Semi-supervised learning for spatiotemporal super-resolution Land Cover segmentation

Boris Flach[a], **Lukáš Brodský**[b], Antonín Hruška[a], Tomáš Dlask[a], Ondřej Novák[b]
[a]Czech Technical University in Prague, [b]Mapradix

ESA Project: **Spatiotemporal Land Change Monitoring Based on Sentinel-2 Time Series and VHR Images**

lukas.brodsky@mapradix.cz

30 May | ESRIN | Frascati, Italy | https://suredos24.esa.int

Spatio-temporal SEN2VHR

Motivation:

- Sentinel-2 has many advantages for national and regional authorities (e.g. National Parks) to monitor changes;
- However, VHR resolution products are required;

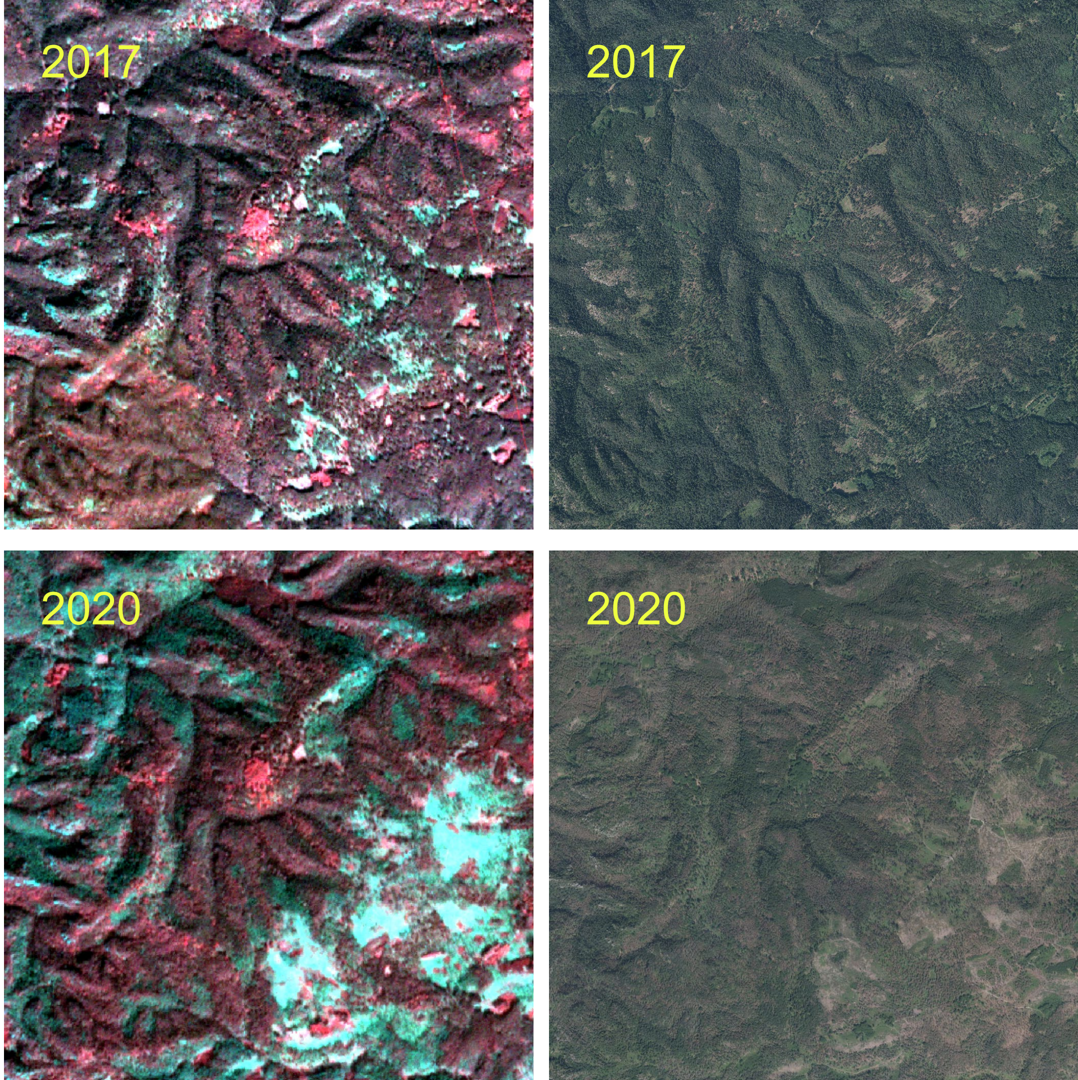
Applications:

- Continuous VHR forest change monitoring
- Continuous VHR urban change monitoring

Spatio-temporal SEN2VHR

Bohemian Switzerland National Park

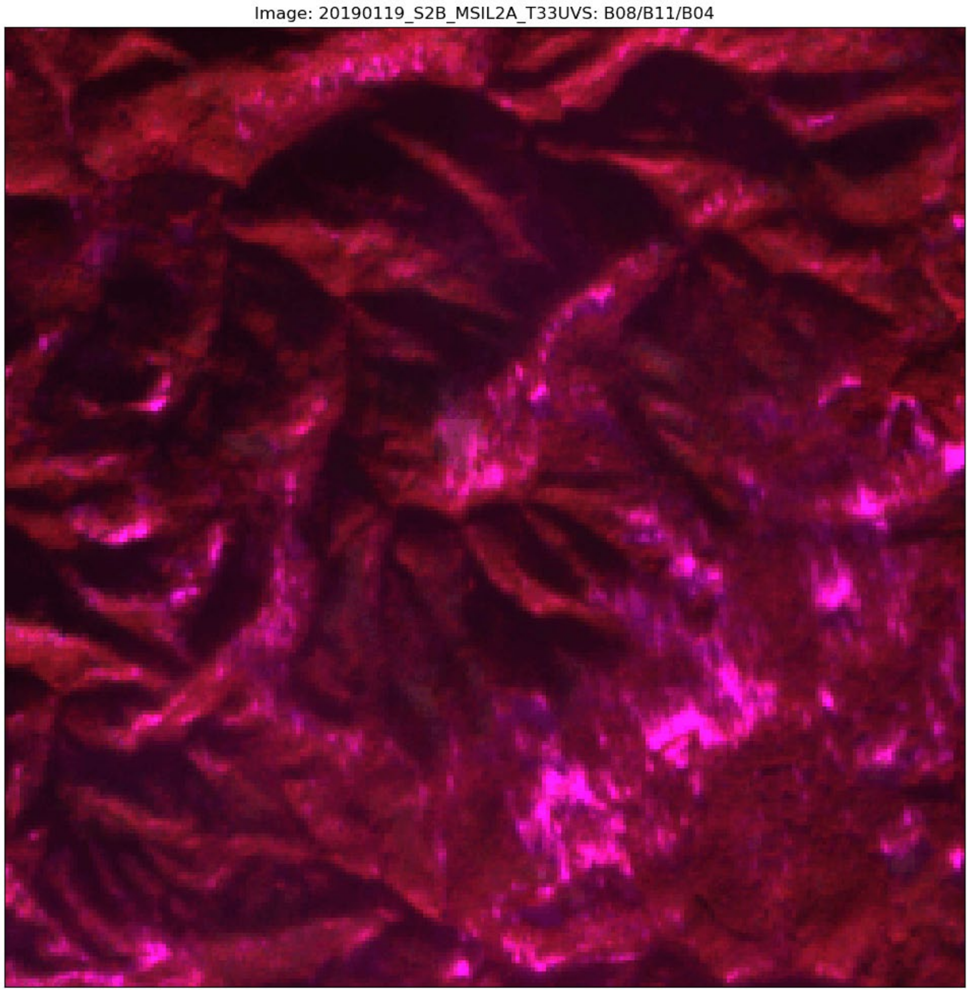
- Sparse VHR images



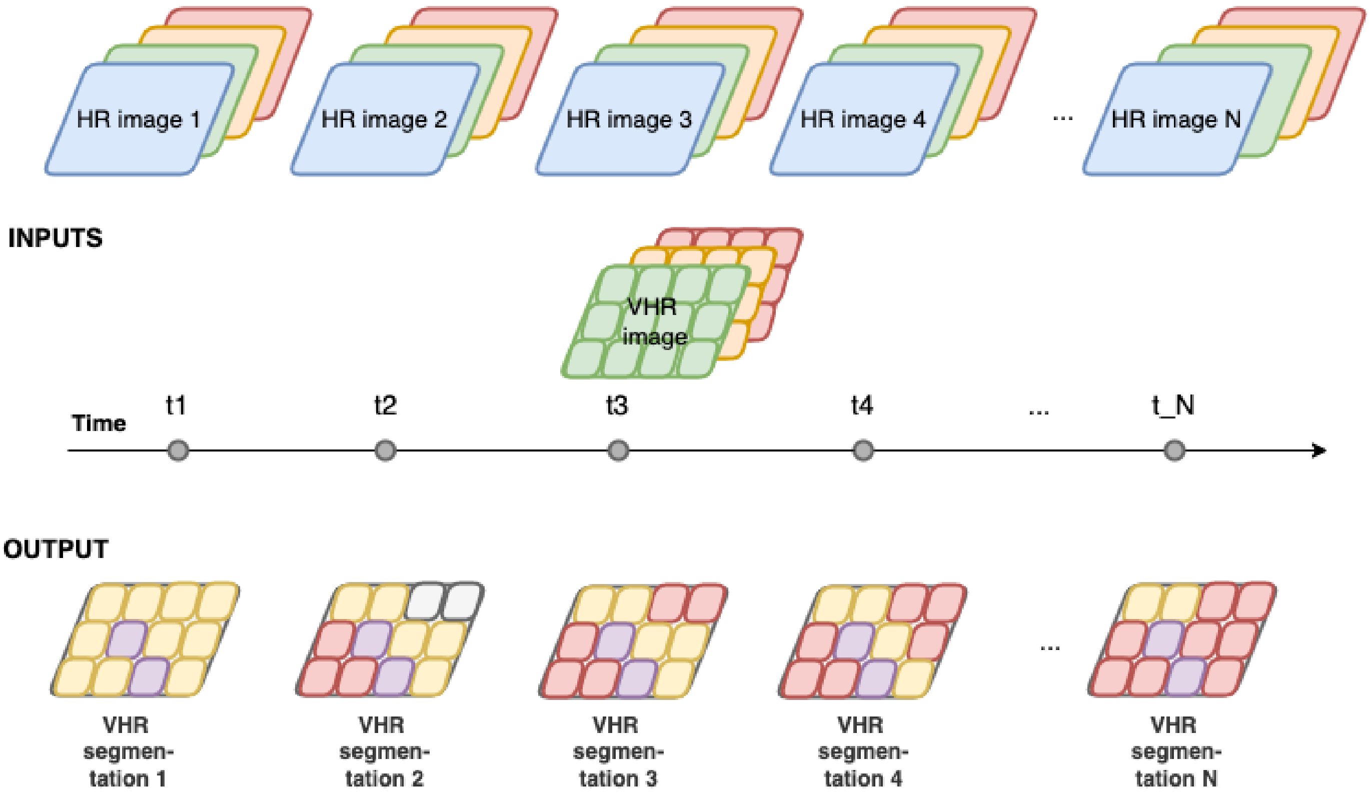
Spatio-temporal SEN2VHR

Bohemian Switzerland National Park

- Sentinel-2 time series



Spatio-temporal SEN2VHR



Spatio-temporal SEN2VHR

Goal:

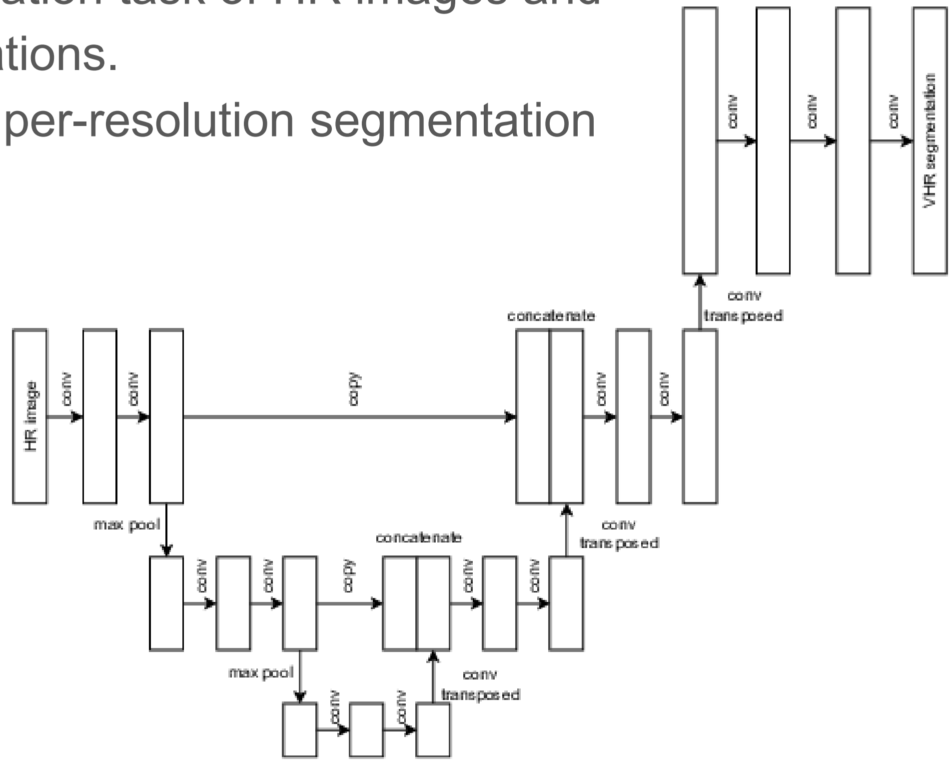
- goal is to design and train a model that takes as input a VHR Land Cover segmentation and an ongoing sequence of Sentinel-2 images, and predict a sequence of VHR Land Cover segmentations.

ST_SEN2VHR Solution

1. **Single-image super-resolution segmentation**
2. **From sparse to dense annotations: active learning**
3. **Spatiotemporal Modelling and Super-resolution Segmentation**
4. **Semi-supervised learning**

1. Single-image super-resolution segmentation

- Jointly solves the semantic segmentation task of HR images and the required lifting to VHR segmentations.
- Asymmetric UNet architecture for super-resolution segmentation



1. Single-image super-resolution segmentation

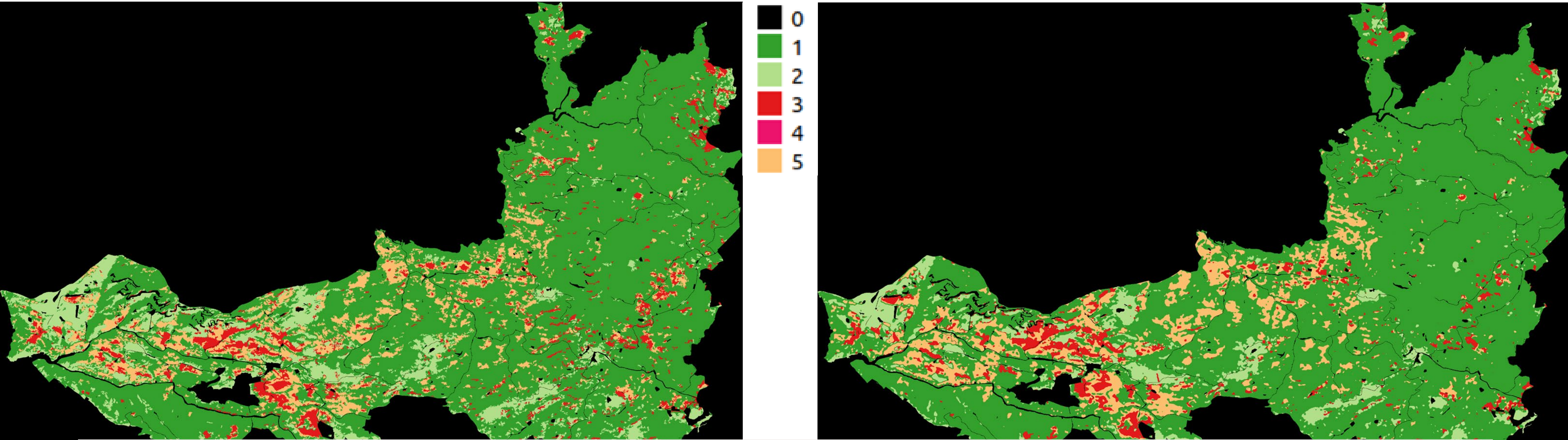
The model is applied and trained in the following way:

- the model segments single HR images without temporal context
- the model is fully convolutional and can be trains in presence of gaps
- training data are the enriched annotations which we assume to be valid for HR images with acquisition dates at most 5 days away from the acquisition date of the VHR data.

Forest monitoring nomenclature:

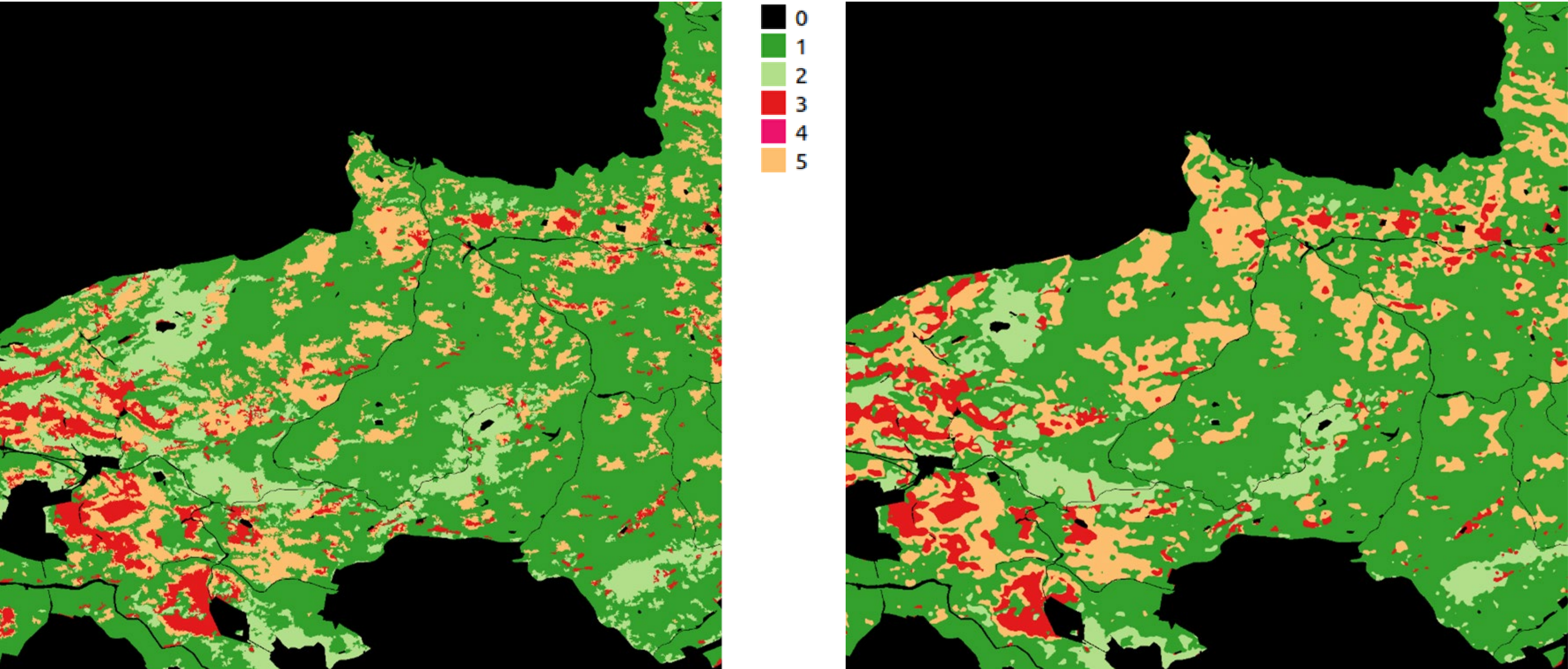
1. coniferous forest
2. broadleaf forest
3. clear-cut
4. grass and herbs in clear-cutted areas
5. dry trees

Left: “ground truth” VHR segmentation, right: VHR segmentation obtained by the model from the closest date HR acquisition (segmentation accuracy of 87%)



Experiments show that the project objectives are feasible, conditioned on an improved spatio-temporal modelling of the HR acquisitions and an improved treatment of missing data (clouds, cloud shadows, etc.).

Left: “ground truth” VHR segmentation, right: VHR segmentation



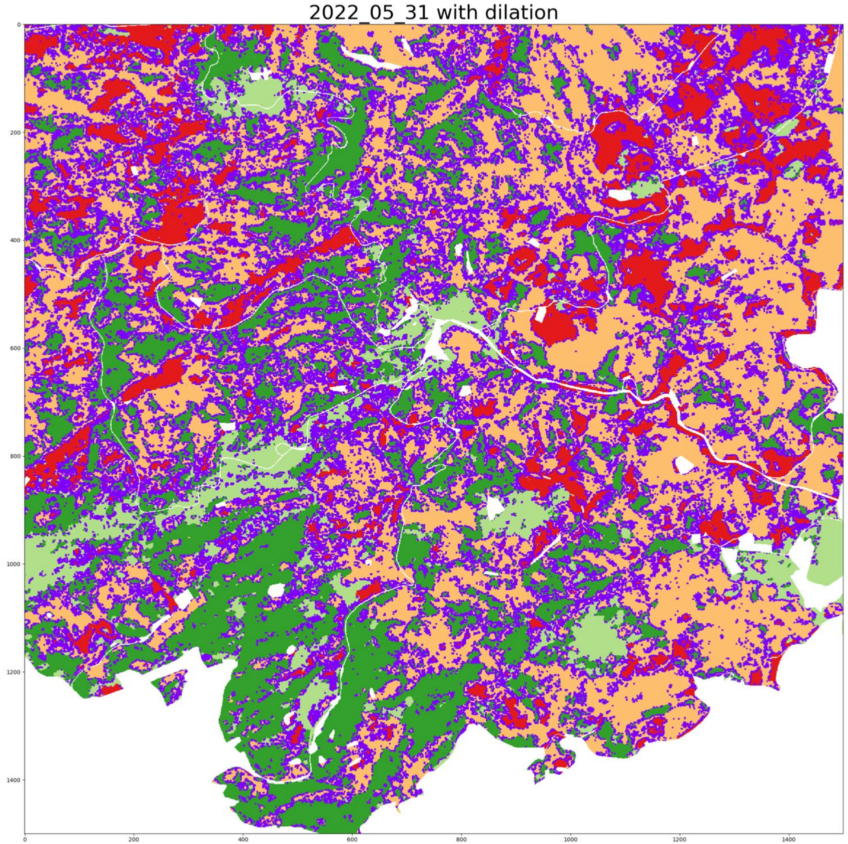
2. From sparse to dense annotations: active learning

Main idea: iterative active learning (Ren et al., 2021).

- 1. Use a RF model and active learning for iteratively enlarging an initially very sparse annotated data set;
- 2. The segmentations obtained from the iteratively retrained random forest will then serve as denser annotations for training a deep neural network with U-Net architecture.

From sparse to dense annotations: active learning

- Transition annotations



3. Spatiotemporal Modelling and Super-resolution Segmentation

Architecture:

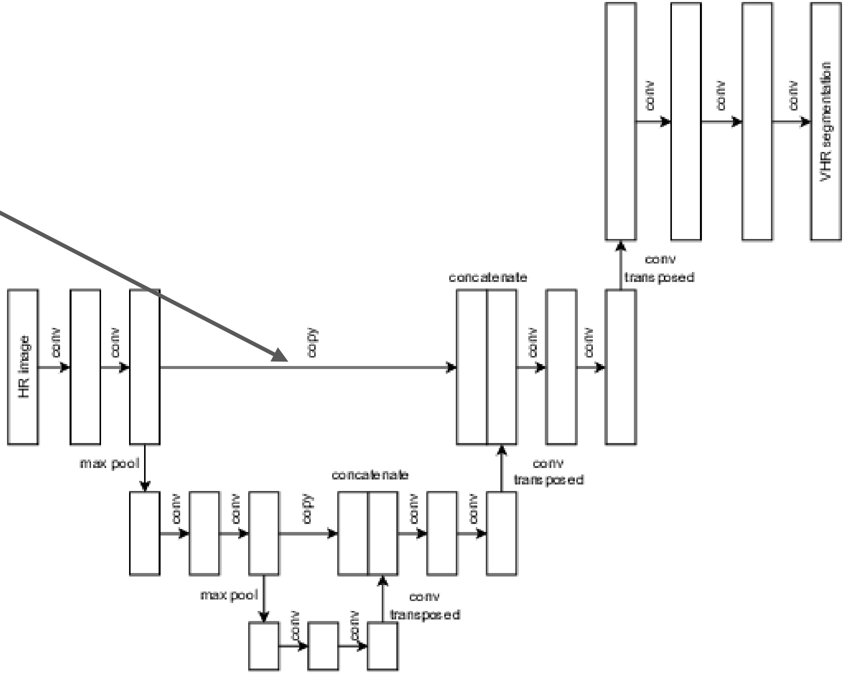
- **Spatiotemporal CNN with asymmetric U-Net**
- **Conditional Markov model** that accounts for the long-term temporal context.

! The core challenge of this module is scarcity of available training data.
! Yet another challenge for this task arises from incomplete data in the HR cube caused by clouds and snow cover.

ASUnet/ME

- ASUnet extension for seasonal variations

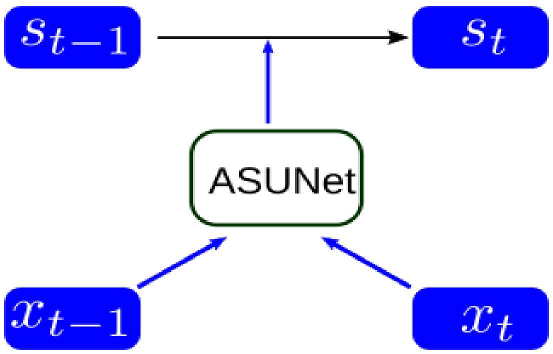
The skip connections are scaled and shifted by outputs of MLP with month embedding inputs.



3. Spatiotemporal Modelling and Super-resolution Segmentation

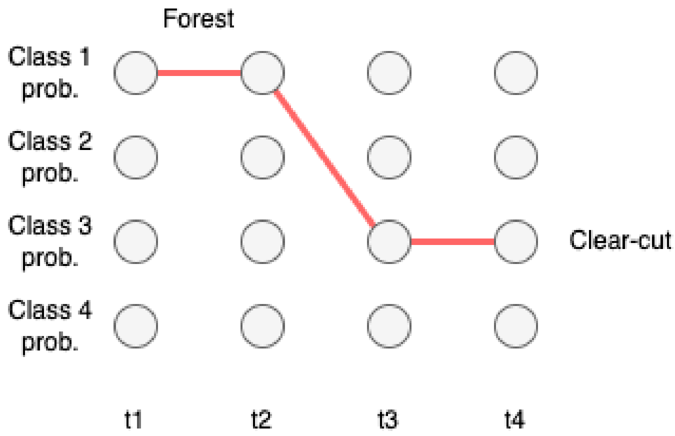
Modelule A: ASUnet/ME

Input: HR sliding window to account for a local spatio-temporal context.
Output: LC transition probabilities .



Modelule B: Markov chain model

Tracks the most probable consistent sequence of LC states (labels), based on transition probabilities.



4. Semi-supervised learning

- ASUnet/ME + Markov model
- Annotations: fully spatial coverage for e.g. one image in 1 a year (transition label sets)

- **Training objective:** added component for temporal smoothness of the predicted transition probabilities

  *L = cross-entropy + lambda * temporal smoothness*

- Training loss: log probabilities of set of transition labels + lambda of temporal smoothness
- Temporal smoothness: KL-divergence between consecutive transition probabilities

Results

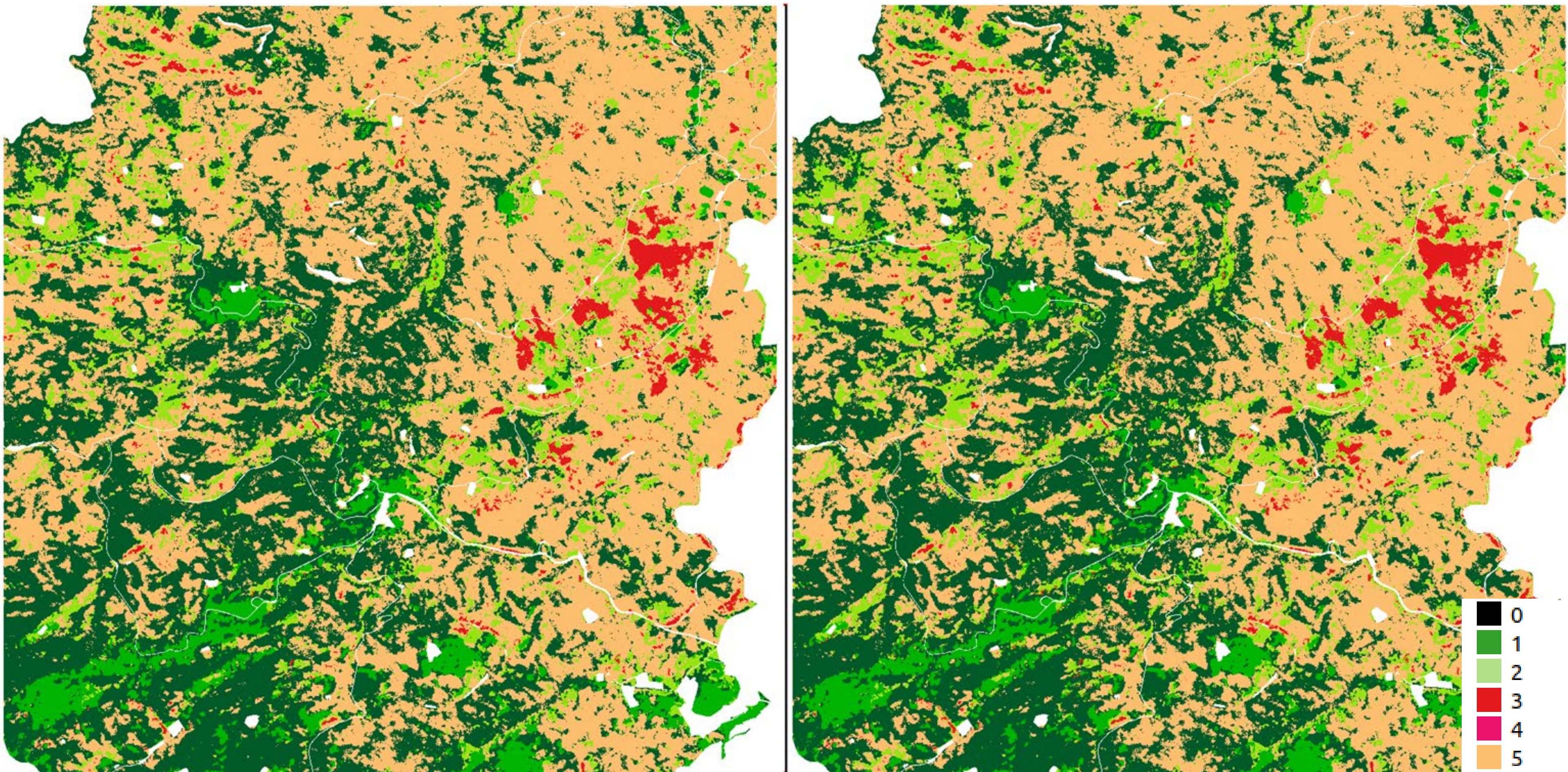
Overall segmentation accuracies

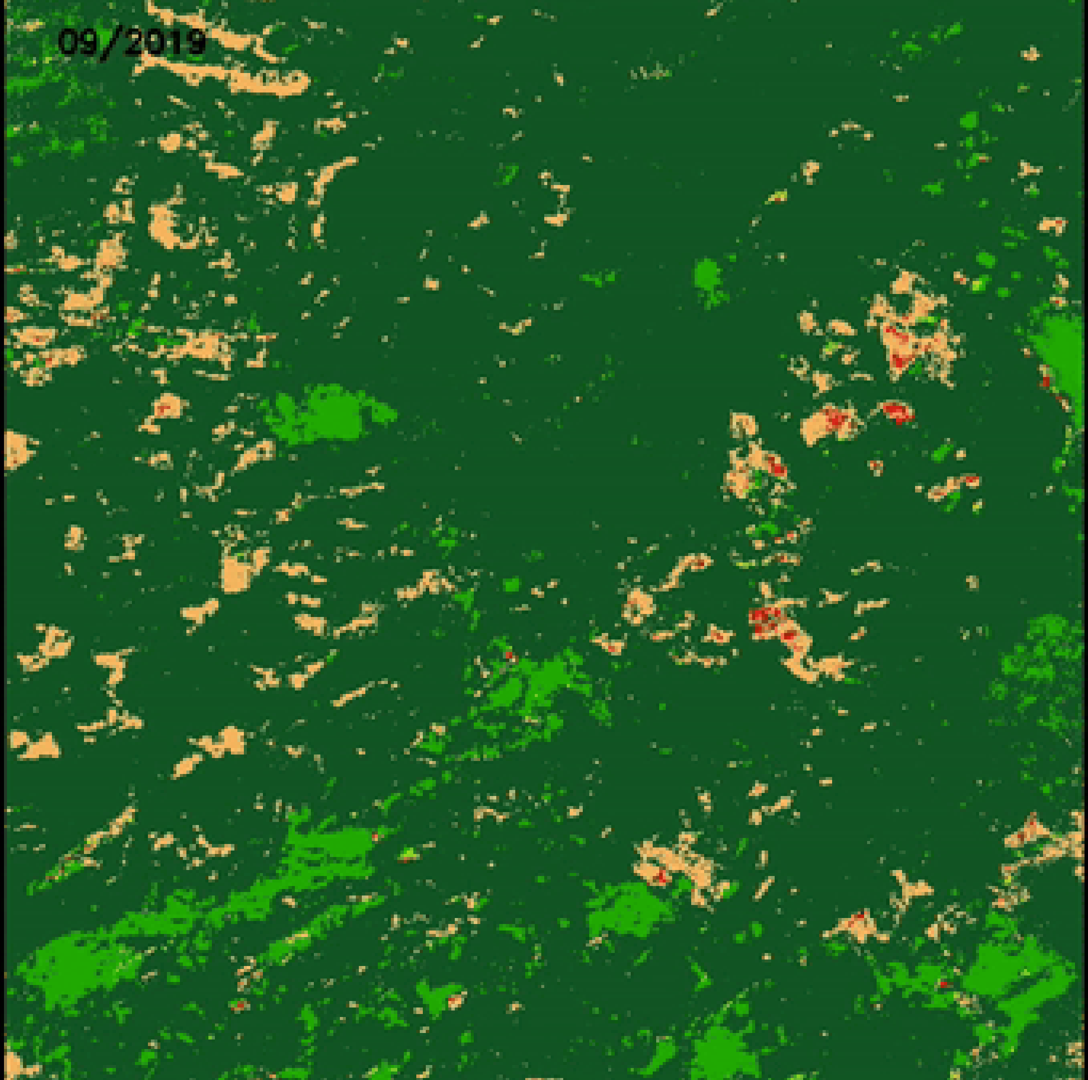
HR date	VHR date	accuracy
2018-08-07	2018-08-07	0.920
2019-09-03	2019-09-04	0.910
2020-09-10	2020-09-09	0.828
2021-09-05	2021-09-06	0.874
2022-07-19	2022-07-20	0.896

- ASUNets with 3-4 resolution levels and ~5.0e6 parameters can achieve an accuracy close to 90%

Results

Left: VHR segmentation obtained by the model, right: VHR annotation (2022-07-20)




09/2019

Conclusion

- Proposed ASUnet/ME proves suitable;
- The network with medium size (1-5 MIO) is trainable on sparse annotations;
- Curren cloud masks (Fmask) leads to possible mistakes both in training and validation;
- Validation accuracy drops when using the time-wise training-validation split, especially when we leave out acquisitions for years with strong Land Cover changes;
- Future:
  - generative model for dealing with the gaps;
  - Ablation study (gemoetric and radiometric normalization);
  - Spatiotemporal validation schemes;

Thank you for your attention!