Land cover mapping from medium-resolution SAR and multi-spectral remote sensing images

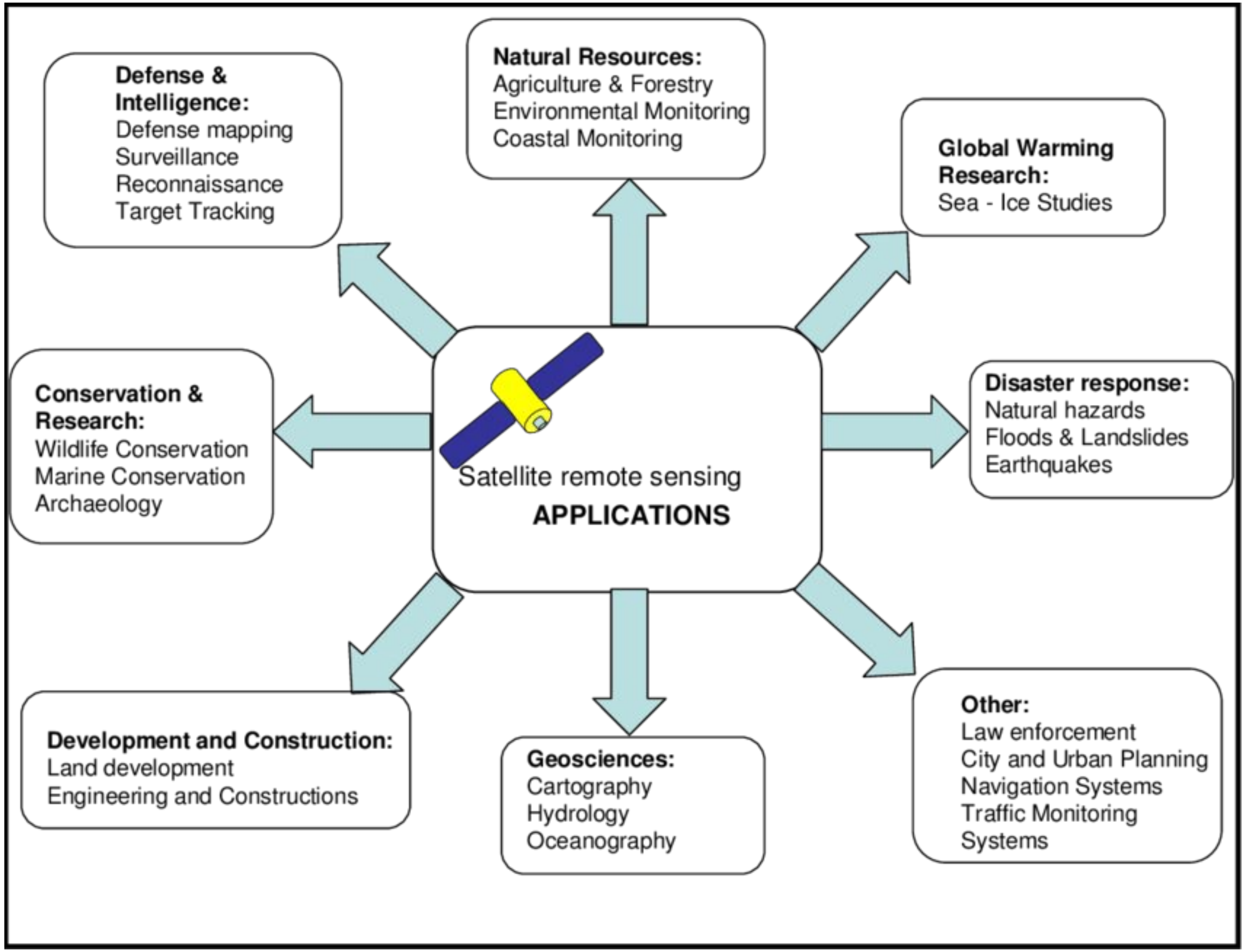
Ivan Dubrovin, Shakir Sofi, Veronika Shirokova, Evgeny Avdotin, Kundyz Onlabek, Arina Ivanova, Aleksandr Gamayunov, Ilya Barskiy

Skoltech

Outline

Skoltech

Introduction

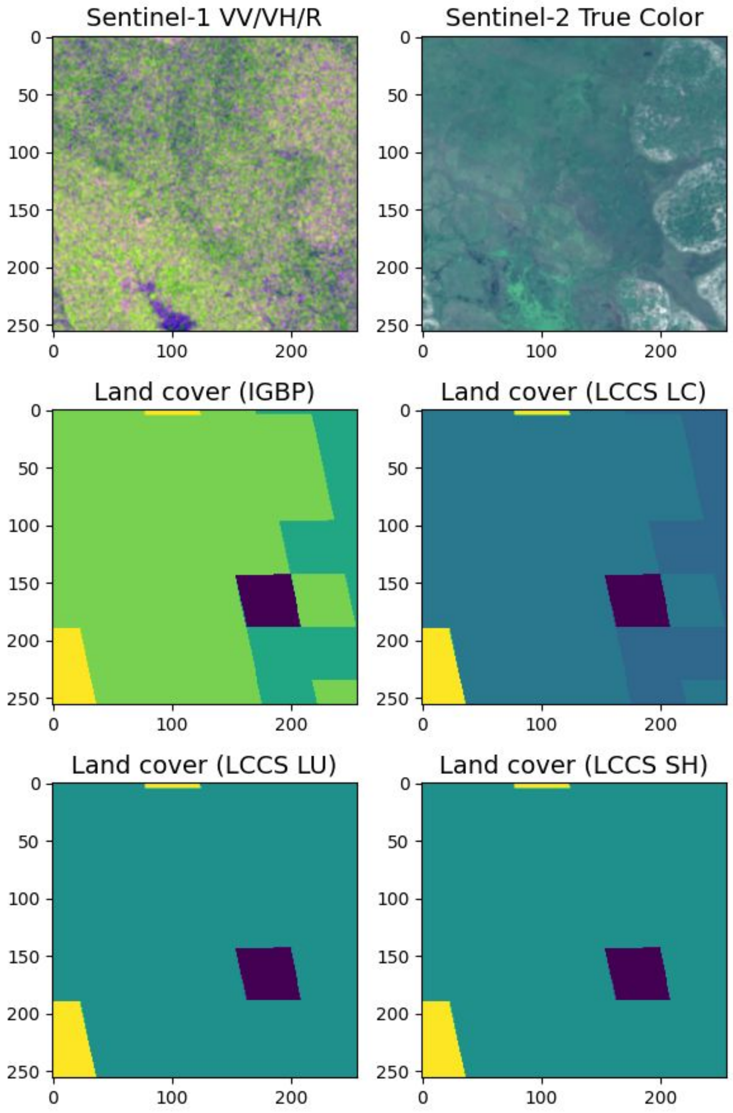


Skoltech

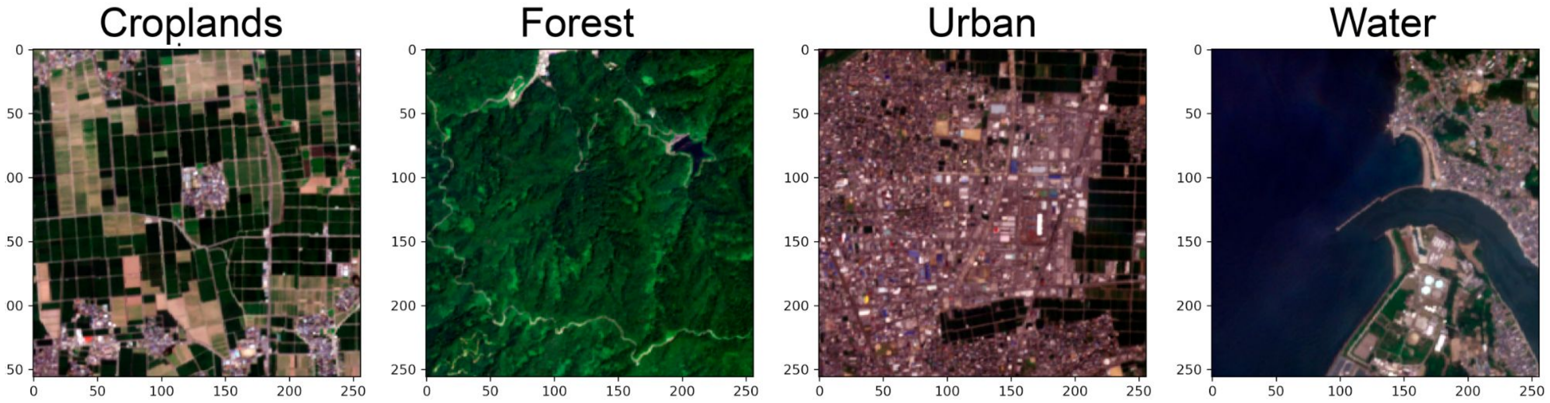
Image segmentation in remote sensing

- Image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze.

Dataset

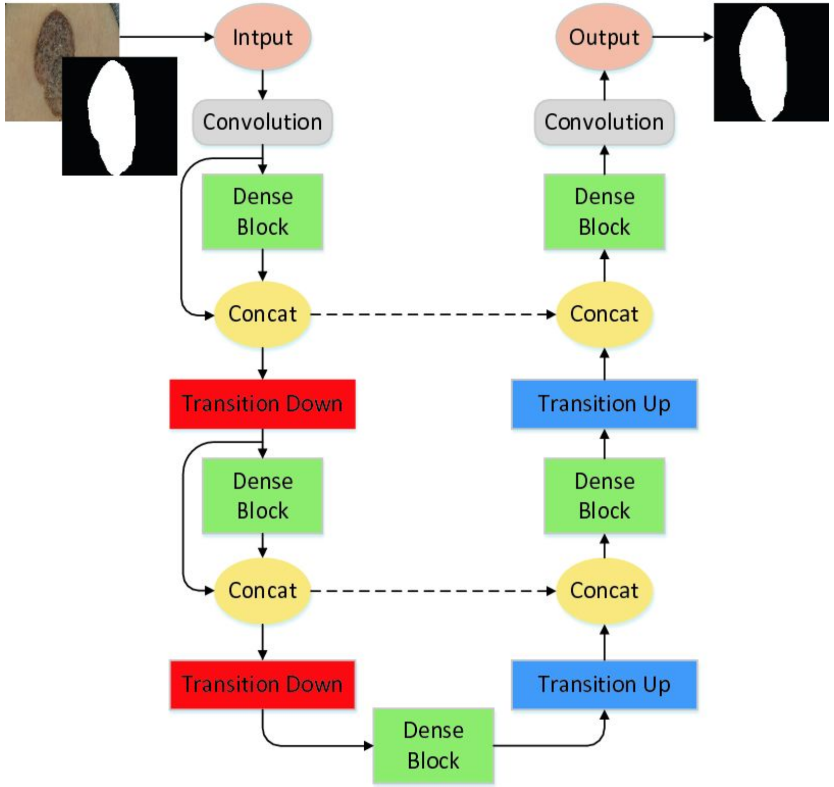


The SEN12MS dataset contains 180662 patch triplets (Sentinel-1 dual-pol SAR, Sentinel-2 multi-spectral, MODIS land cover) in the form of multi-channel Geo-Tiff images with 17 classes initially. We reduced the number of classes to the main 10 classes.



Baseline Models

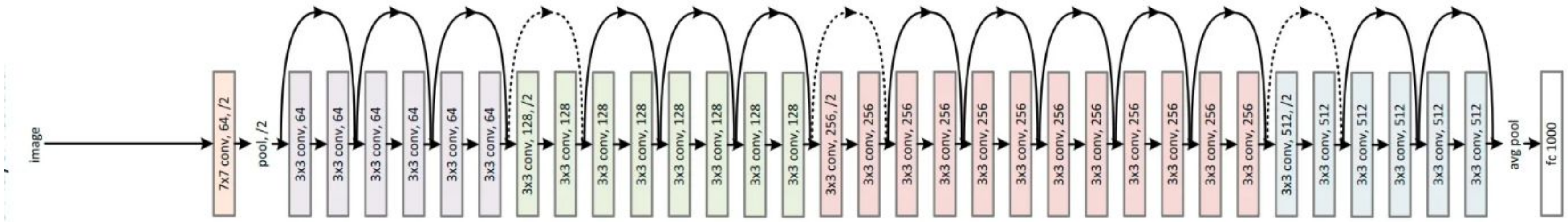
DenseNet



These models were proposed in the article by the authors of the dataset, which we used in this work

-ResNet was too slow and obsolete for our task

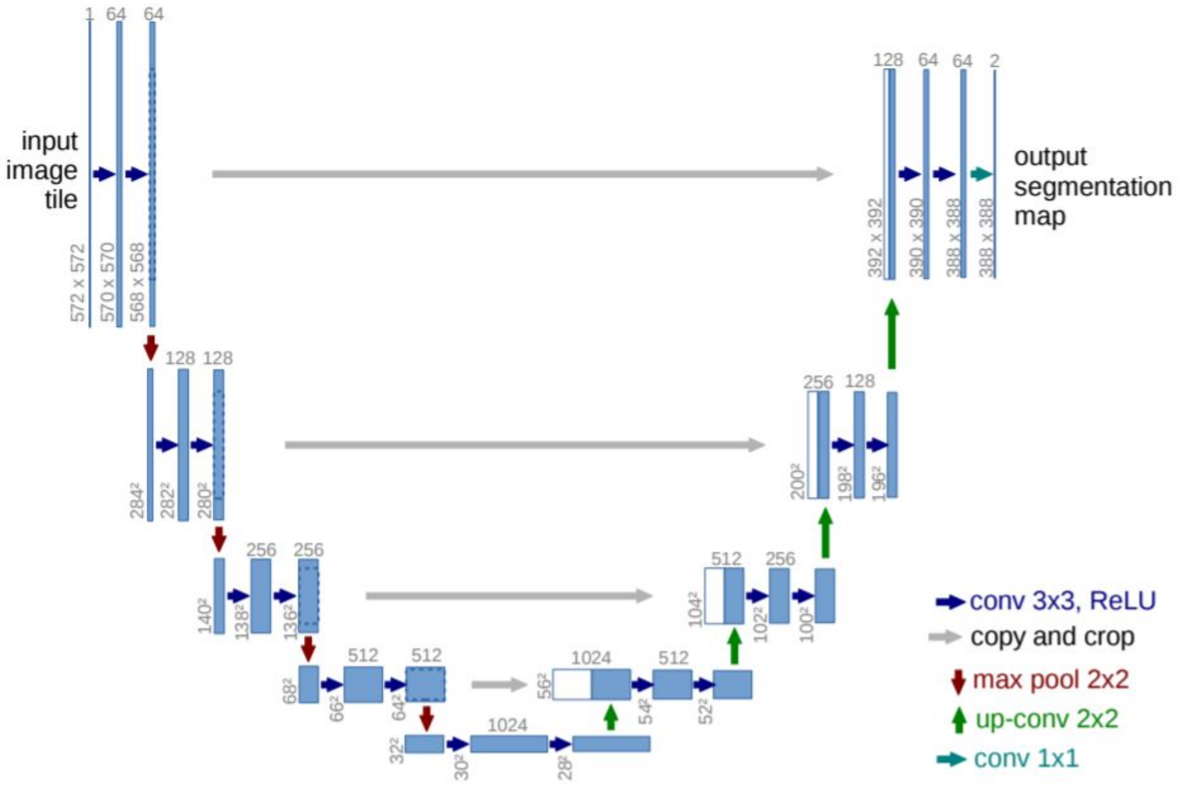
ResNet



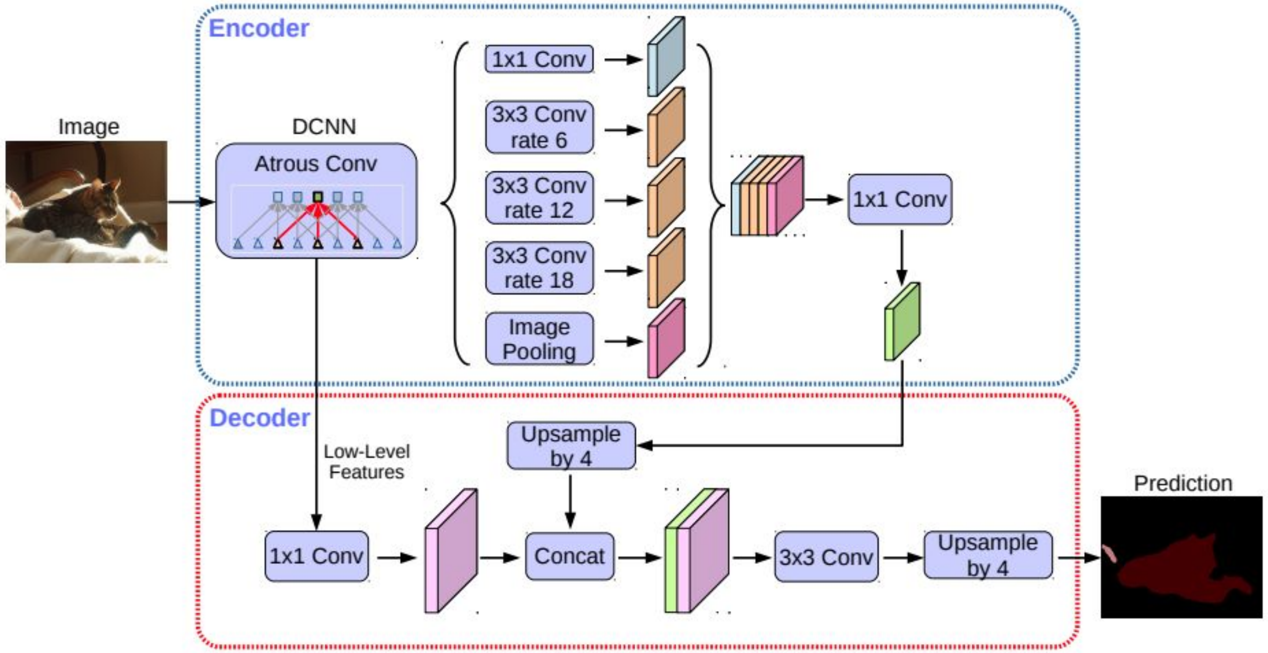
Skoltech

Models for experiments

U-net



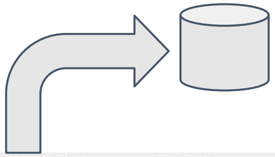
DeepLab v3plus

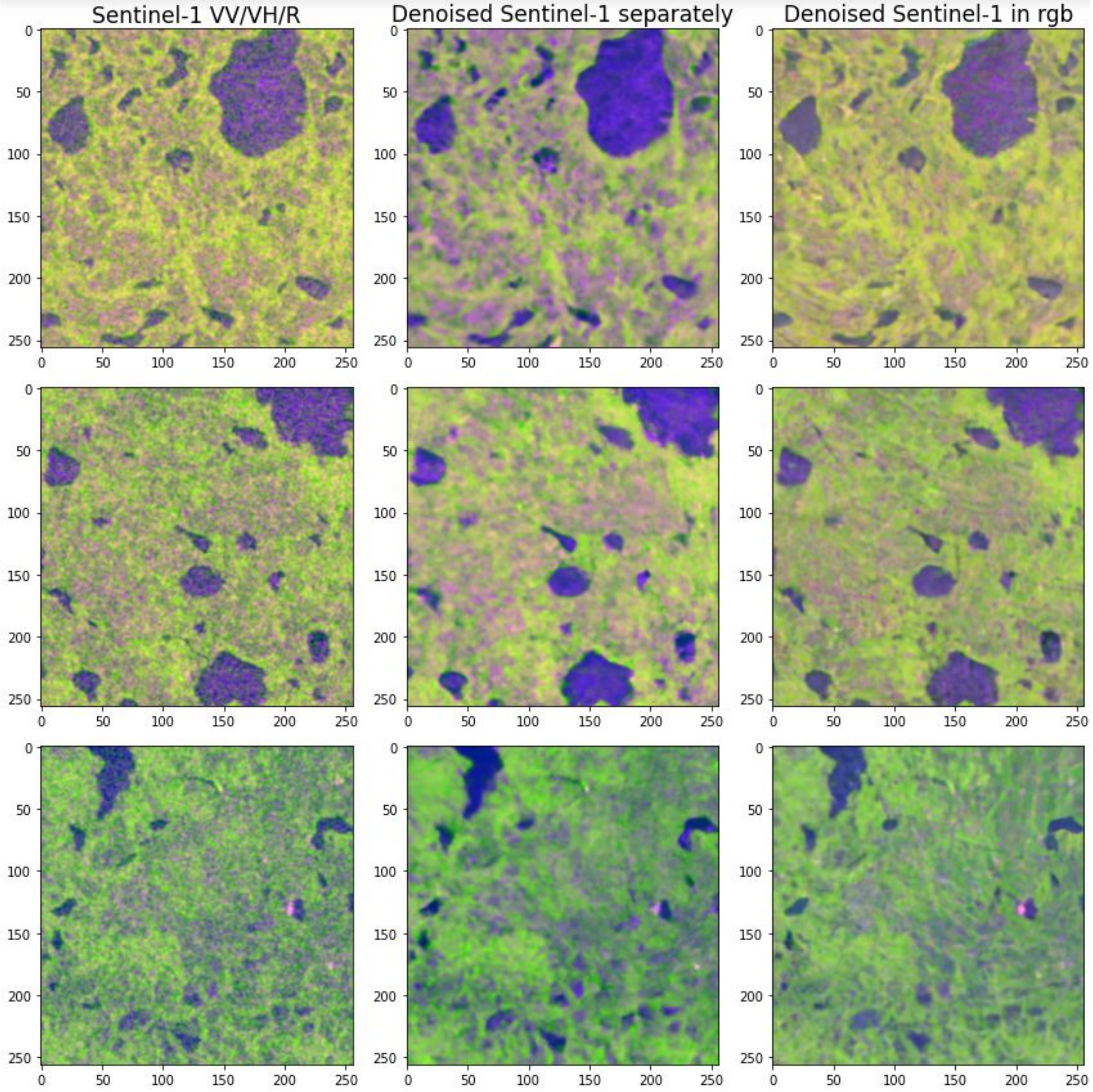


Skoltech

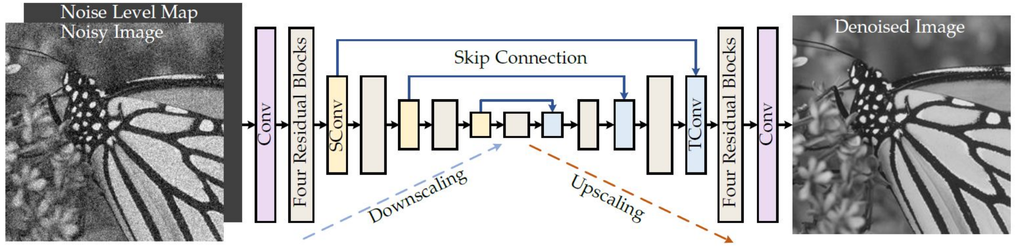
Denoising model model

enrich
the dataset





DRUNet

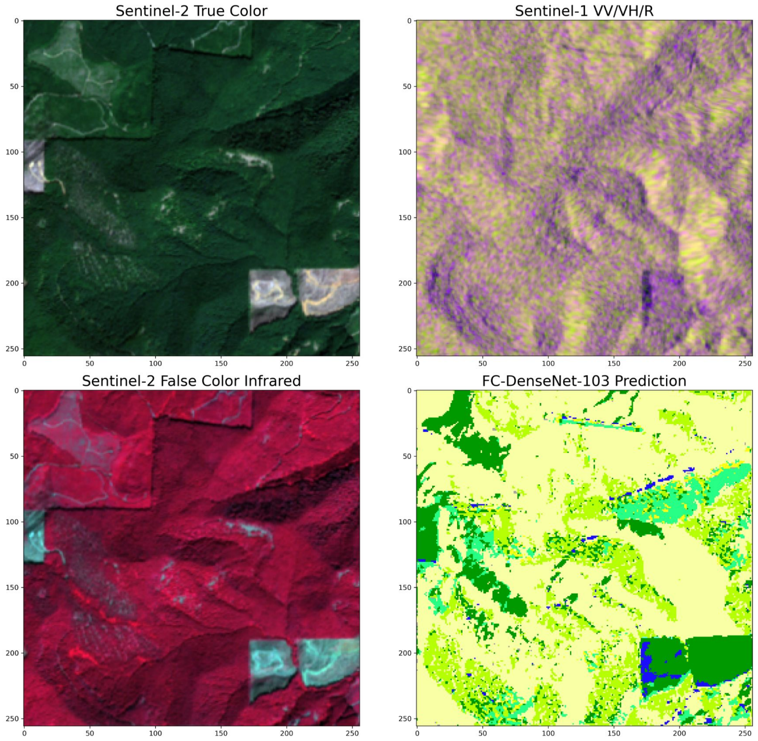


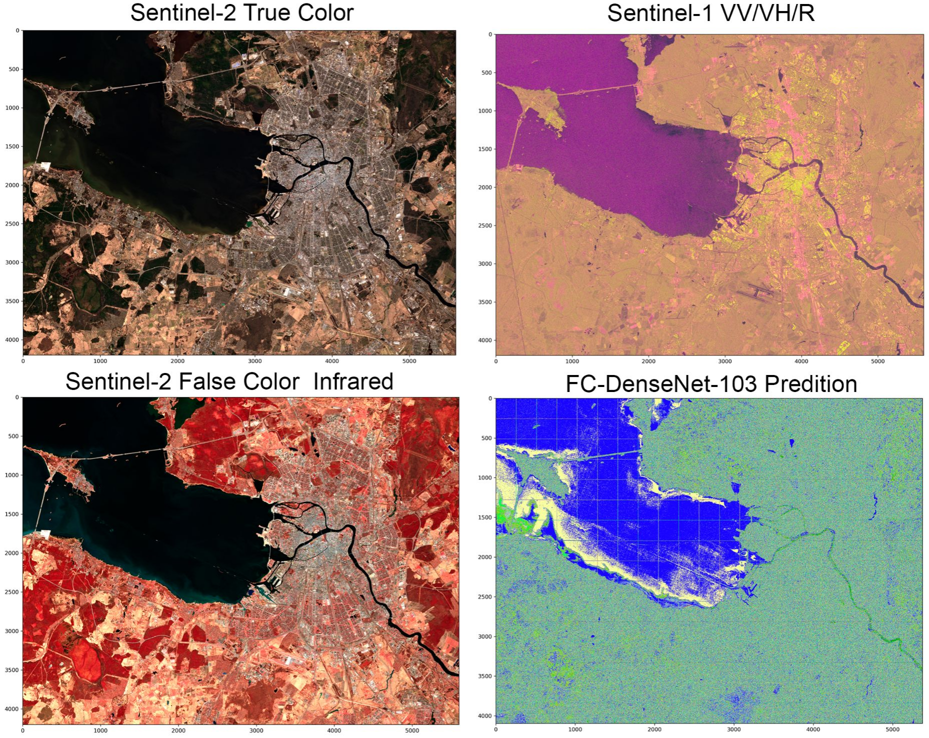
DRUNet gray
for denoise channels separately

DRUNet color
for denoise channels together in RGB space

Skoltech

Experiments: ResNet and DenseNet

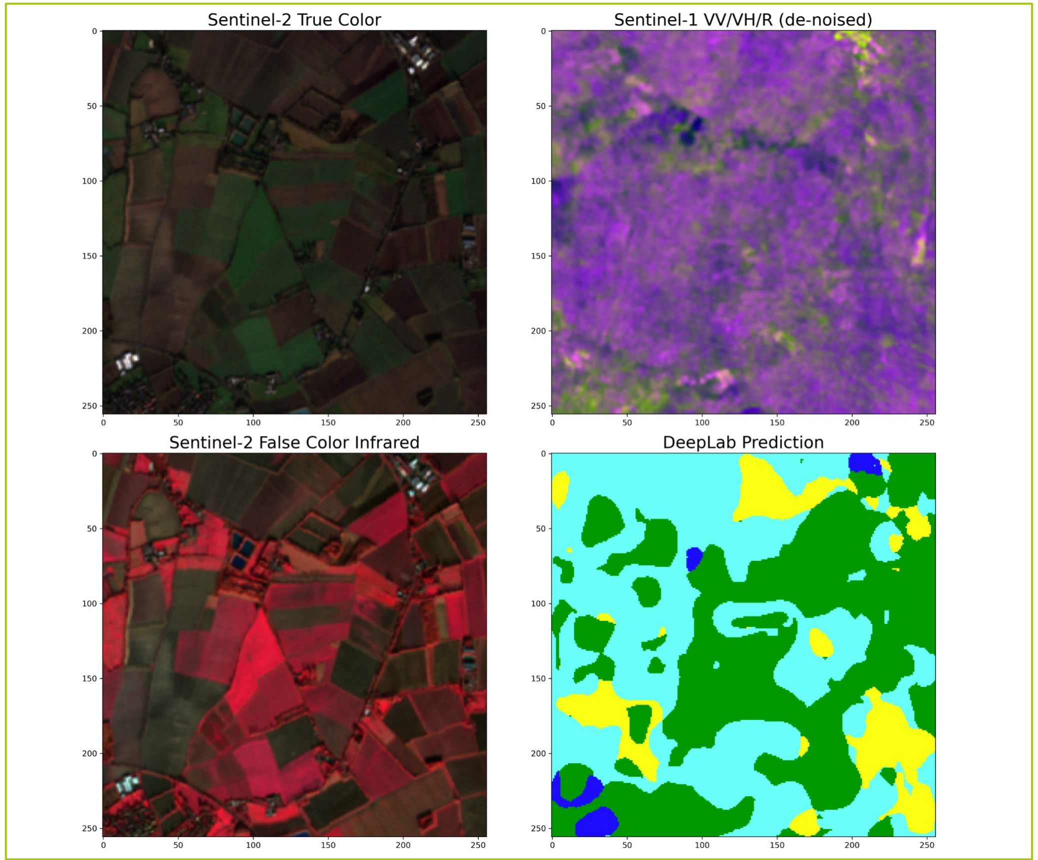




2 cases of the input:

1. 10 bands of Sentinel-2 (10 input channels);
2. Denoised Sentinel-1 + 10 bands of Sentinel-2 (12 input channels);

Experiments: DeepLab



Parameters:

Input size: bs x 256 x 256 x 10(12)
Batch size: 4
Loss: Categorical cross-entropy
Optimizer: Nesterov Adam
Initial LR: 0.0001

2 cases of the input:

1. 10 bands of Sentinel-2 (10 input channels);
2. Denoised Sentinel-1 + 10 bands of Sentinel-2 (12 input channels);

Skoltech

Results

Experiment	OA St. Petersburg	OA Rome	OA Munich
FC-DenseNet-103	26.13 %	0%	0.44%
FC-DenseNet-103 + de-noising	0.65%	0.13%	0.29%
DeepLab	0.12%	0%	0.038%
DeepLab + de-noising	33.27 %	0.06%	0.84%

Bad

Skoltech

Experiments: U-Net

- To define the best input set for U-Net segmentation model, applied to the SEN12MS

6 cases of the input:

1. Sentinel-1 + 3 bands (RGB) of Sentinel-2 (5 input channels);
2. Sentinel-1 + 3 bands (RGB) of Sentinel-2 with transformations (5 input channels) ;
3. filtered Sentinel-1 + 3 bands (RGB) of Sentinel-2 with transformations (5 input channels);
4. Sentinel-1 + 10 bands of Sentinel-2 (12 input channels);
5. Sentinel-1 + 10 bands of Sentinel-2 with transformations (12 input channels);
6. filtered Sentinel-1 + 10 bands of Sentinel-2 with transformations (12 input channels);

Parameters:

1. Transformations (non-destructive): *horizontal* and *vertical flips, rotation by 90* with probability=0.5;
2. Filter: *Lee filter* - to handle speckles of Sentinel-1;
3. Train/validation/test split: *75%/15%/10%*;
4. Batch size: 2
5. Loss function: *Cross-entropy*;
6. Optimizer: *Adam* with initial learning rate *lr = 0.0001*;
7. Scheduler: *ReduceLROnPlateu* with factor=0.5 and patience=4.

Experiments: U-Net

Quantitative results: Overall accuracies

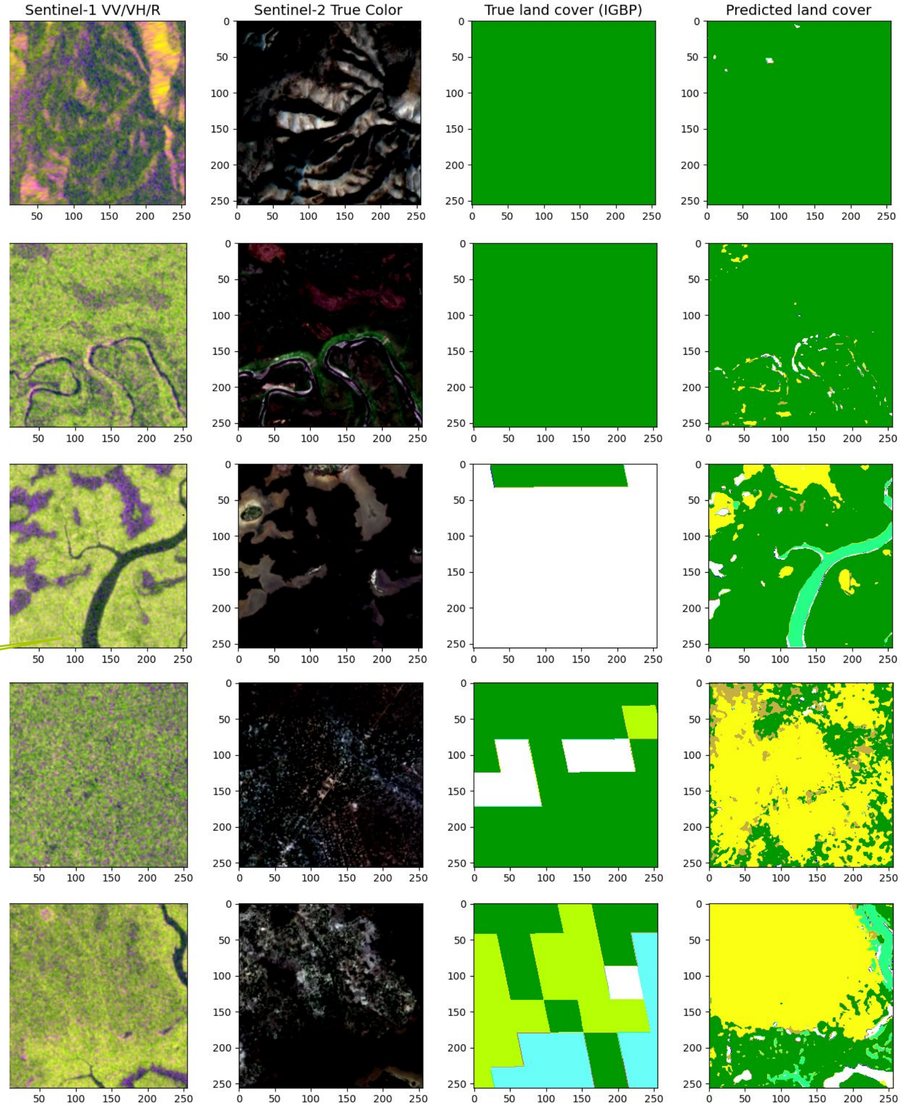
Case	Train set, %	Validation set, %	Test set, %
1	23.41	15.25	18.1
2	22.99	15.27	13.7
3	22.24	21.24	14.2
4	22.39	12.84	21.6
5	23.32	22.11	16.8
6	23.47	20.98	21.7

- The best OA accuracies on test correspond to the 4th and 6th input cases;
- The minimal val-test gap relates to the 6th case.

- Poor convergence with true labels;
- BUT. The model tends to comply with input data.

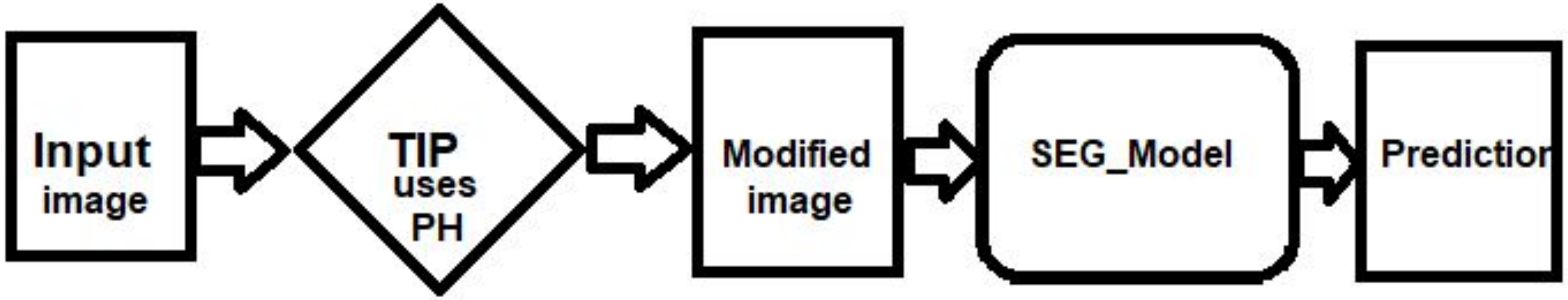
Problem of true labels (IGBP) reliability. Maybe, it's better to try manually assigned ones?

Qualitative results: Predicted maps for 6th case



Skoltech

Topological Image Processing (TIP)

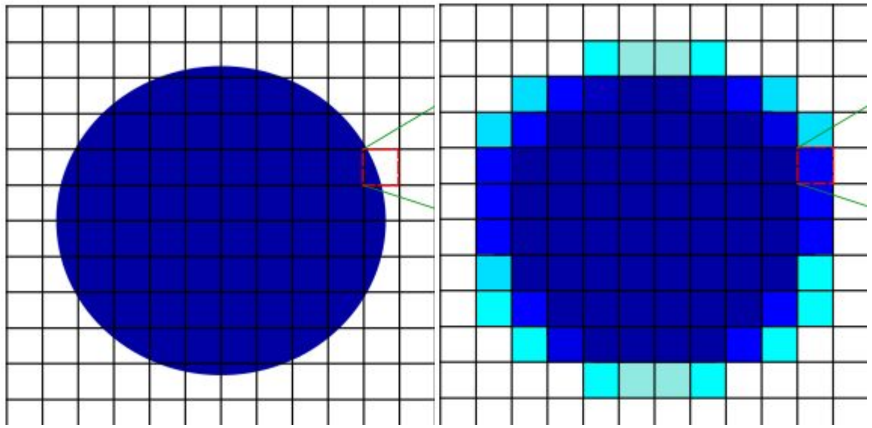


- The Idea is to use simple UNet model which takes topologically processed images.

- Artificially destruct irrelevant objects, and construct new images with known topological properties in irrelevant regions of an image, using Persistent Homology, etc.

Persistent Homology: Important tool of Algebraic Topology which, captures Global Structures in images/data.

Topological Processing



1. **Image Smoothening**

   Images smoothening is known to
   to get global structures in image, as it
   removes point-wise focus.

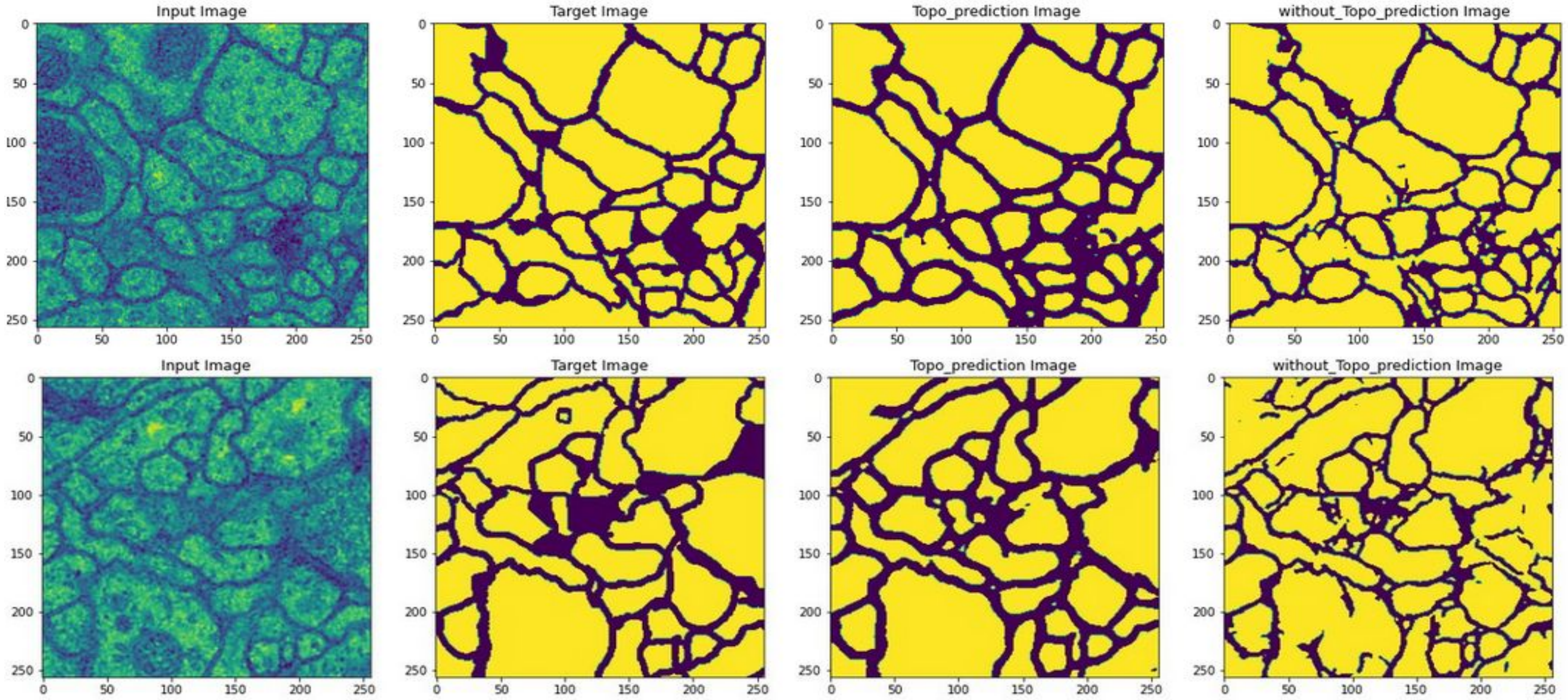
2. **Border - Modification**

   Using the topological information, like Betti-numbers (which give us information about the number of connected components, circular shapes, holes in images) and other Persistent diagrams.

   We modify the input images, In modified images irrelevant information is removed.

Skoltech

Result with Topological Image processing (TIP)



Model	Accuracy	Dice	Betti_Error
Without TIP	0.900	0.938	0.72
With TIP	0.908	0.938	0.20

Skoltech

Conclusions

According to the values of computed overall accuracies, the best input set was defined as filtered Sentinel-1 + 10 bands of Sentinel-2 with non-destructive transformations.

We still, however, believe that we have arrived at some important insights into the topic during the time we spent preparing and running our experiments and analyzing the results.

The first self-evident result is that even though the labels are of significantly lower resolution compared to the images, the nature of convolutional neural networks makes the resulting segmentation maps match the input in resolution.

We have come to an understanding that fully convolutional networks adapted for semantic segmentation are the weapon to choose when tackling the problem of semantic segmentation of remote sensing images.

thx.

Skoltech