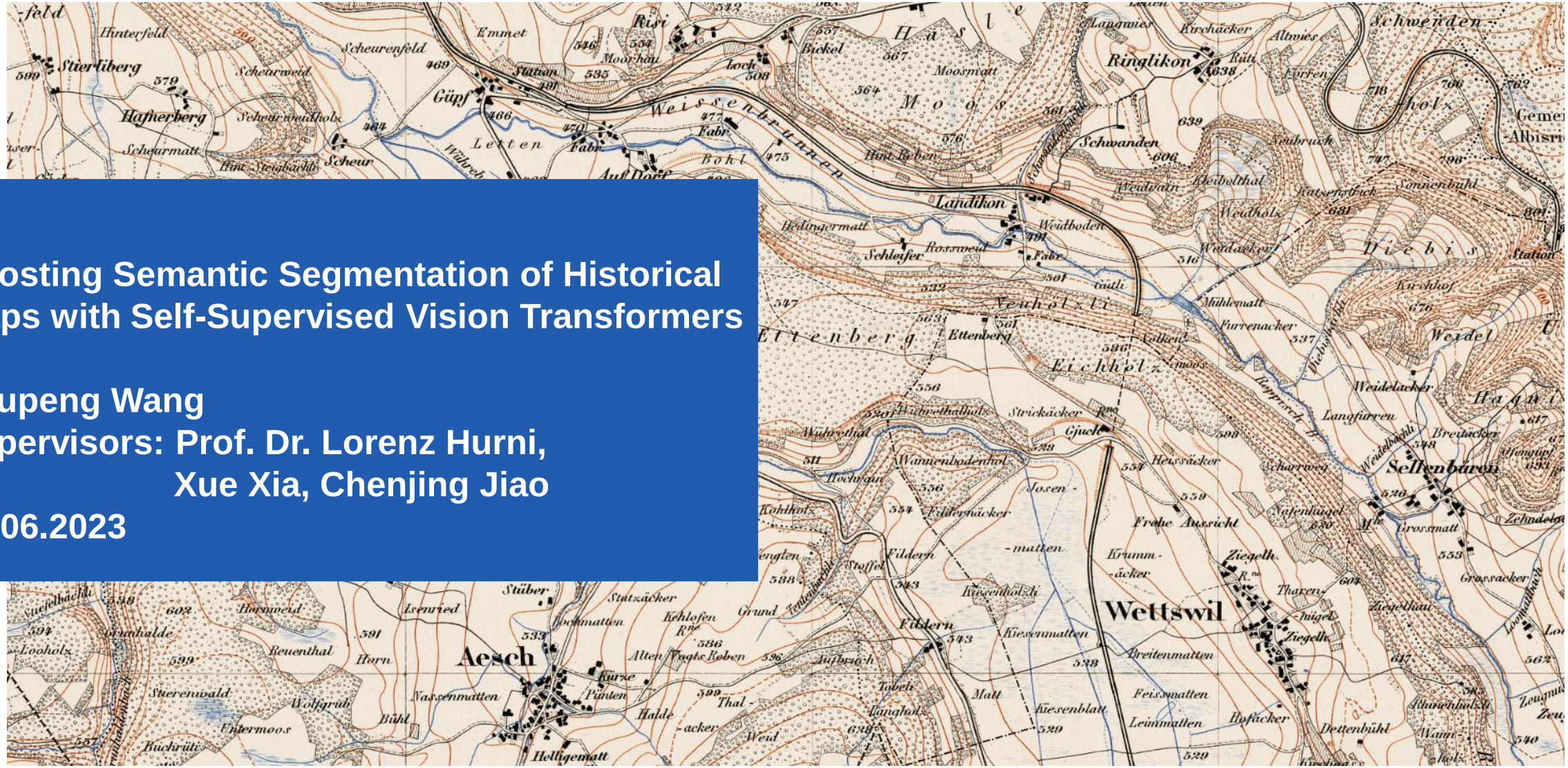
ETH zürich

IKG
Institut für Kartografie und Geoinformation
Institute of Cartography and Geoinformation



Boosting Semantic Segmentation of Historical Maps with Self-Supervised Vision Transformers

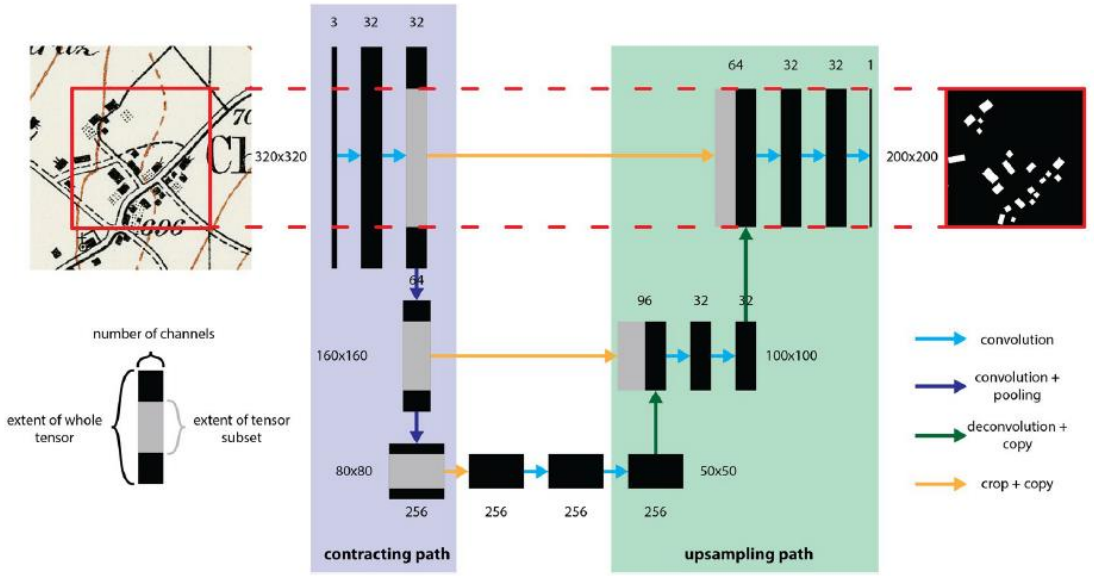
Shupeng Wang
Supervisors: Prof. Dr. Lorenz Hurni,
Xue Xia, Chenjing Jiao

01.06.2023

Agenda

1. Background and goals
2. Data
3. Methodology
4. Result
5. Conclusion and future directions
6. Reference

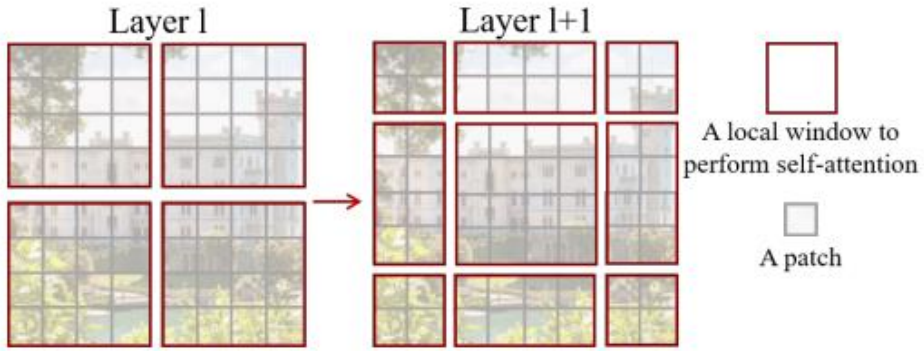
Background and goals



(Heitzler & Hurni, 2020)

CNNs

?



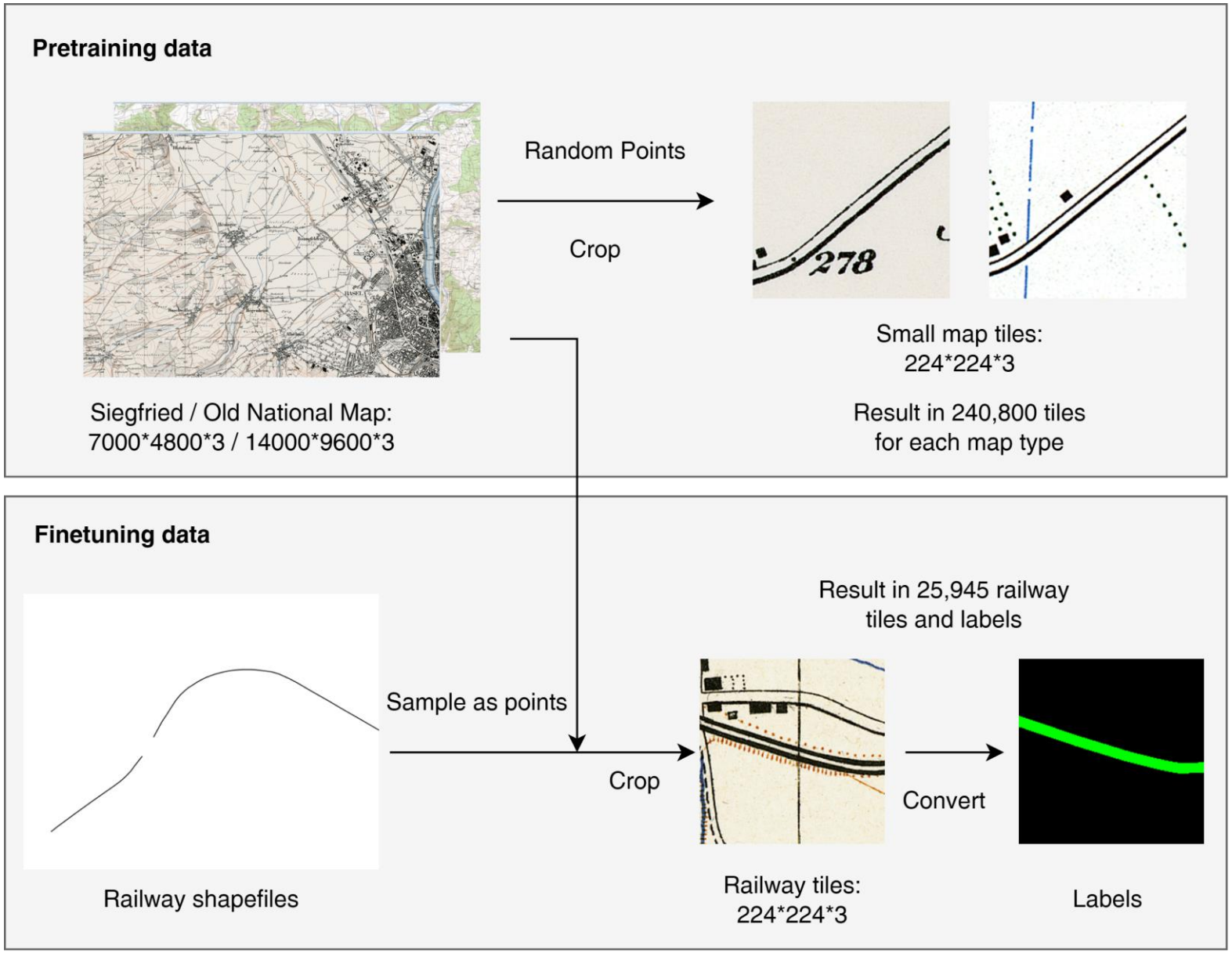
(Liu et al., 2021)

Swin Transformer

(Understanding Contrastive Learning)

Contrastive Learning

Data

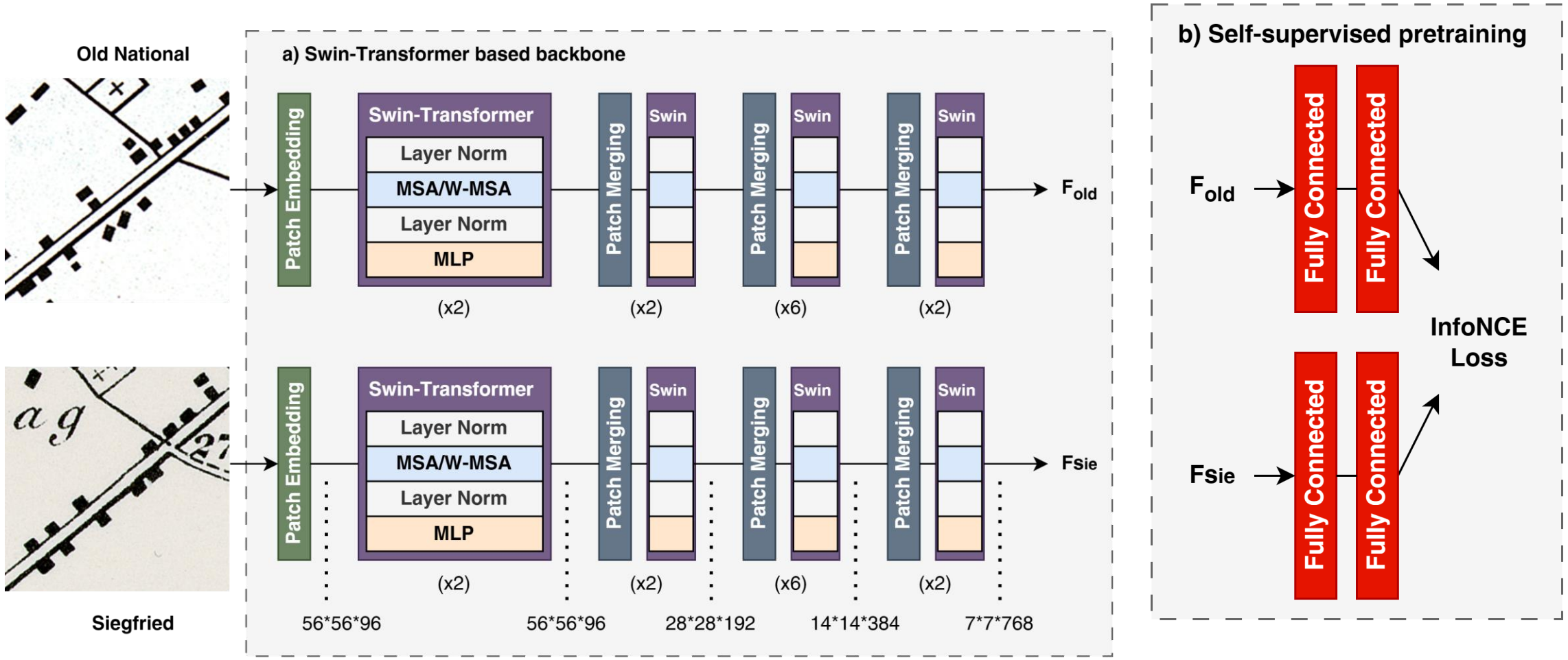


Also mixed with road tiles.

Small datasets are also prepared (10%, 5%, 2.5%)

Methodology

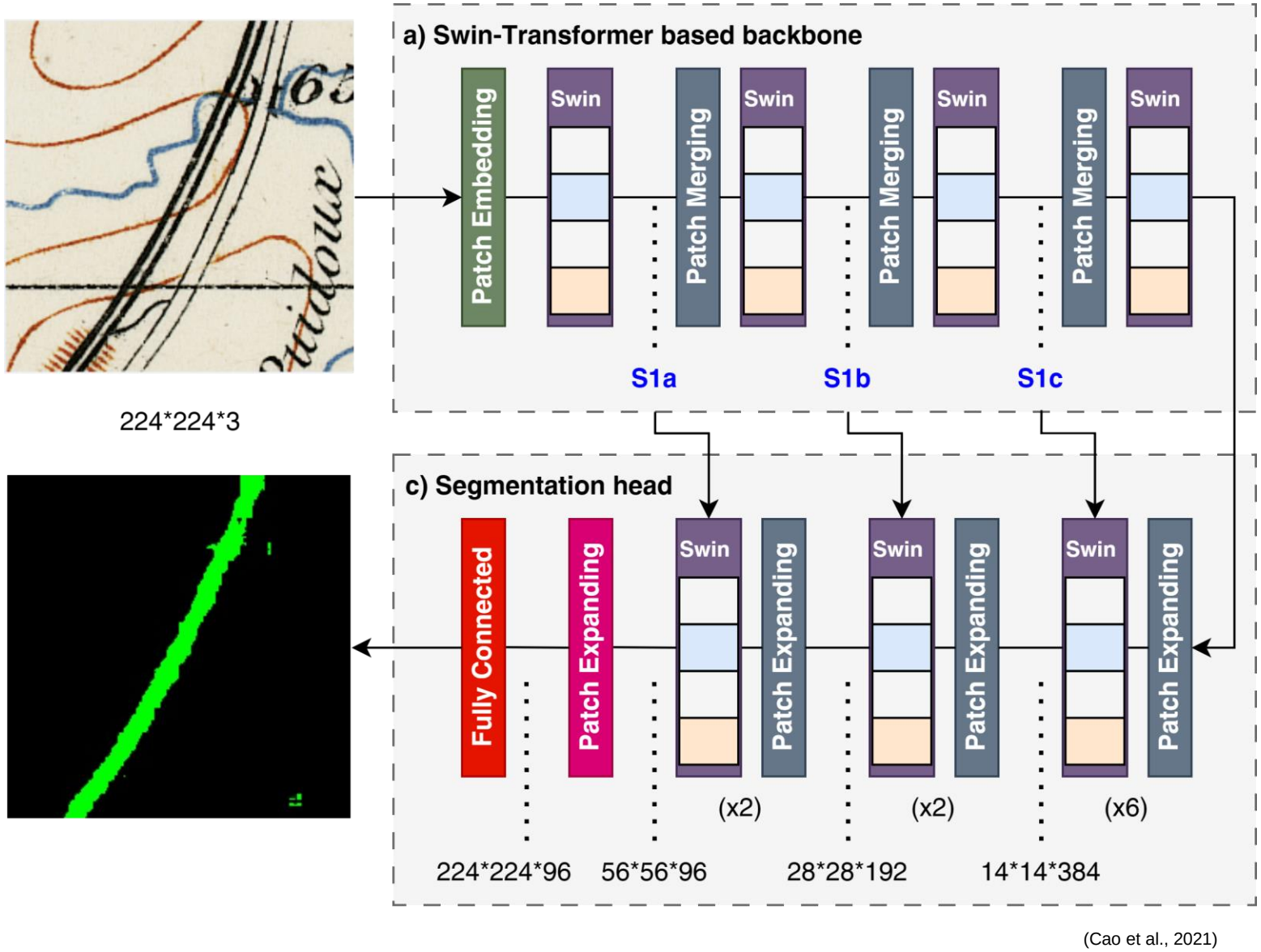
- Pretraining part



(Chen et al., 2020; Liu et al., 2021; Scheibenreif et al., 2022)

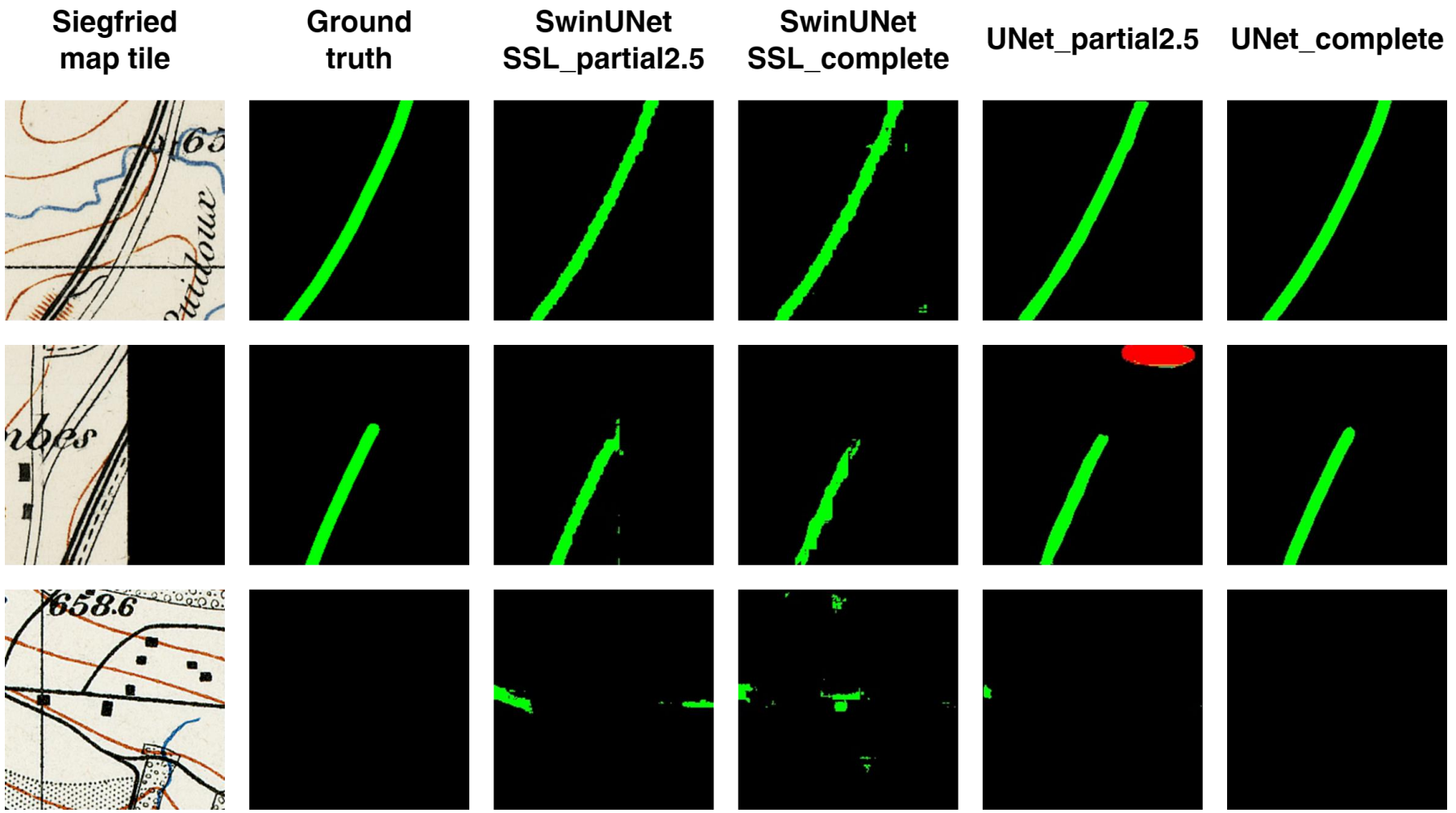
Methodology

- Finetuning part



(Cao et al., 2021)

Result and discussion



Model performance on the semantic segmentation of normal railways

Result and discussion

Conclusion and future directions

Strength

- Good performance with small training dataset
- Efficient training speed

Limitation

- Fair performance with large training dataset
- Difficulty in handling imbalanced training classes

Future directions

- Improve SimCLR performance (batch size + data augmentation)
- Explore alternative pretraining strategies (BEiT or MAE)
- Adopt different loss functions during finetuning (focal loss)

Reference

- Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2021). *Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation*. http://arxiv.org/abs/2105.05537
- Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). *A Simple Framework for Contrastive Learning of Visual Representations*. http://arxiv.org/abs/2002.05709
- Heitzler, M., & Hurni, L. (2020). Cartographic reconstruction of building footprints from historical maps: A study on the Swiss Siegfried map. *Transactions in GIS*, *24*(2), 442–461. https://doi.org/10.1111/tgis.12610
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). *Swin Transformer: Hierarchical Vision Transformer using Shifted Windows*. http://arxiv.org/abs/2103.14030
- Scheibenreif, L. M., Hanna, J., Mommert, M., & Borth, D. (2022). *Self-supervised Vision Transformers for Land-cover Segmentation and Classification*. *https://openaccess.thecvf.com/content/CVPR2022W/EarthVision/papers*

Shupeng Wang
shupwang@student.ethz.ch

ETH Zürich
Poster Presentation
Spring Semester 2023