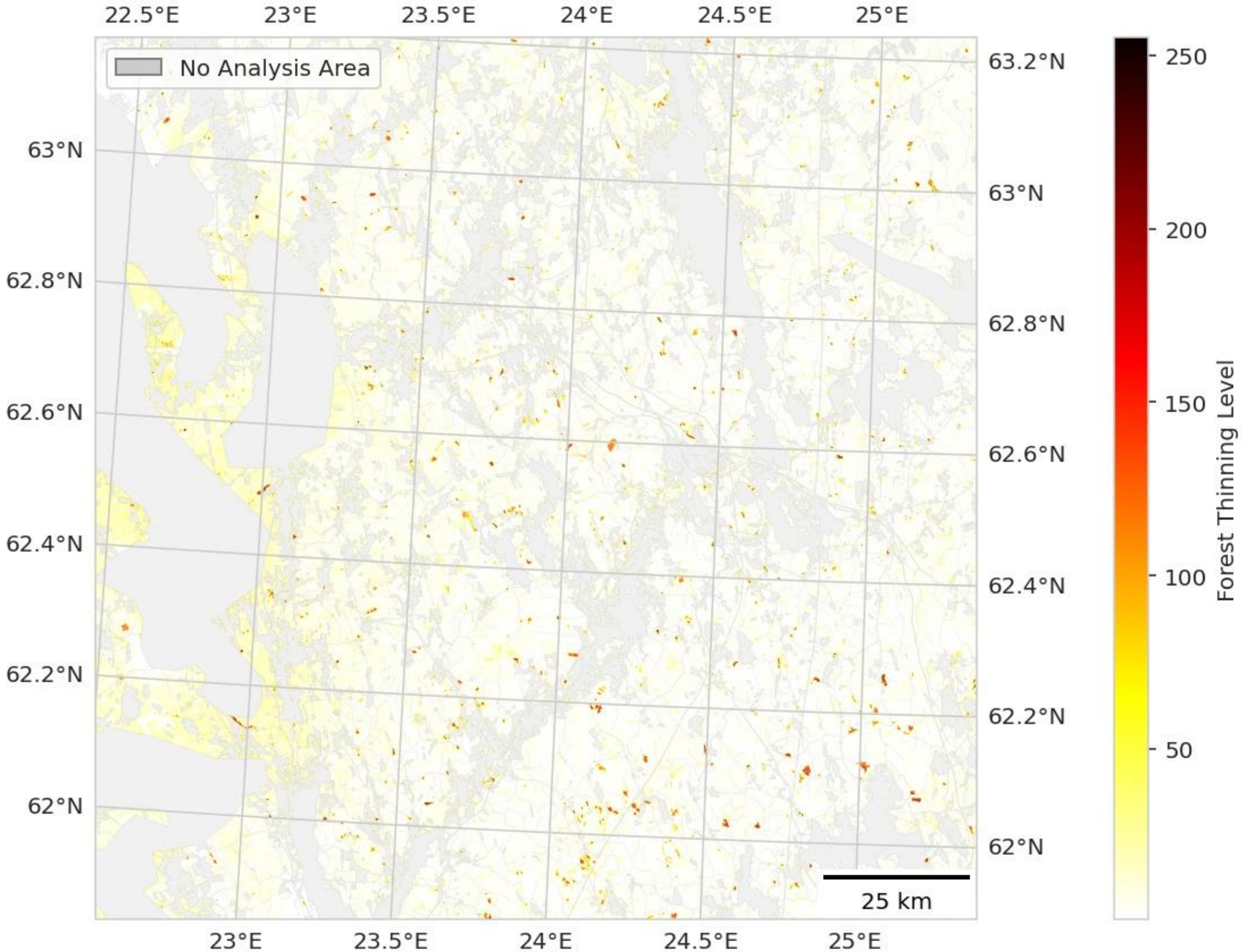
Experiments: Windstorm Damage

Prediction results when the pretrained Resnet-18 has been utilized as feature extractor:

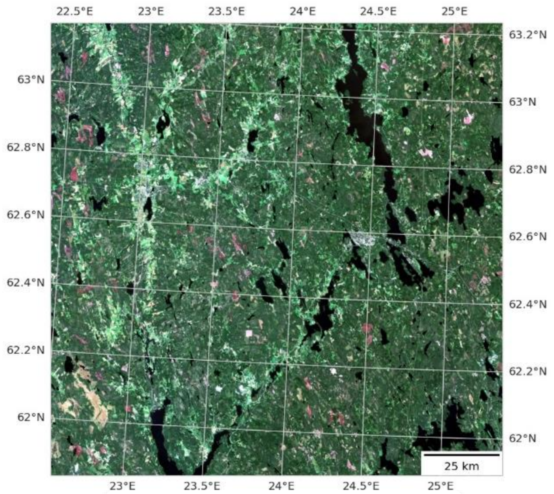


FALSE POSITIVES are real changes even observed by eye in S2 images, but not labelled as Windstorm.

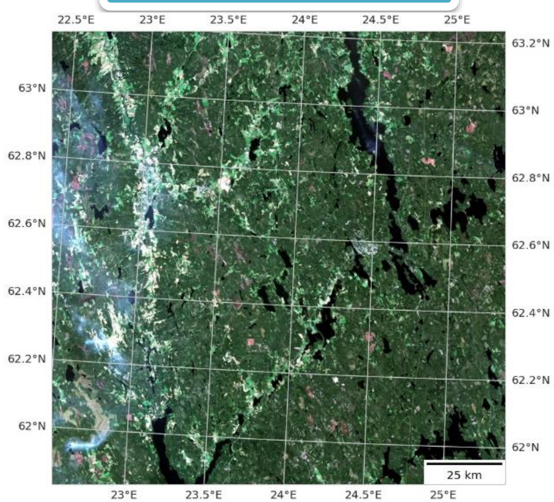
Experiments: Clearcut Damage



Ground-truth



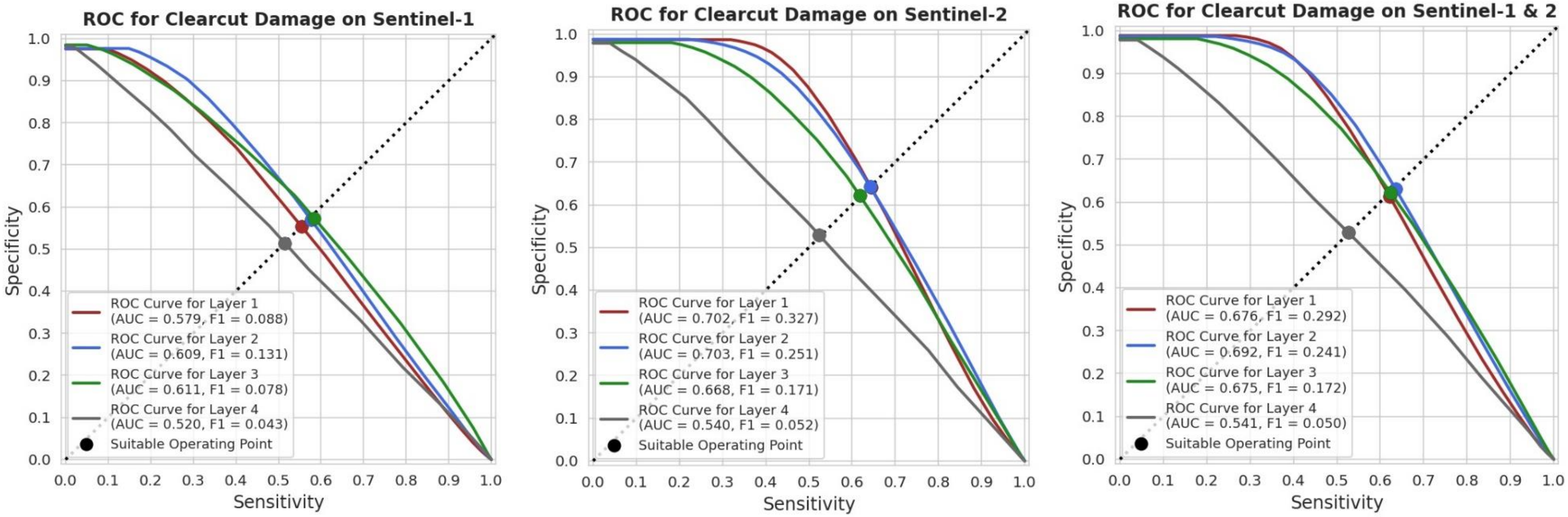
Before the event



After the event

Experiments: Clearcut Damage

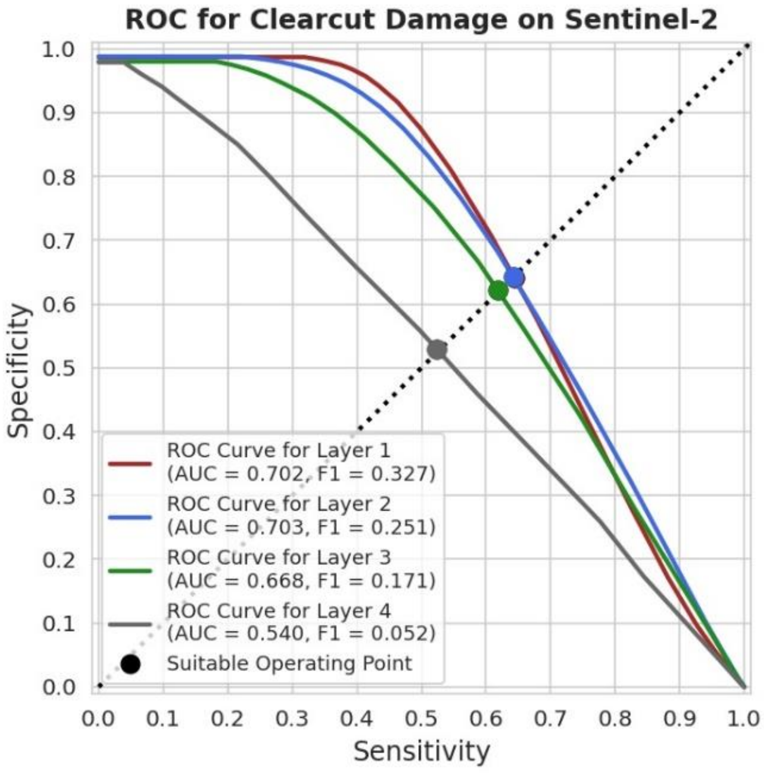
Comparisons of input types for detecting windstorm damage on BigEarthNet based model

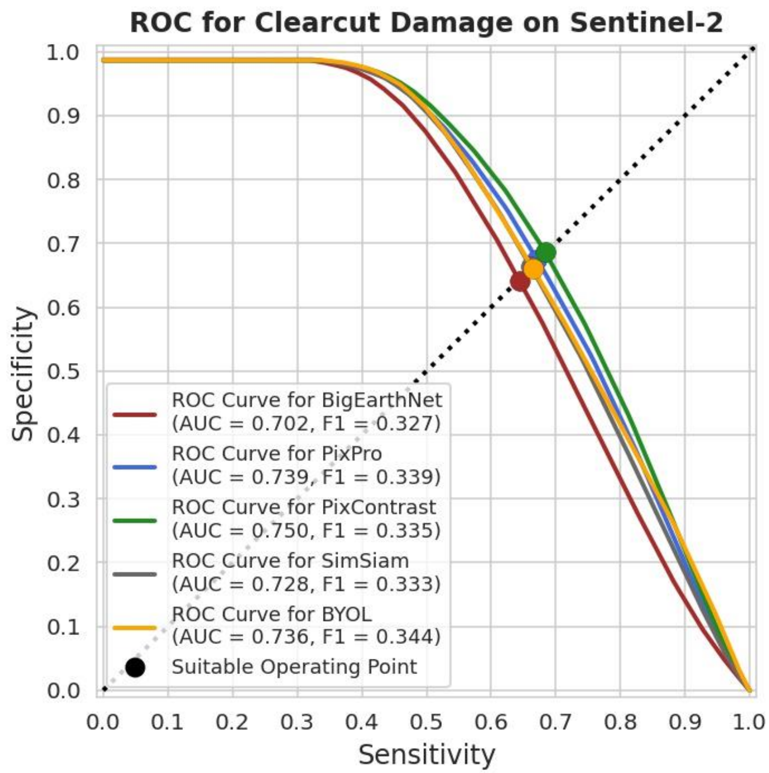


Sentinel-2 is better performing for clearcut damage assessment

Low-level layers of Resnet-18 result in better prediction

Experiments: Clearcut Damage

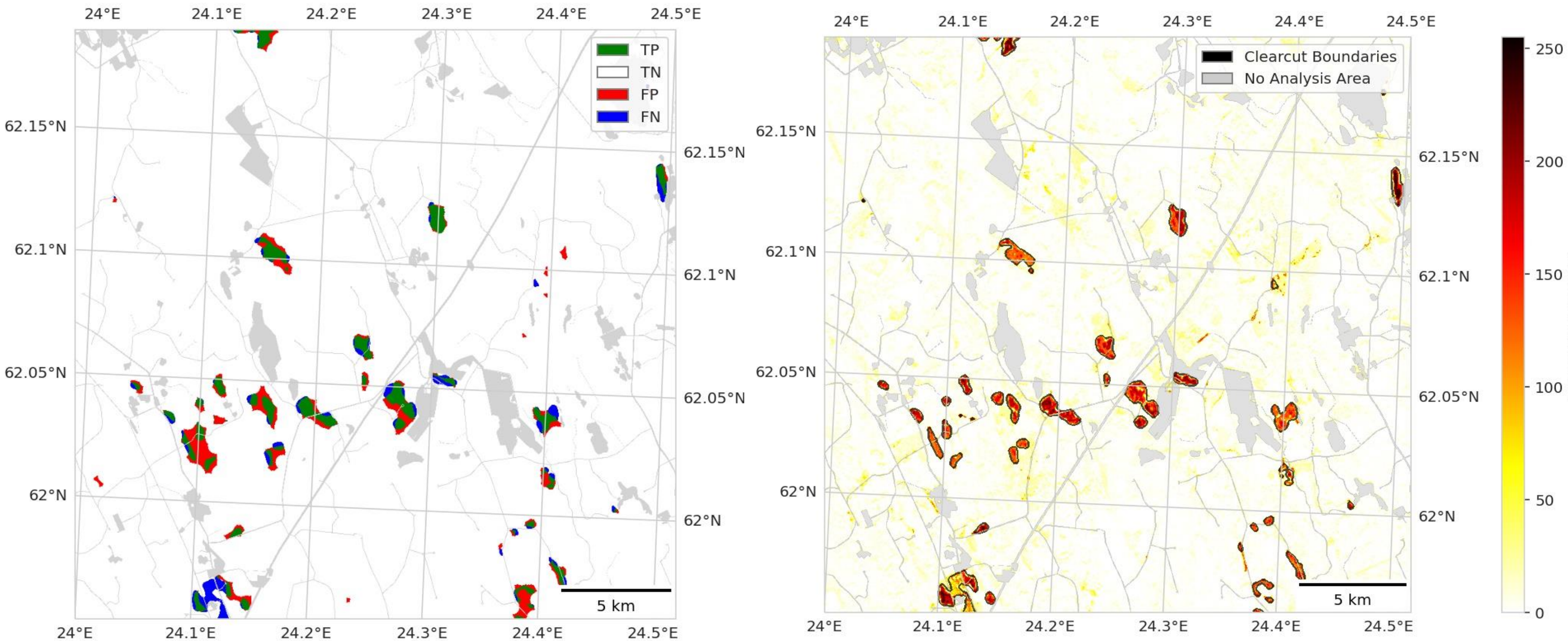




PixContrast weights outperforms the baseline **BigEarthNet** weights in Clearcut damage assessment.

Experiments: Clearcut Damage

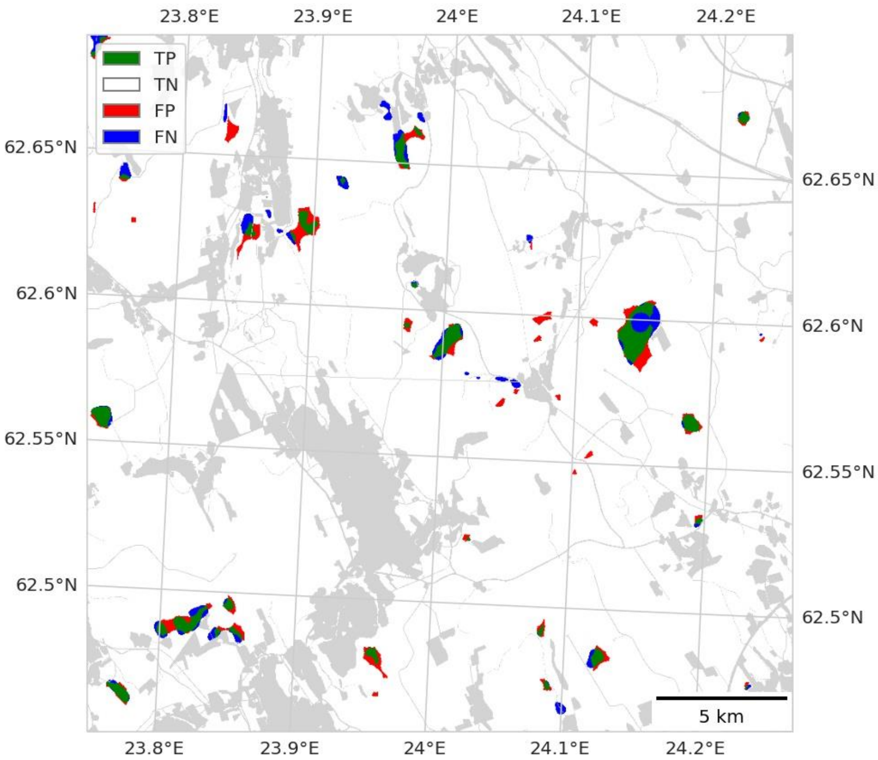
Prediction results when the pretrained Resnet-18 has been utilized as feature extractor:

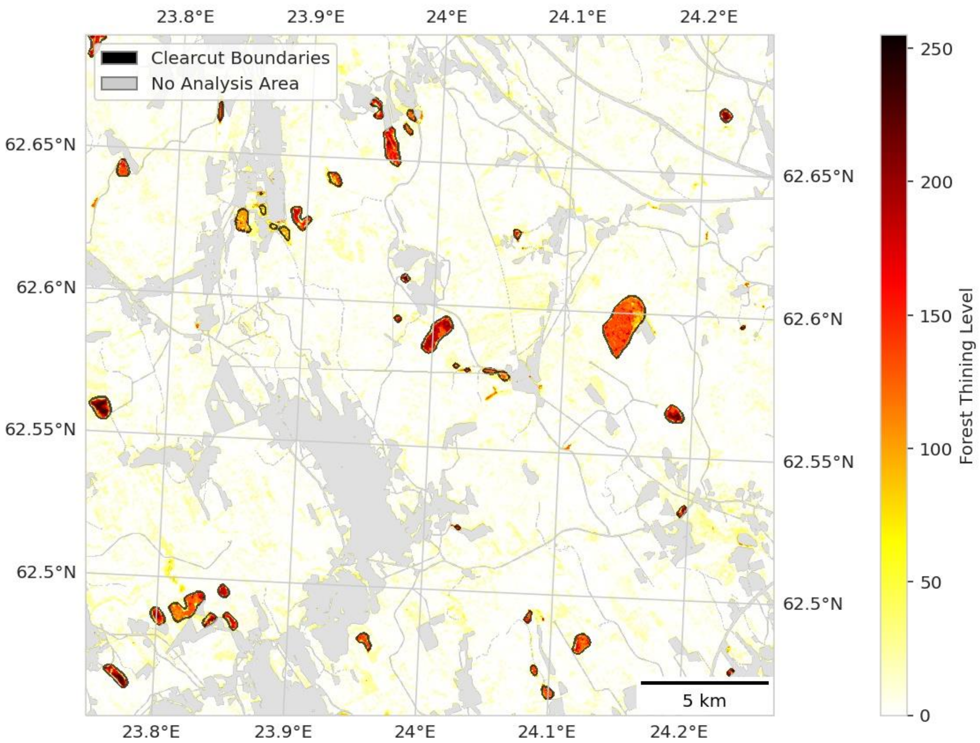


FALSE POSITIVES are real changes even observed by eye in S2 images, but not labelled as Windstorm.

Experiments: Clearcut Damage

Prediction results when the pretrained Resnet-18 has been utilized as feature extractor:





FALSE POSITIVES are real changes even observed by eye in S2 images, but not labelled as Windstorm.

Conclusions

Training SSL only for 10-15 epochs on Forestry dataset is outperforming in most of the cases, even with ~10 times smaller data

	Snow Damage		Windstorm Damage		Clearcut Damage	
	F1	AUC	F1	AUC	F1	AUC
PixContrast	**0.588**	**0.870**	0.211	0.784	0.335	**0.750**
PixPro	0.547	0.844	**0.214**	**0.790**	0.339	0.739
SimSiam	0.529	0.833	0.207	0.789	0.333	0.728
BYOL	0.562	0.839	0.204	0.781	**0.344**	0.736
BigEarthNet	0.584	0.844	0.211	0.782	0.327	0.702

Conclusions

Snow damage

- Competitive results for unsupervised bitemporal change detection, comparable to supervised methodologies utilizing Sentinel-1 time series data and “hand-engineered” features
- Ways to adapt the methodology to image time series need further exploration.

Windstorm damage

- DCVA was successful in detecting change areas, though specificity can be improved. Performance of DCVA is **dependent on the capabilities of feature extractor**.
- Further, potential of **forest-specific feature extractor can be examined to improve** the retrieved features and thus prediction performance.

Thank you!