Contrastive Self-supervised Data Fusion for Satellite Imagery

Artificial Intelligence
Machine Learning
Universität St.Gallen

Linus Scheibenreif, Michael Mommert, Damian Borth

AIML Lab, School of Computer Science, *firstname.lastname@unisg.ch*

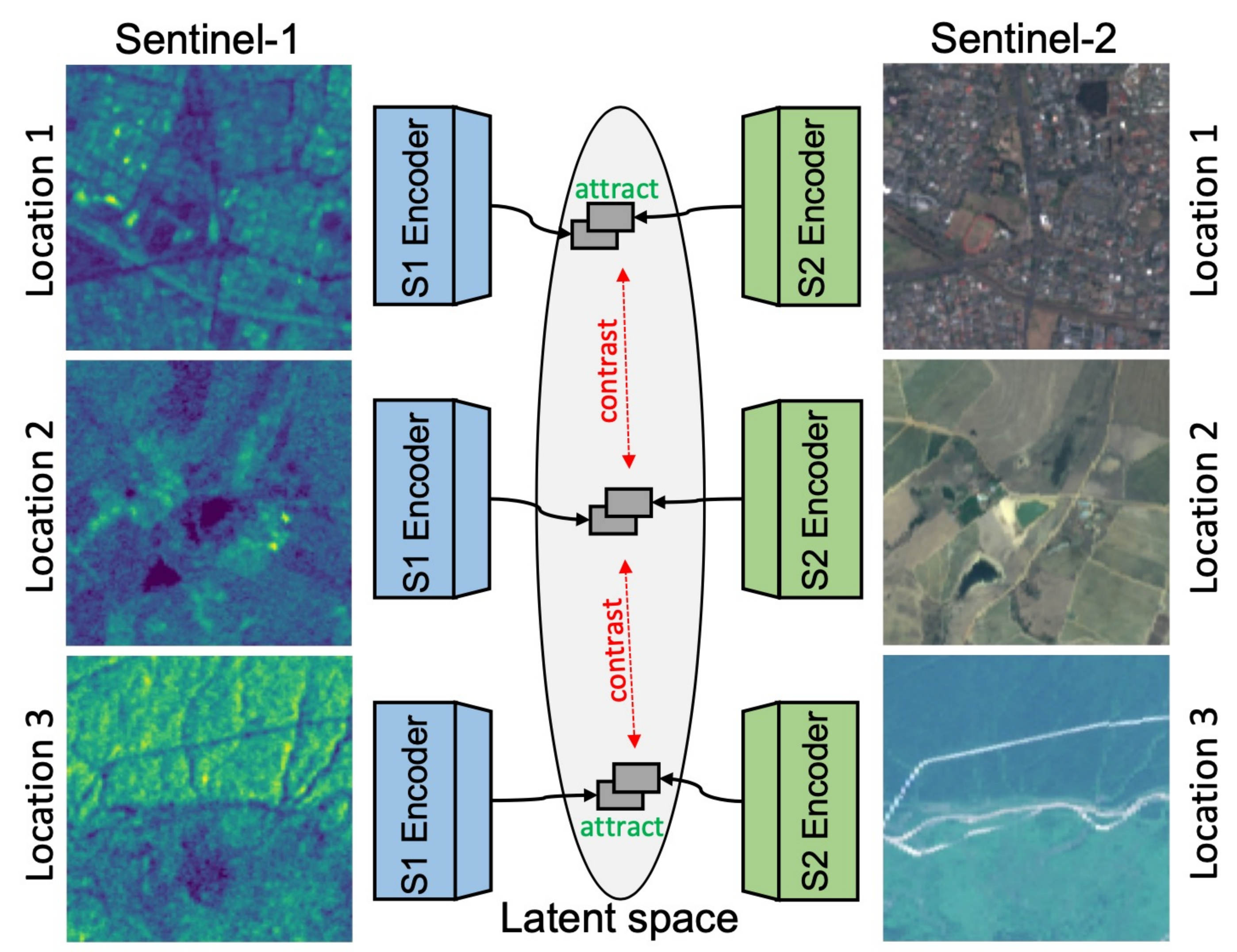
Motivation

- Unlabeled observations are abundant in the remote sensing domain
- Labeling this data is difficult, slow and expensive
- Realistically, it is not possible to label all data that satellites acquire
- This makes **self-supervised learning** a promising approach to leverage vast amounts of unlabeled data in the remote sensing domain

Idea

- We adapt the SimCLR [1] contrastive self-supervised learning approach to co-located Sentinel-1 and Sentinel-2 imagery
- Instead of randomly augmenting data samples, our technique treats the different imaging modalities as positive pairs of views for the same scene
- This enables self-supervised data fusion and yields strong performance on land-cover classification downstream tasks

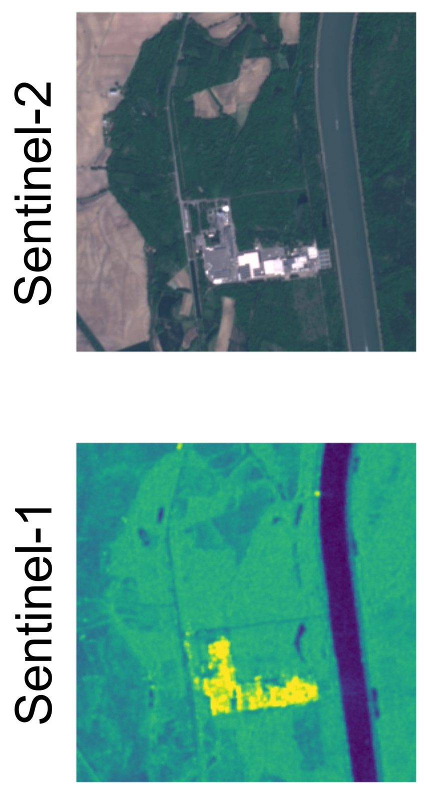
Overview



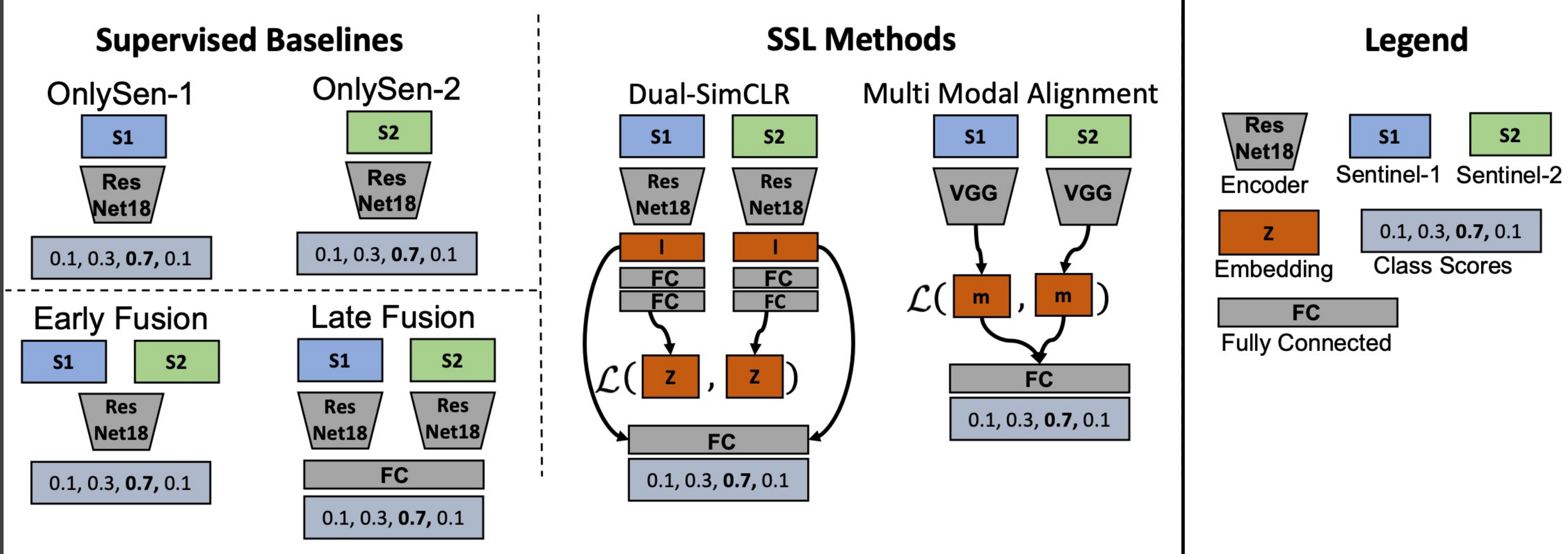
Dataset

Sentinel-2 is a multi-spectral ESA Earth observation mission. The Sentinel-2 satellites provide imagery with up to 10m resolution.
Sentinel-1 actively senses the Earth with synthetic aperture radar (SAR) at approximately 10m resolution.

We obtain co-located Sentinel-1/2 images from the **SEN12MS** [2] and **DFC2020** [3] datasets for self-supervised pre-training and downstream classification tasks, respectively.



Model Architectures



Evaluation on Downstream Tasks

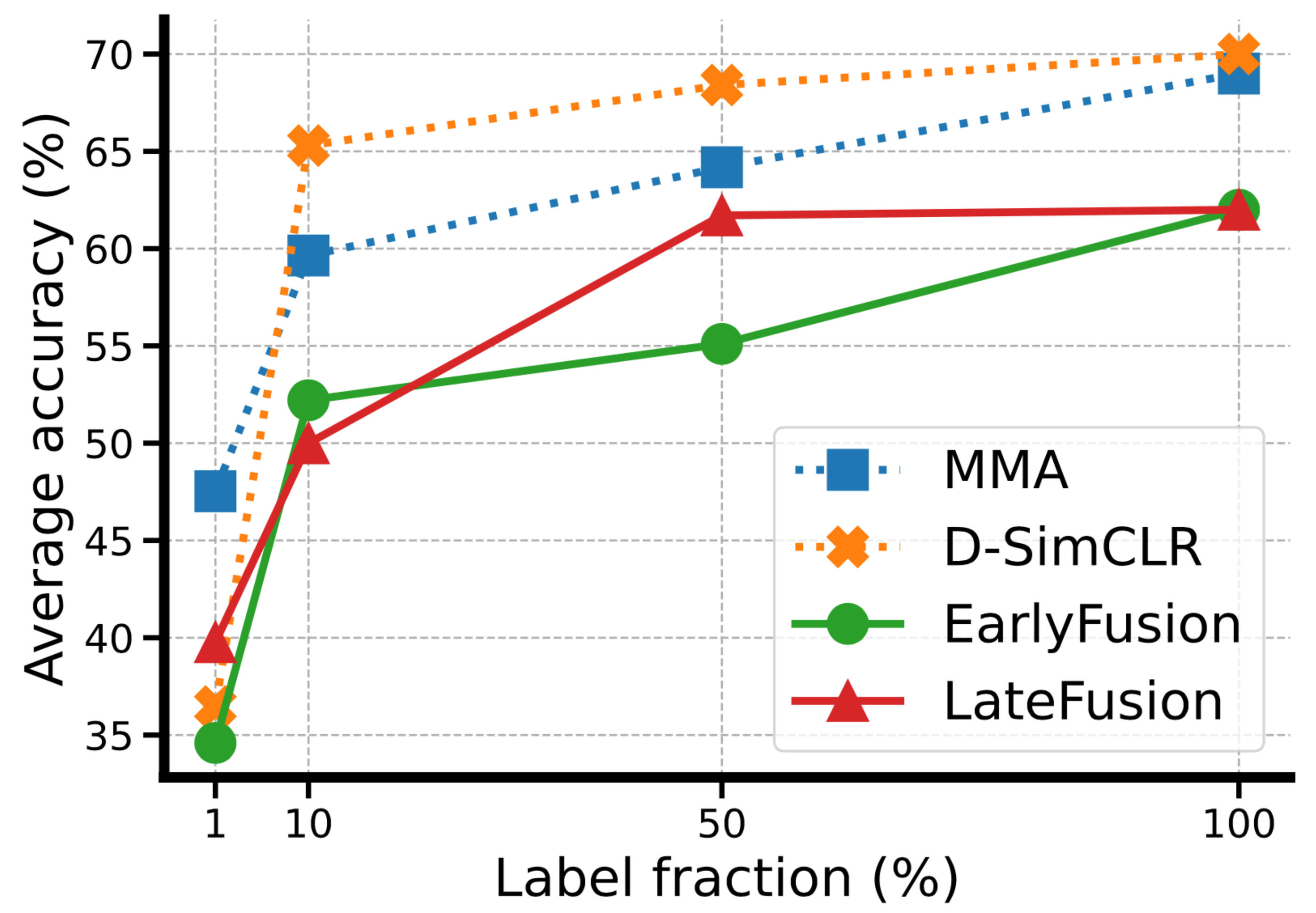
Single-label classification

Accuracy (%)	Forest	Shrubland	Grassl.	Wetl.	Cropl.	Urban	Barren	Water	**Average**	**OA**
OnlySen-1	80 ± 15	57 ± 2	18 ± 17	0 ± 0	75 ± 10	67 ± 9	58 ± 2	97 ± 2	57 ± 3	62 ± 1
OnlySen-2	43 ± 26	78 ± 12	45 ± 29	11 ± 6	59 ± 9	62 ± 5	61 ± 18	96 ± 6	57 ± 6	62 ± 5
EarlyFusion	60 ± 12	66 ± 37	62 ± 8	1 ± 1	66 ± 10	73 ± 6	66 ± 18	99 ± 0	62 ± 4	66 ± 2
LateFusion	62 ± 23	76 ± 14	51 ± 18	1 ± 2	64 ± 11	71 ± 5	75 ± 9	100 ± 1	62 ± 4	65 ± 3
SimCLR (RGB)	11 ± 12	69 ± 13	45 ± 14	3 ± 3	66 ± 22	26 ± 23	77 ± 14	99 ± 1	49 ± 3	58 ± 4
D-SimCLR	78 ± 11	84 ± 6	62 ± 10	10 ± 6	63 ± 3	84 ± 4	82 ± 7	99 ± 0	$\mathbf{70 \pm 2}$	$\mathbf{70 \pm 1}$
MMA	68 ± 17	89 ± 5	53 ± 13	8 ± 9	71 ± 7	80 ± 6	81 ± 7	100 ± 0	69 ± 2	69 ± 1

Multi-label classification

F1 Score (%)	Forest	Shrubland	Grassl.	Wetl.	Cropl.	Urban	Barren	Water	**Average**	**O-F1**
OnlySen-1	69 ± 2	46 ± 6	29 ± 5	8 ± 8	68 ± 7	81 ± 3	60 ± 8	96 ± 1	57 ± 2	62 ± 2
OnlySen-2	37 ± 20	51 ± 14	43 ± 20	23 ± 18	76 ± 2	79 ± 6	63 ± 10	94 ± 2	58 ± 3	63 ± 2
EarlyFusion	48 ± 10	53 ± 7	45 ± 13	13 ± 11	69 ± 5	84 ± 4	71 ± 4	94 ± 1	60 ± 3	62 ± 3
LateFusion	56 ± 6	45 ± 11	33 ± 9	18 ± 24	64 ± 3	69 ± 16	53 ± 15	96 ± 1	54 ± 7	61 ± 5
SimCLR (RGB)	3 ± 4	49 ± 11	24 ± 16	10 ± 8	63 ± 24	40 ± 36	49 ± 15	73 ± 6	39 ± 10	49 ± 6
D-SimCLR	62 ± 2	61 ± 3	53 ± 7	31 ± 2	72 ± 3	87 ± 0	77 ± 1	96 ± 1	$\mathbf{67 \pm 1}$	$\mathbf{69 \pm 1}$
MMA	58 ± 5	57 ± 5	35 ± 8	10 ± 6	77 ± 3	89 ± 1	73 ± 5	97 ± 0	62 ± 2	66 ± 1

Effect of Label Fraction



Outperforms supervised training with 10% of labeled observations

Summary

- Self-supervised Sentinel-1/2 data fusion
- Strong performance on land-cover classification downstream tasks
- Outperforms supervised training with 10% of labeled samples

References

[1] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning*(pp. 1597-1607). PMLR.
[2] Schmitt, M., Hughes, L. H., Qiu, C., & Zhu, X. X. (2019). SEN12MS–A CURATED DATASET OF GEOREFERENCED MULTI-SPECTRAL SENTINEL-1/2 IMAGERY FOR DEEP LEARNING AND DATA FUSION. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, 153-160.
[3] Yokoya, N., Ghamisi, P., Hansch, R., & Schmitt, M. (2020). Report on the 2020 IEEE GRSS data fusion contest-global land cover mapping with weak supervision [technical committees]. *IEEE Geoscience and Remote Sensing Magazine*, *8*(4), 134-137.