Project Seminar, 2024

Land Cover Classification based on Multiscale Time Series of Satellite and Aerial Images

Supervisors:
M.Sc. Hubert Kanyamahanga,
Dr.-Ing. Mareike Dorozynski

Students:
Mahsa Ghandali
Sunanda Menon
Shiva Timilsina
Mohammad Mansouri
Suman Khanal

Overview

- Introduction
- Methodology
  - Baseline architecture
  - Self-Supervised Learning (SSL)
- Experiments
  - Dataset and experimental setup
- Results
- Conclusion

Introduction

- **Motivation**

  - To achieve **high quality Land Cover (LC)** predictions.
  - Advantage of combining temporal and spatial information.
  - Land Cover Classification for **Hameln** region.

Satellite Image Time Series (SITS)

Aerial image

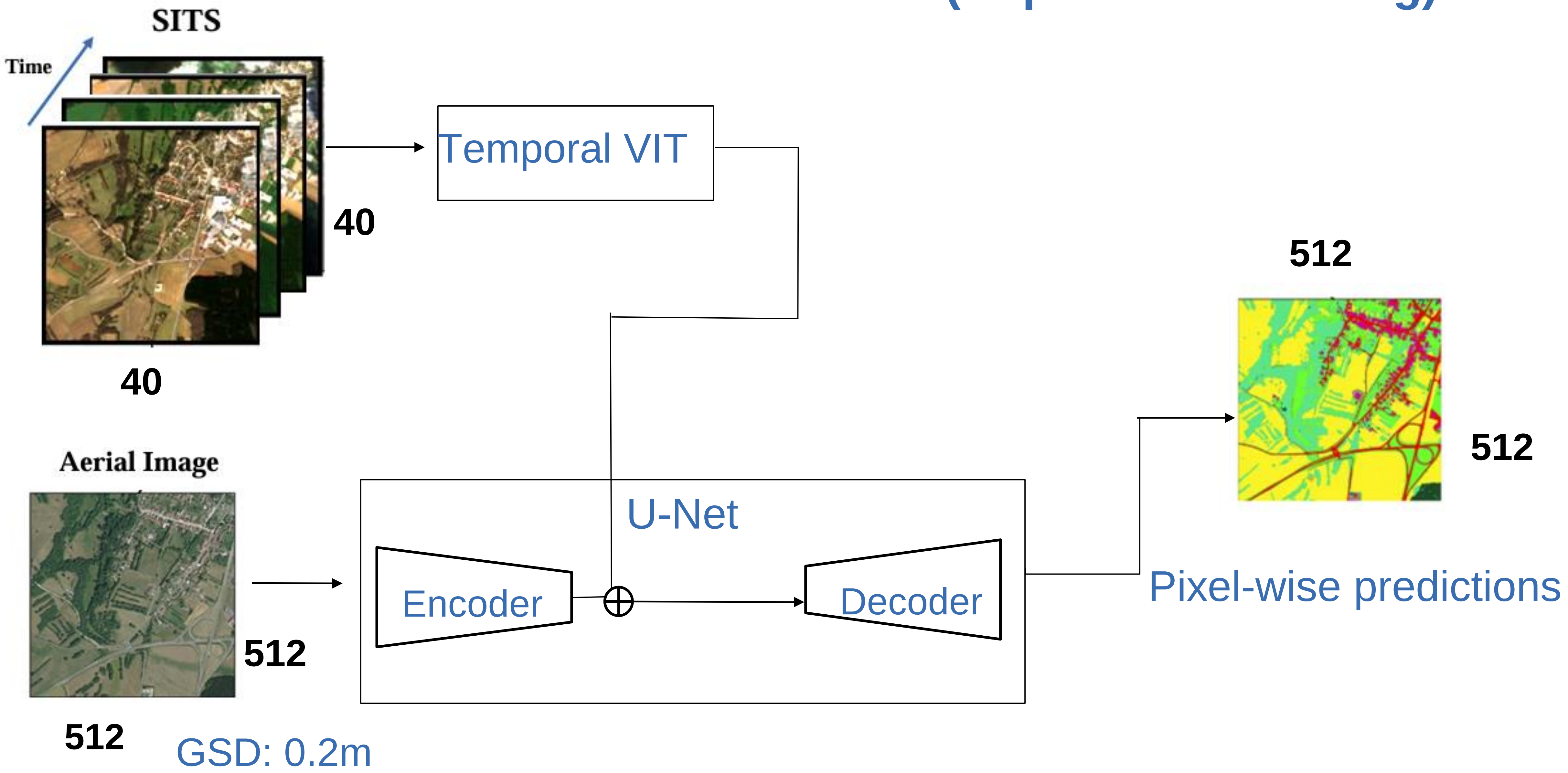
Introduction

- **Challenges**
  - Integrating **multi sensor** data.
  - Availability of very **few** high-quality training labels.

- **Our approach**
  - **Self-Supervised Learning** (SSL)
    - ❖ Large unlabelled data for pre-training and small labelled data for fine-tuning.
    - ❖ SSL has been shown to outperform Supervised Learning method in the literature.

GSD: 10m

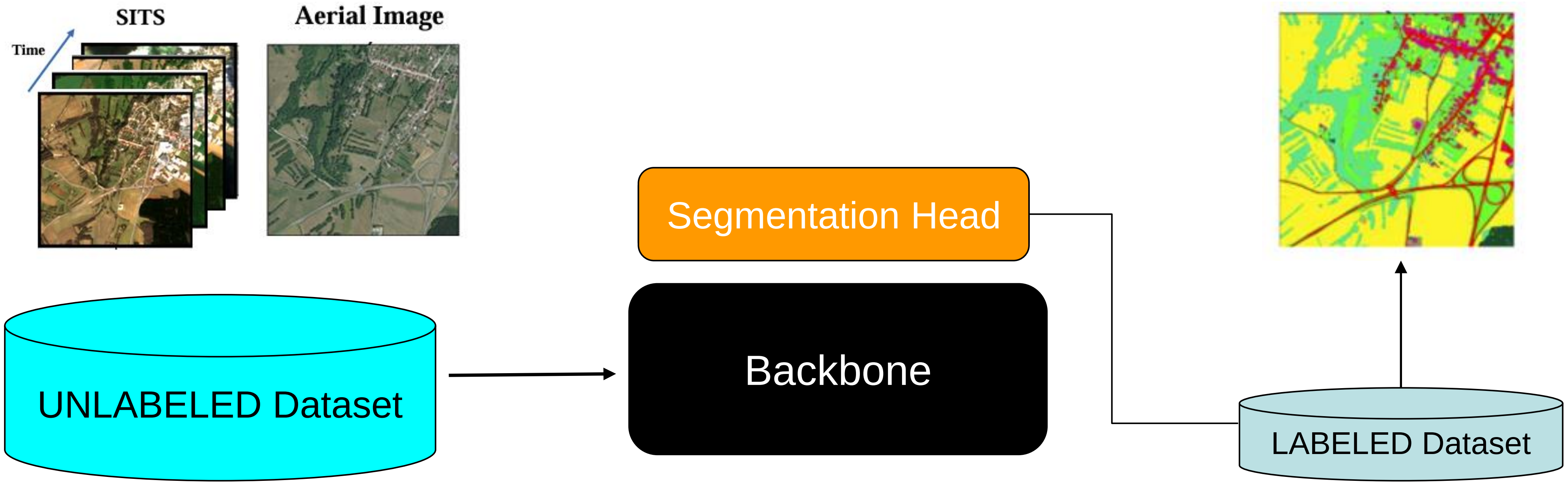
Baseline architecture (Supervised learning)



Self-Supervised Learning (SSL)

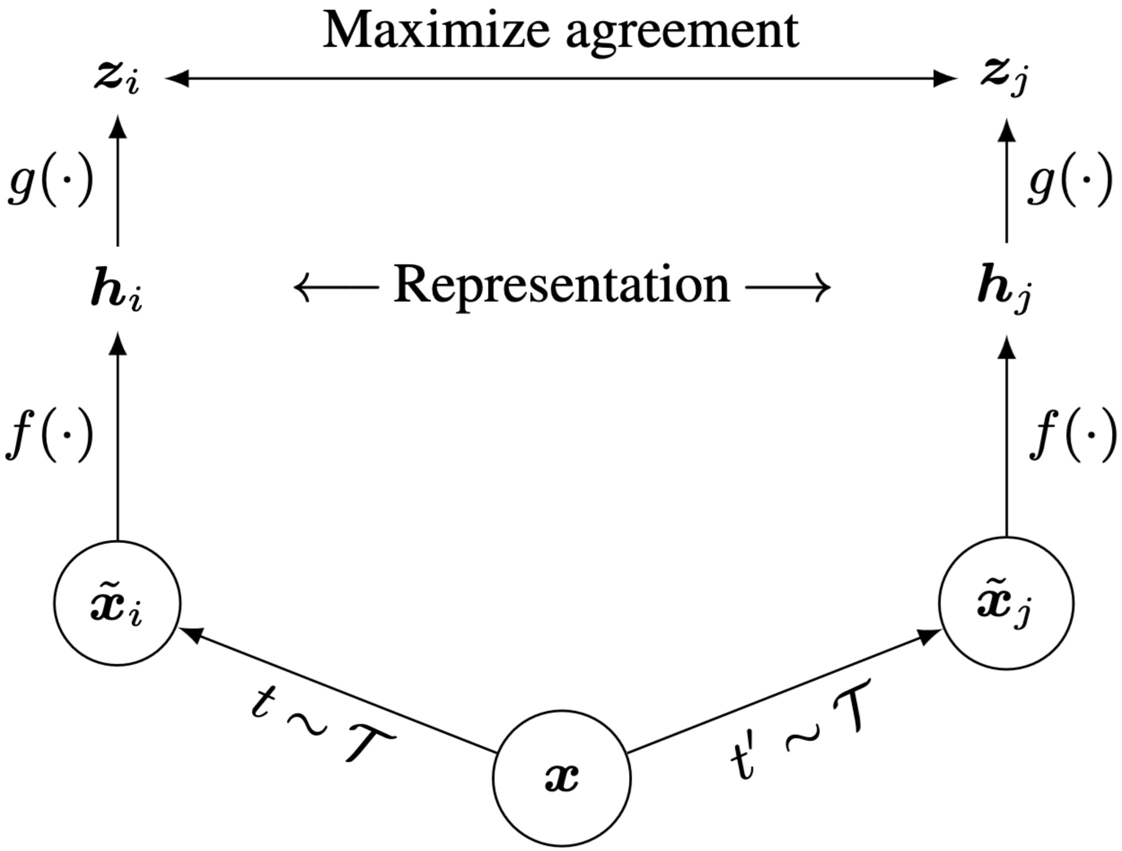
1. Self-supervised pre-training

2. Supervised fine-tuning

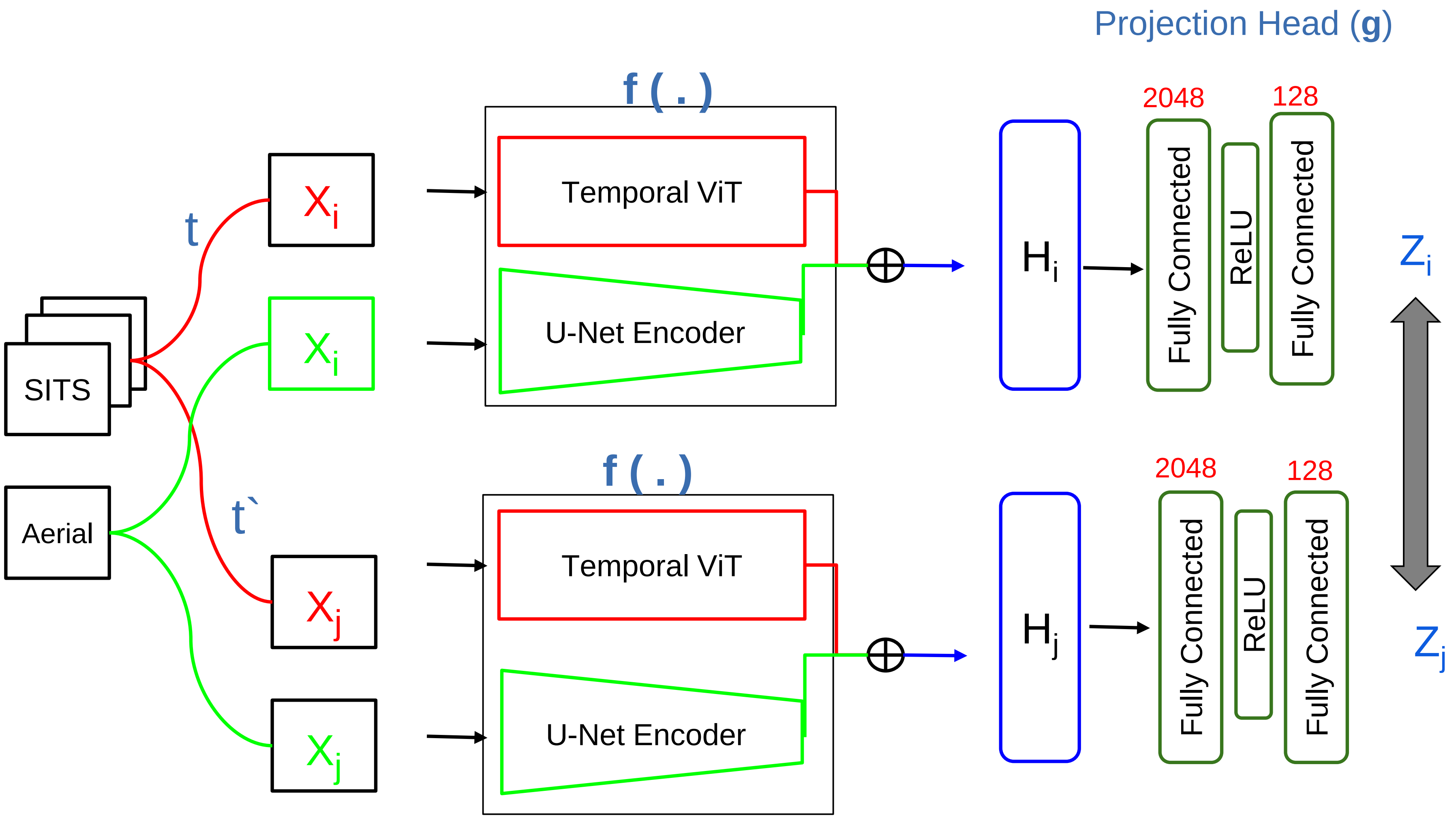


Contrastive Learning

A Simple Framework for Contrastive Learning of Visual Representations (**SimCLR**)



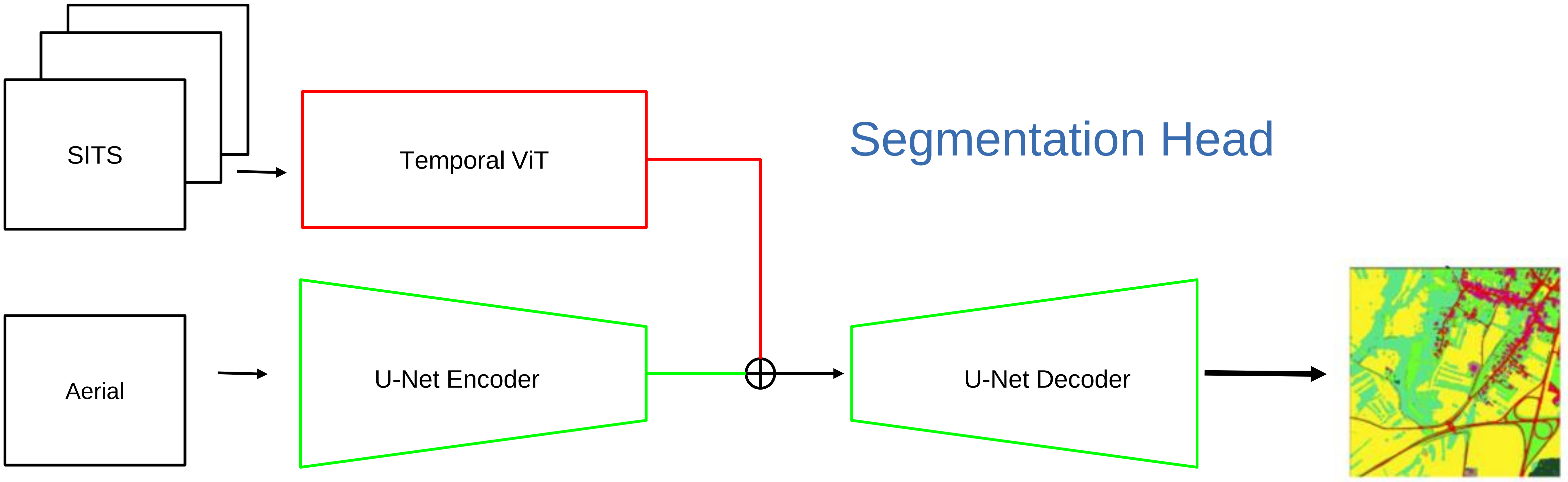
- **Transformation: t , t`**
- **Views: x_i , x_j**
- **Encoder: f (.)**
- **Feature map: h_i , h_j**
- **MLP(Projection): g(.)**
- **Feature Map(MLP): z_i , z_j**



Contrastive Learning

- **Transformation : t , t`**
- **Views: x_i , x_j**
- **Encoder : f (.)**
- **Feature map : h_i , h_j**
- **MLP(Projection) : g(.)**
- **Feature Map(MLP): z_i , z_j**

Supervised fine-tuning



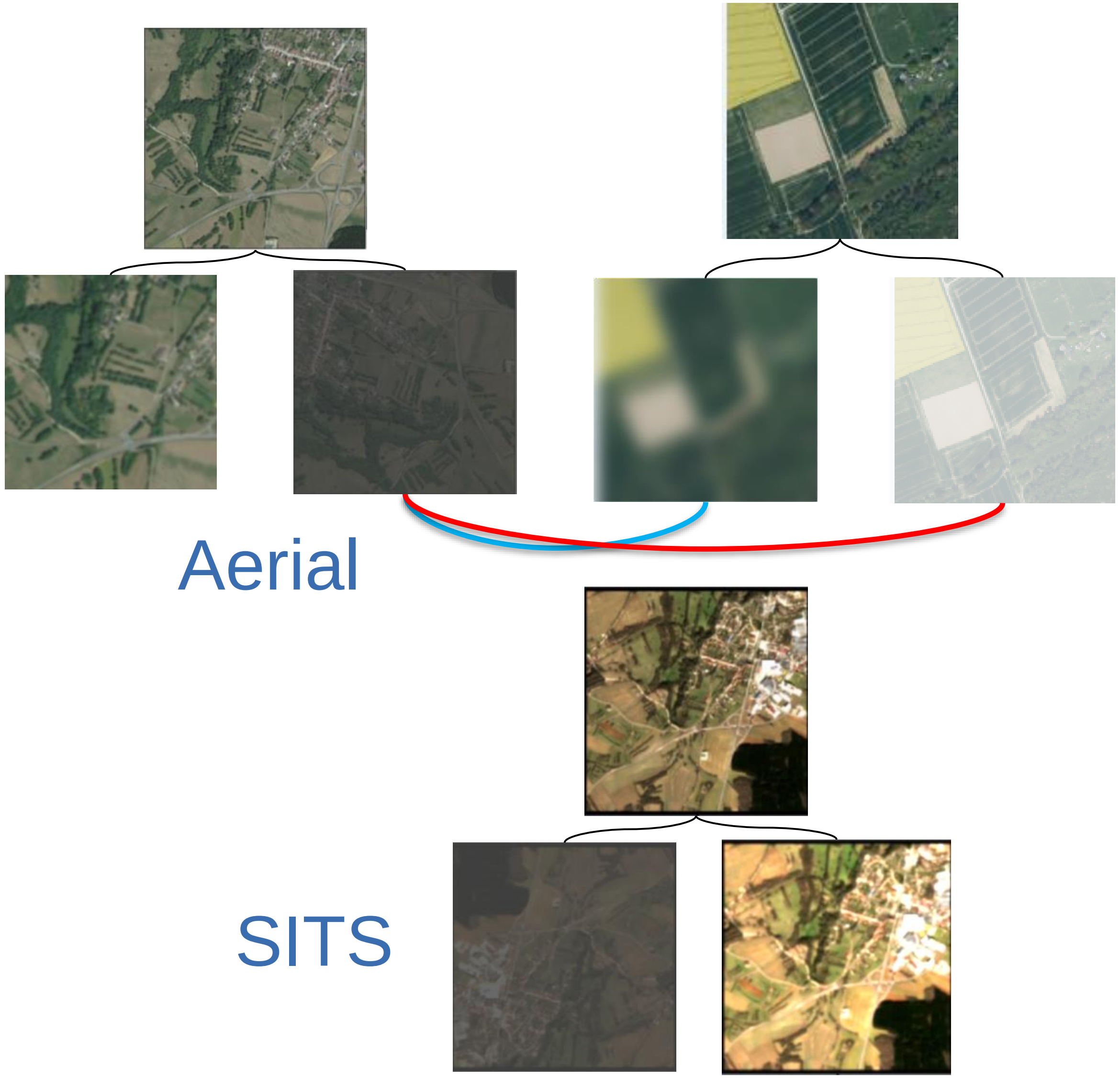
Pixel-wise predictions

Positive and Negative pair generation

Data Augmentation

Geometric and Radiometric transformation

- **Cropping randomly and resizing (20%)**
- **Gaussian blur**
- **Erasing part of image**
- **Brightness and contrast changes**
- **Simulation of different lighting condition**



Contrastive learning loss

$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)}$$

- **$l_{i,j}$** : Loss for the image positive pair (i, j)
- **sim (z_i , z_j)** : Similarity measure between feature vectors z_i and z_j (cosine similarity)
- **τ**: Temperature parameter that scales the similarities
- **2N**: Total number of images in the batch
- **$1_{[k\neq i]}$** : Indicator function that equals 1 if k≠i, and 0 otherwise
- **exp(·)** : Exponential function

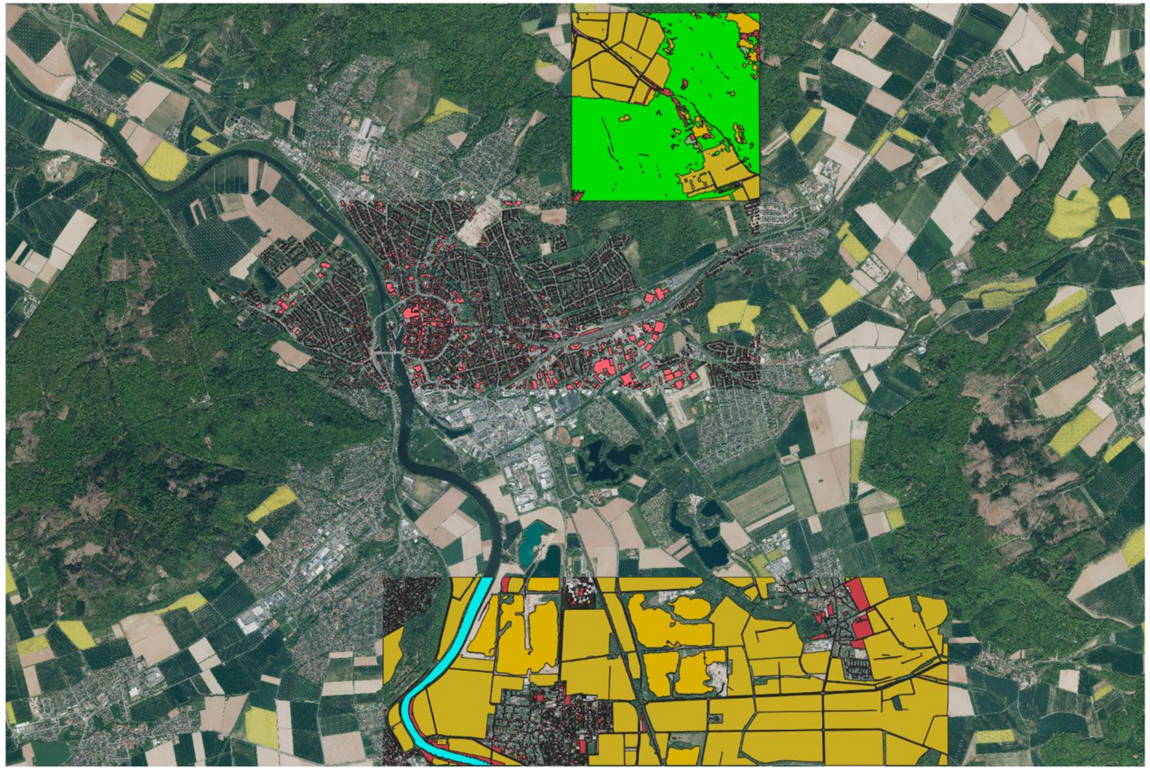
Data

Hameln:

Pre-training

9322 Patches

Fine-Tuning



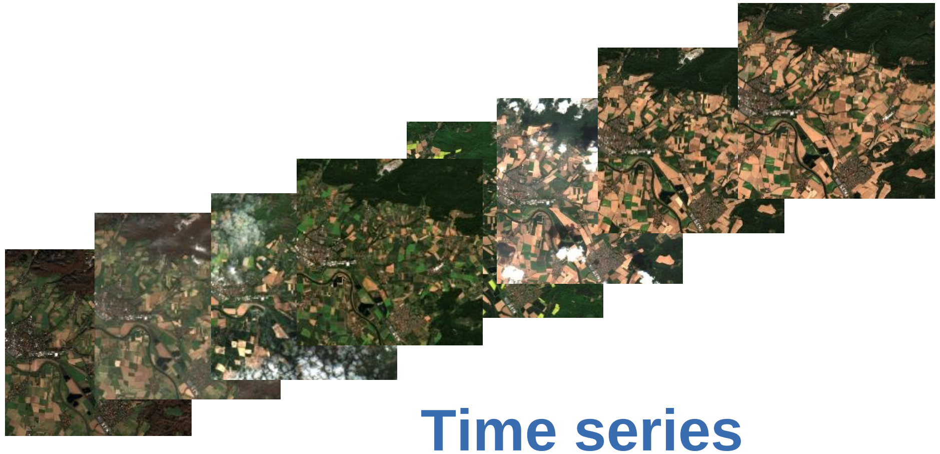
2077 Patches

Data

Hameln:

- ❖ SITS
  - ➢ Feb - Nov
  - ➢ GSD: 10 m
  - ➢ Channels: 10
  - ➢ Patch size: 40*40 pixels
  - ➢ length of time series:8

- ❖ Aerial
  - ➢ Mono-Temporal
  - ➢ GSD: 0.2 m
  - ➢ Channels: 4(R,G,B,IR)
  - ➢ Patch size: 512*512 pixels



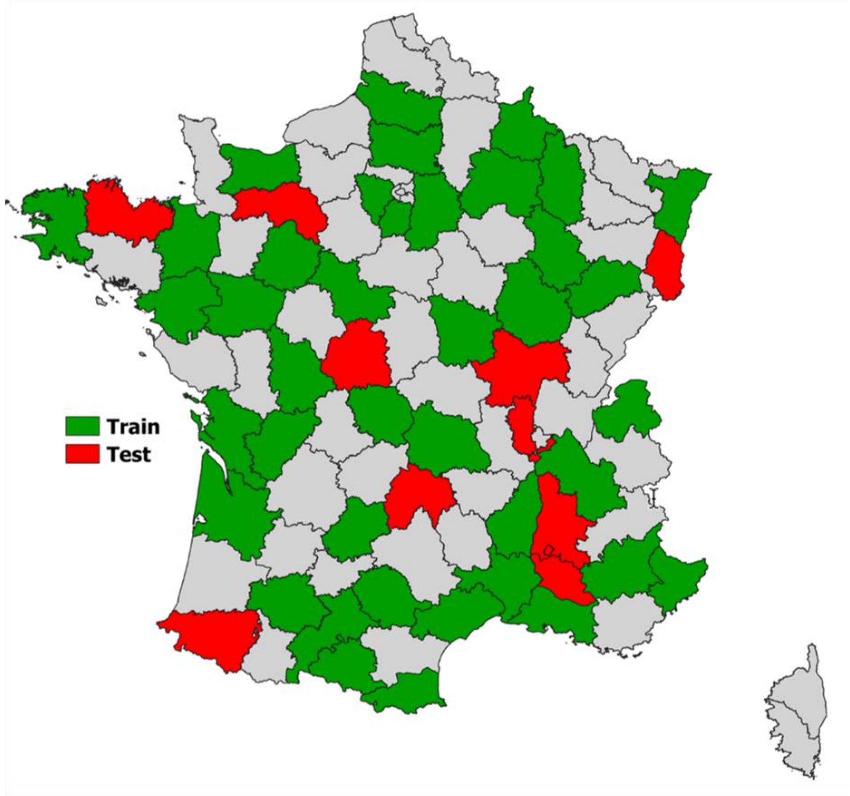
Time series

Aerial patch

Data

FLAIR #2

- ➢ Open-source dataset (France)
- ➢ Task: LC classification
- ➢ Contains aerial and SITS images
- ➢ Labelled dataset



50 areas in France

Experiments

SSL

#	Pre-Training Data	Pre-Training Epochs	Fine-tuning Data	Fine-tuning Epochs	Batch	Temp
1	Baseline (FLAIR 2)	30	Baseline	5	16	0.1
2	Hameln	40	Hameln	20	16	0.1
3	Hameln	40	Hameln	20	8	0.7

Supervised Learning

#	Dataset	Epochs
1	Baseline	30
2	Hameln	50

Setup

- Optimiser: Stochastic gradient descent with learning rate of 0.001
- Supervised fine-tuning : Cross entropy loss with adjusted weights for imbalance class handling.
- Batch size: 8, 16
- Projection head: 2-layers FCN, ReLU activation
- Temperature parameter : 0.7, 0.1
- Evaluation metrics: IoU

Results

Hameln

#	Buildings	Underground working	Loose material	grass	reed	crop	Deciduous trees	Coniferous trees	grove	shrubbery	Running water	Standing water
SSL1	**63.00**	**38.07**	0.00	**10.58**	**19.20**	**30.00**	**33.04**	**3.01**	0.00	0.00	0.00	0.00
SSL2	**73.00**	**39.00**	0.00	**15.00**	**25.6**	**37.22**	**35.80**	**6.80**	0.00	0.00	0.00	0.00
SUP	**72.00**	**40.00**	0.00	**13.53**	**20.4**	**43.50**	**35.80**	**9.50**	0.00	0.00	0.00	0.00

Baseline

#	Buildings	Pervious surface	Impervious surface	Bare soil	water	coniferous	deciduous	brush wood	vineyard	herbaceous vegetation	land	plowed land
SSL	53.00	32.00	52.00	21.00	65.00	34.00	60.00	17.00	40.00	36.00	43.0	32.0
SUP	77.80	45.90	68.00	33.80	82.20	29.50	64.7	21.4	56.1	41.00	49.2	37.60

Results

Dataset and method	mIoU	OA
Baseline SL	**50.63**	**66.80**
Hameln SL	**19.56**	**47.74**
Baseline SSL	**40.91**	**61.09**
Hameln SSL	**16.41**	**38.42**
Hameln SSL	**19.37**	**41.45**

Limitations

- Very few labeled images (2077) available.
- Manual quality check of produced annotations.
- Problem with batch size
    - Our batch size : 8
    - Reference paper minimum batch size 256

Conclusion

- Implemented multimodal SSL on existing baseline.
- Utilized the unlabelled data for model pre-training.
- Supervised learning has better mIoU and OA than SSL in our experiments.
- There is room for improvements.

Outlook

- Getting more training data with labels.
- Using method that does not rely on negative samples or require few negative samples.
- Fusion module can be improved (Attention based)
    - Squeeze-Excitation
    - Residual-Excite

References

- Garioud, Anatol, Apolline De Wit, Marc Poupée, Marion Valette, Sébastien Giordano, and Boris Wattrelos. "FLAIR: French Land cover from Aerospace ImageRy." arXiv, 2023. https://arxiv.org/abs/2305.14467.
- Ting Chen, Simon Kornblith etc, A Simple Framework for Contrastive Learning of Visual Representations https://arxiv.org/abs/2002.05709
- Linus Scheibenreif Jo¨elle Hanna Etc, Self-supervised Vision Transformers for Land-cover Segmentation and Classification https://ieeexplore.ieee.org/document/9857009
- Yongshuo Zong, Oisin Mac Aodha Etc, Self-Supervised Multimodal Learning: A Survey https://arxiv.org/pdf/2304.01008