Subpixel Classification of anthropogenic features using Deep Learning on Sentinel-2 Data

Nikolaos Ioannis Bountos[1,2], Melanie Brandmeier[1]

1. Esri Deutschland GmbH, Ringstraße 7, D-85402 Kranzberg; 2: TUM, Boltzmannstr. 3, D-85748 Garching
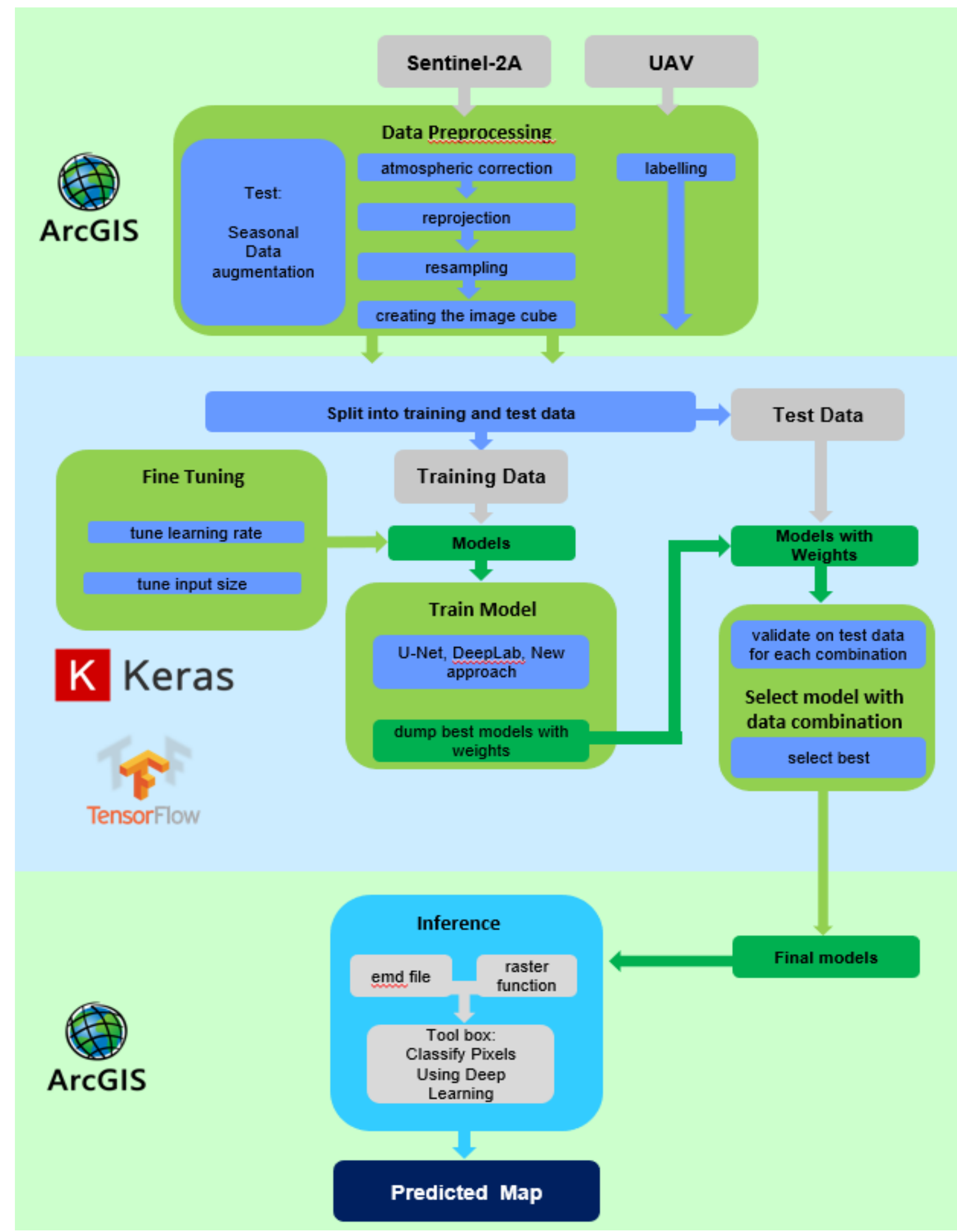
Background & Workflow

Fig 1. Workflow

Motivation

- Constantly changing urban areas.
- Periodical utilization of unmanned aerial vehicles for high resolution images is expensive.
- Increasing availability of satellite images such as Sentinel-2.
- Exploitation of the multispectral information contained in Sentinel-2 images.

Data and Workflow

- Data: Sentinel-2 resampled to 1m spatial resolution
- Features: Buildings, Rivers, Railways and roads.
- The workflow is shown in Fig. 1. In addition, we test the effect of using multitemporal data

Deep Learning models

- DeepLab[1] Model with three channel Input (RGB)
- DeepLab with nine channels as input
- A novel late fusion approach based on DeepLab
- U-Net[2] with three channels as input
- U-Net with nine channels as input
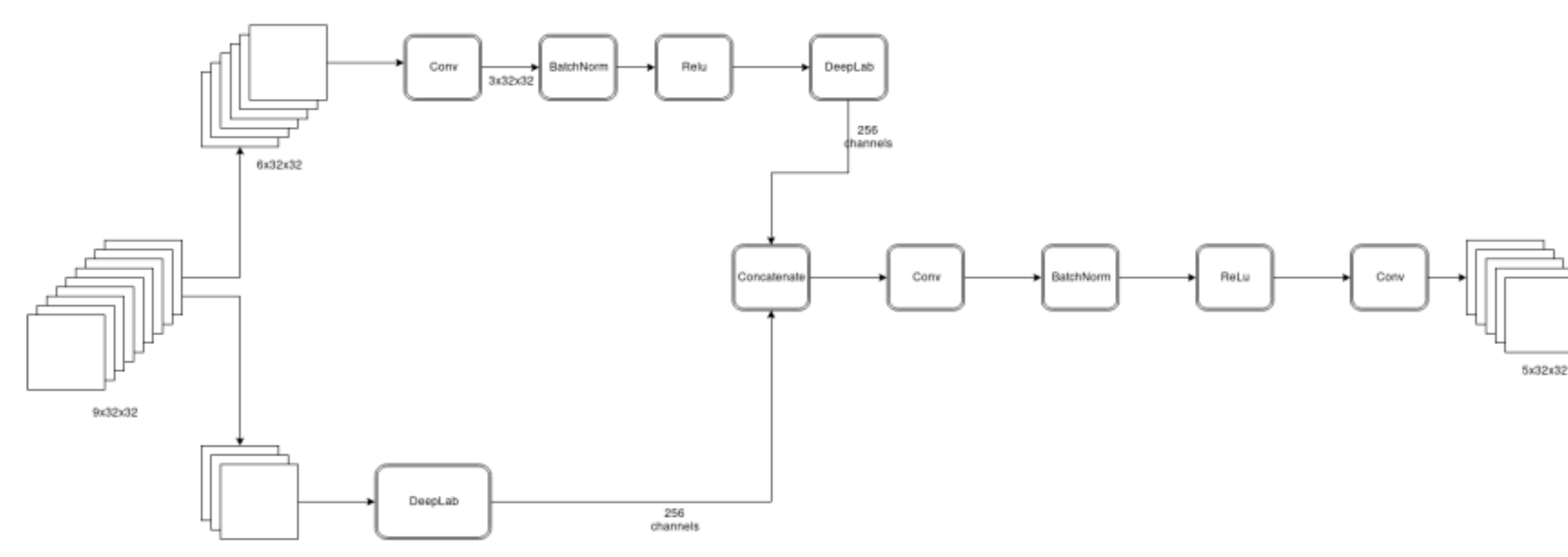
Fig 2. Late Fusion DeepLab

Results for our initial area
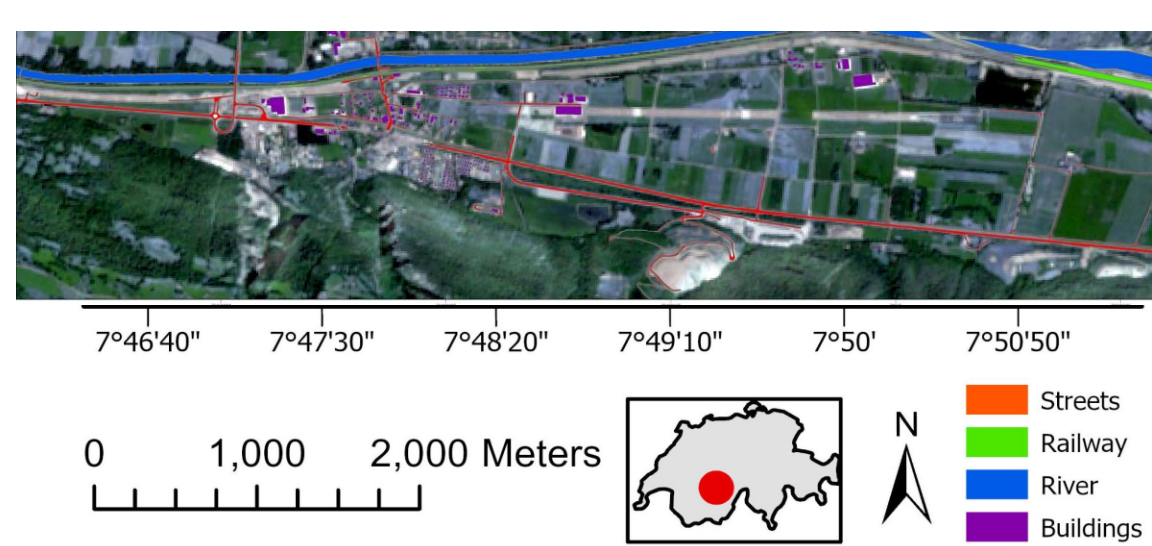
Fig 3. Ground truth

Fig 4. Nine Channel DeepLab

Fig 5. Random Forests prediction

Initial Dataset
Multitemporal Dataset

Effect of multitemporal data on predictions

Method	Overall Accuracy	Buildings	Streets	Railway	River	Background
U-Net 3 channels	10%	24%	19%	3%	49%	2%
U-Net 9 channels	11%	58%	2%	5%	14%	8%
DeepLab 3 channels	88%	50%	54%	69%	93%	94%
DeepLab 9 channels	89%	60%	59%	73%	92%	94%
Late Fusion DeepLab	88%	55%	51%	71%	93%	94%

Table 1. Overall accuracies of models trained on the initial dataset

Method	Overall Accuracy	Buildings	Streets	Railway	River	Background
U-Net 3 channels	16%	5%	13%	54%	15%	15%
U-Net 9 channels	39%	0%	6%	37%	11%	49%
DeepLab 3 channels	87%	36%	44%	67%	94%	94%
DeepLab 9 channels	87%	33%	48%	74%	95%	94%
Late Fusion DeepLab	87%	30%	50%	71%	94%	93%

Table 2. Overall accuracies of models trained on multitemporal data

Transferability to a different area

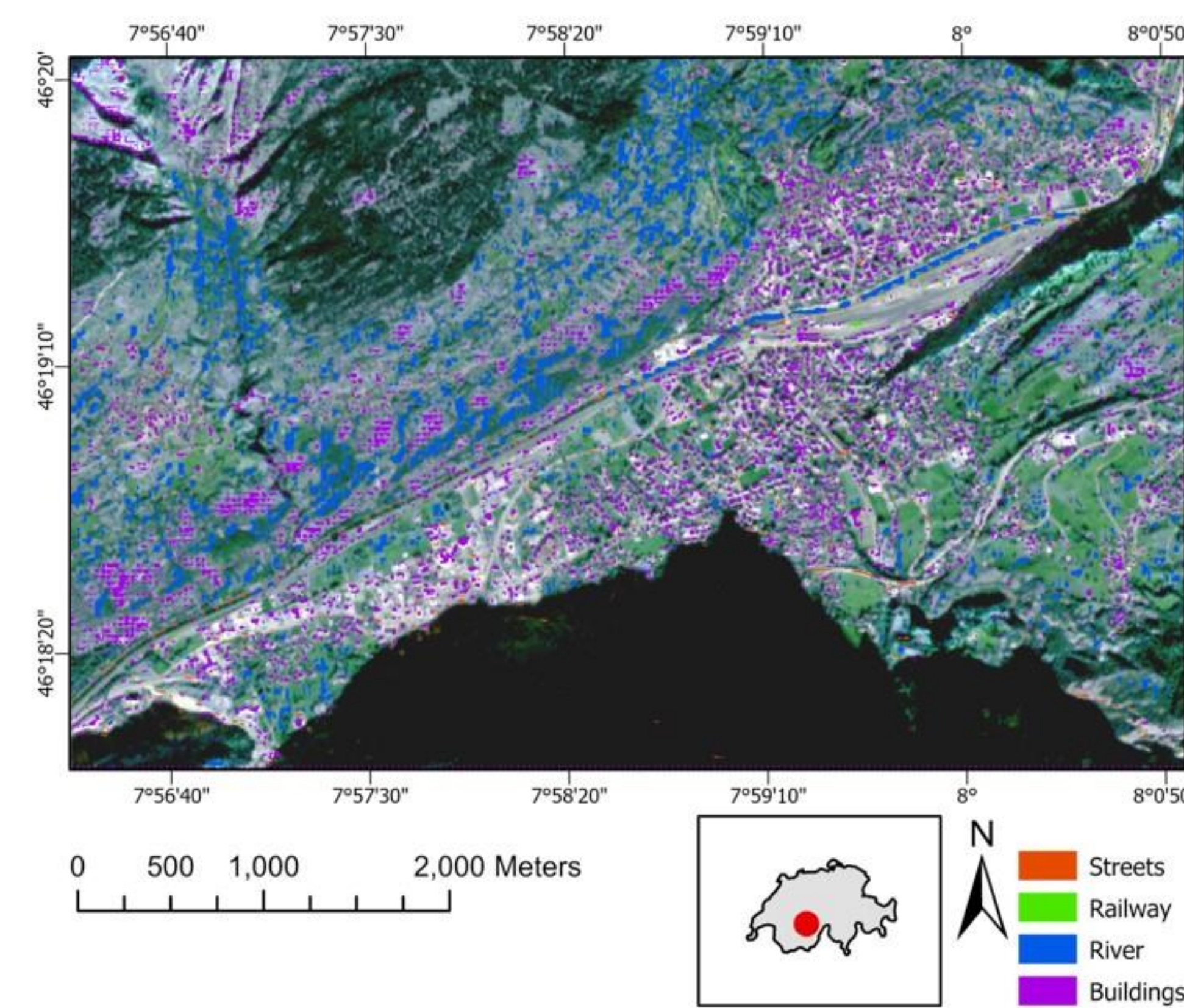### Example: Nine Channel DeepLab

Initial Dataset
Multitemporal Dataset

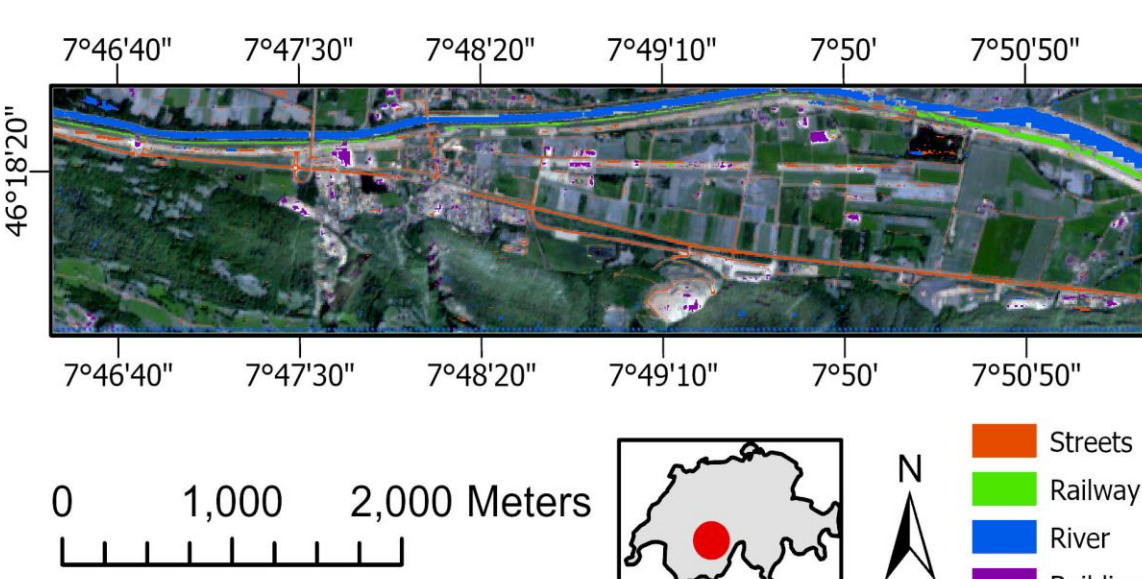Fig 7. Nine channel DeepLab prediction, trained on initial dataset

Fig 8. Nine channel DeepLab prediction, trained with augmented dataset.
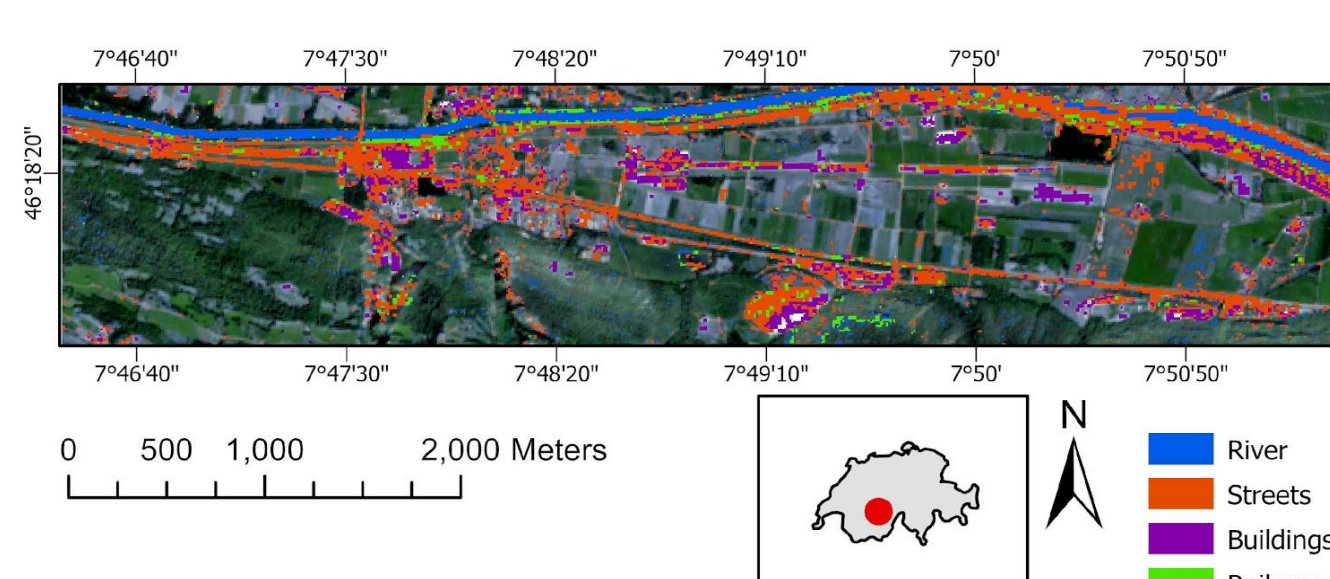
Conclusions

- It is possible to identify features such as Buildings, Rivers, Streets and Railways that are partly represented on a subpixel level in Sentinel-2 using advanced deep-learning approaches.
- Accuracy, however, is still not enought for monitoring purposes.
- Using multitemporal data significantly improves the power of generalization of the architectures
- The integration into GIS could help planning an support decision-making

[1] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017.

[2] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.