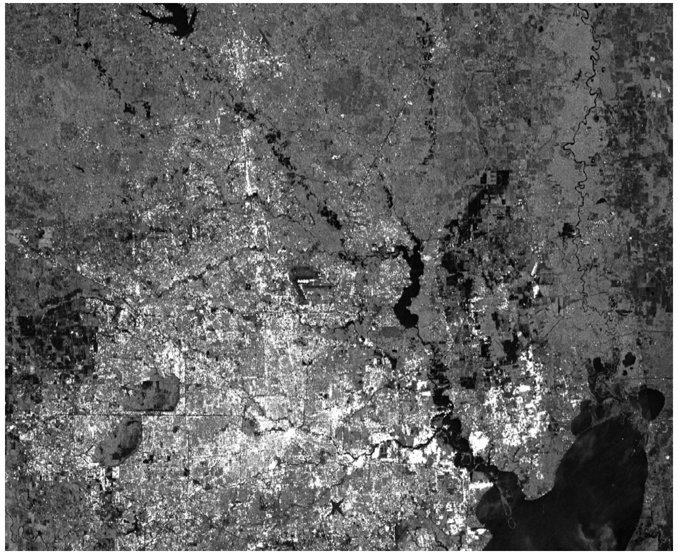
Efficient Learning for Earth Observation Applications with Self-Supervised Learning

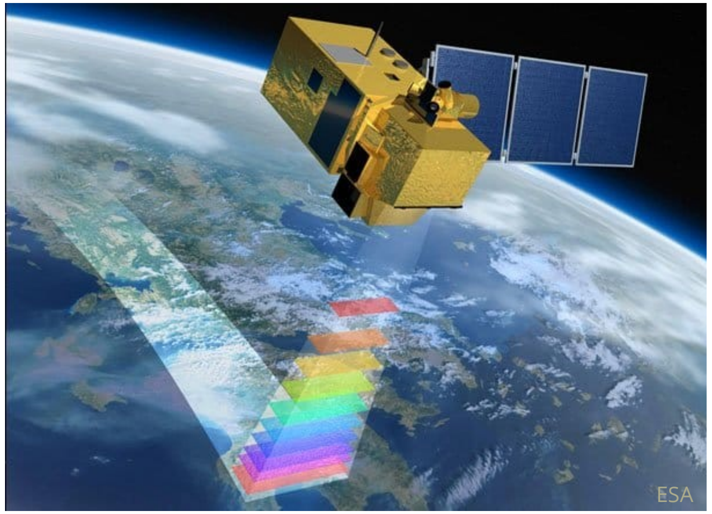
Michael Mommert, Linus Scheibenreif, Joëlle Hanna, Damian Borth
University of St. Gallen

Deep Learning for Earth observation

Deep Learning for Earth observation

Earth observation data are highly complex (unstructured, multi-modal).

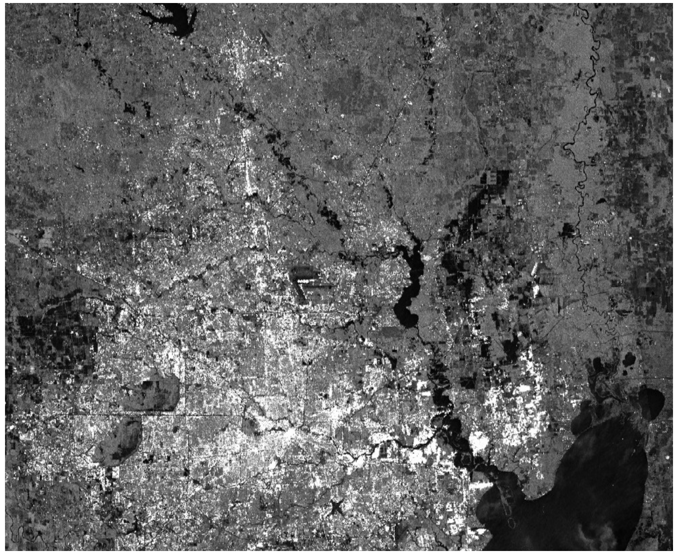


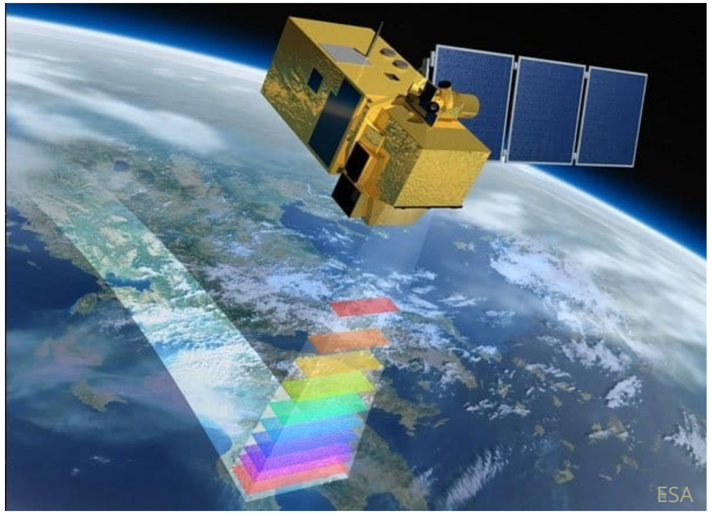


Deep Learning for Earth observation

Earth observation data are highly complex (unstructured, multi-modal).

How can we analyze these vast amounts of data?



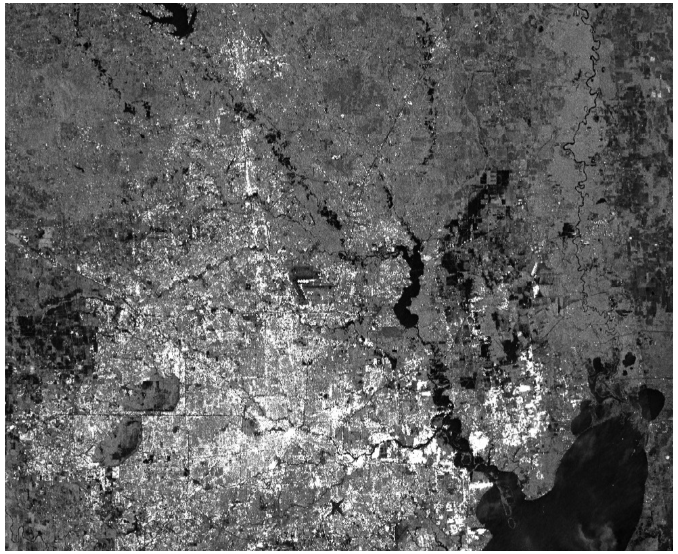
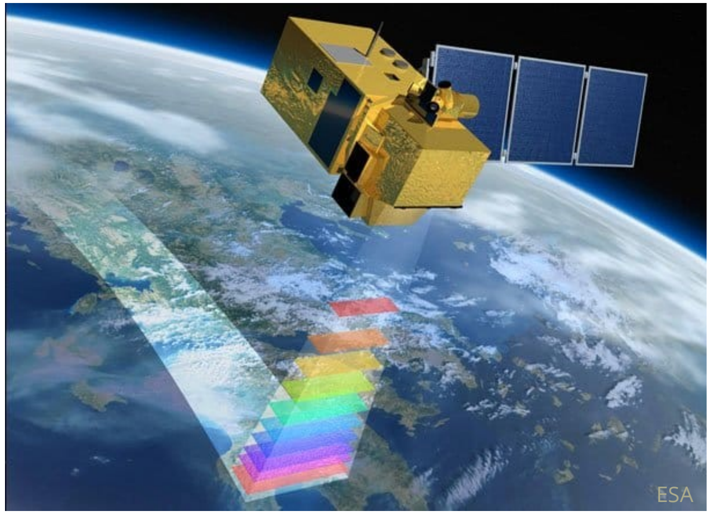


Deep Learning for Earth observation

Earth observation data are highly complex (unstructured, multi-modal).

How can we analyze these vast amounts of data?

Deep Learning offers the **scalability** to analyze large amounts of data.

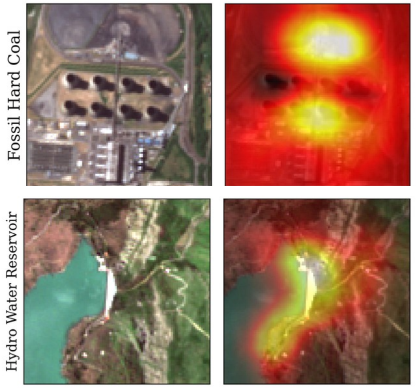

Deep Learning for Earth observation

Earth observation data are highly complex (unstructured, multi-modal).

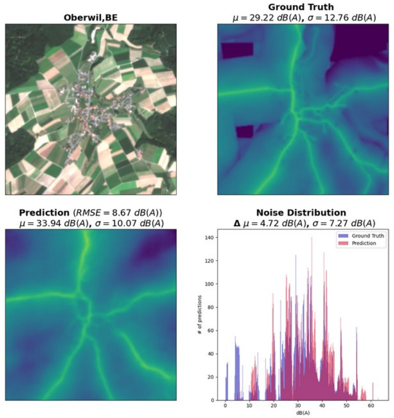
How can we analyze these vast amounts of data?

Deep Learning offers the **scalability** to analyze large amounts of data.

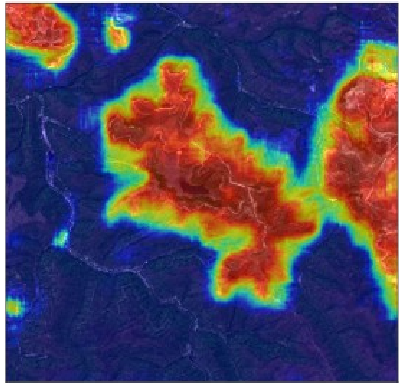
Deep Learning also offers the **flexibility** to deal with a range of different tasks.



Classification



Regression



Segmentation

Object Detection

Deep Learning for Earth observation

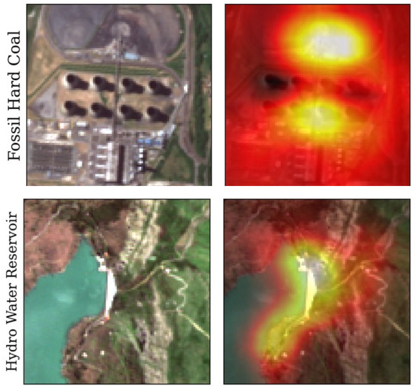
Earth observation data are highly complex (unstructured, multi-modal).

How can we analyze these vast amounts of data?

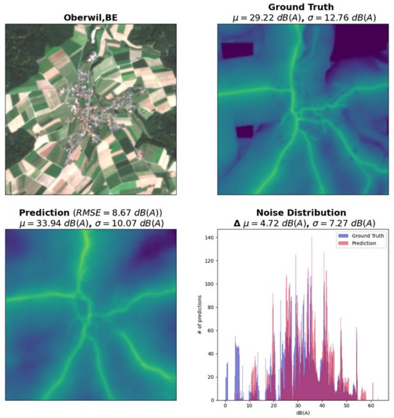
Deep Learning offers the **scalability** to analyze large amounts of data.

Deep Learning also offers the **flexibility** to deal with a range of different tasks.

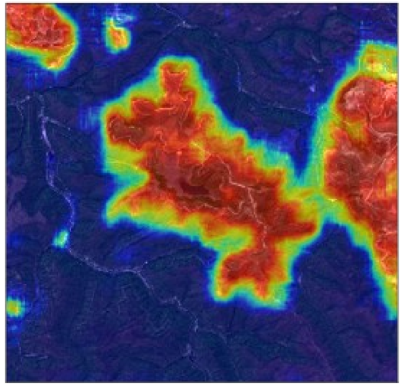
How does it work?



Classification



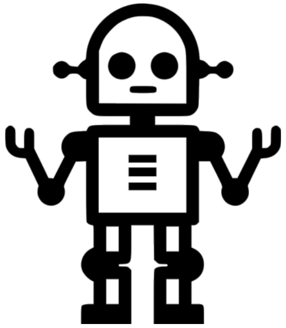
Regression



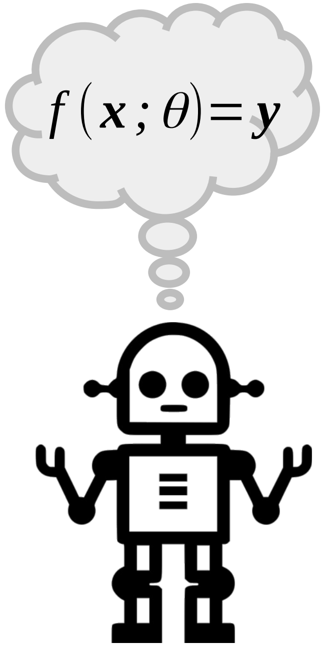
Segmentation

Object Detection

Supervised learning with Neural Networks



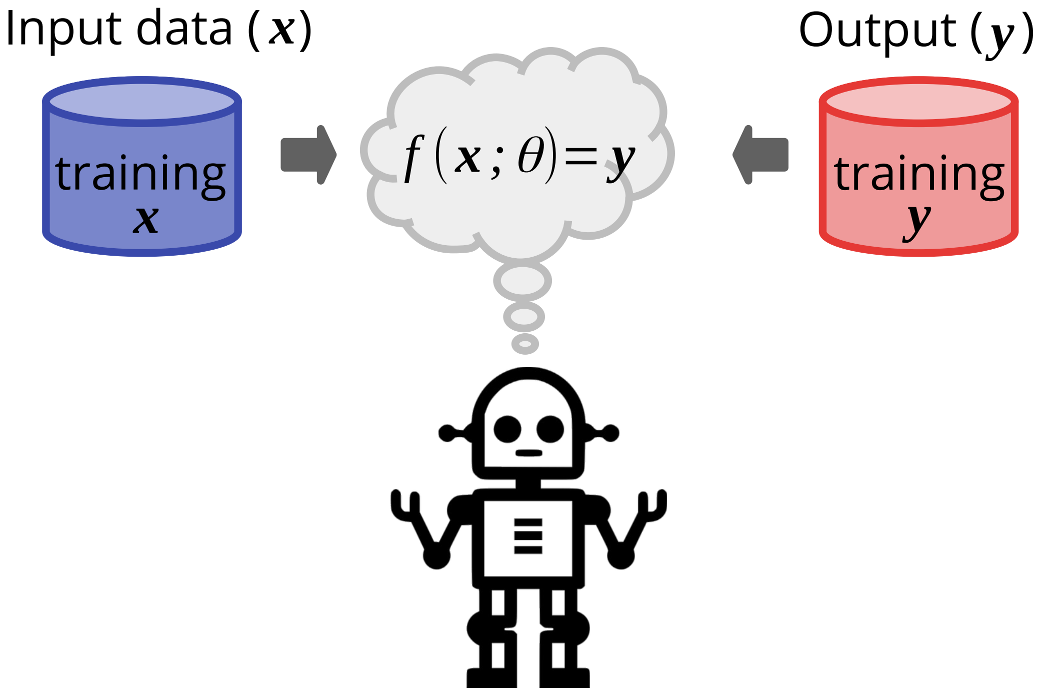
Supervised learning with Neural Networks



A machine learns a task from **annotated examples**.

Mathematically, it learns a function, f, that maps input data, $\boldsymbol{x}$, to the output, $\boldsymbol{y}$.

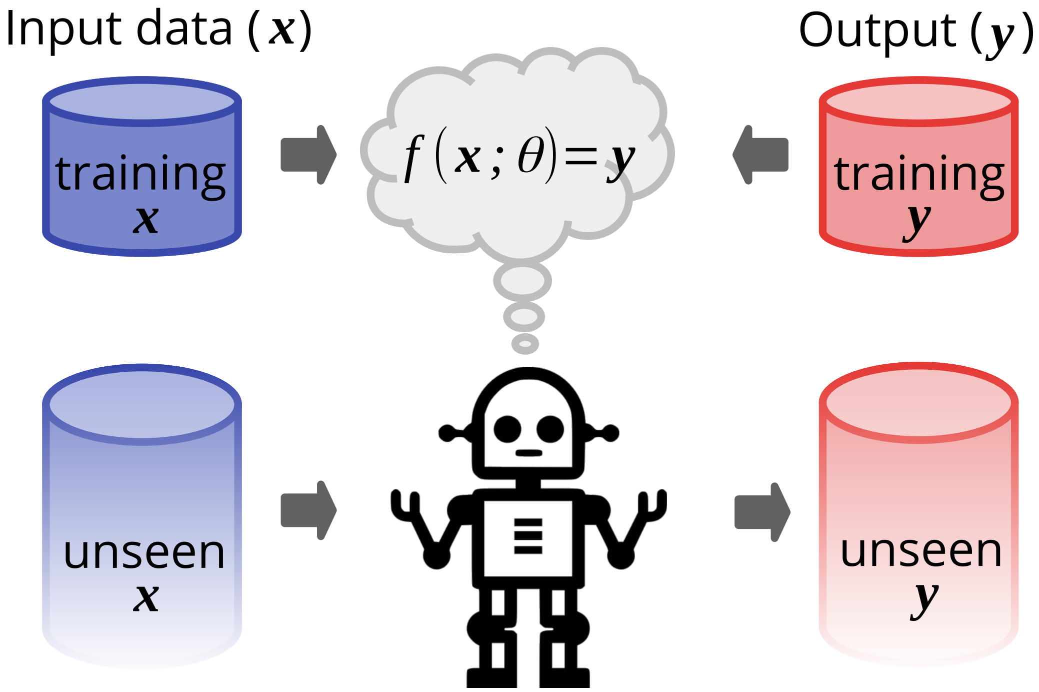
Supervised learning with Neural Networks



A machine learns a task from **annotated examples**.

Mathematically, it learns a function, f, that maps input data, $\boldsymbol{x}$, to the output, $\boldsymbol{y}$.

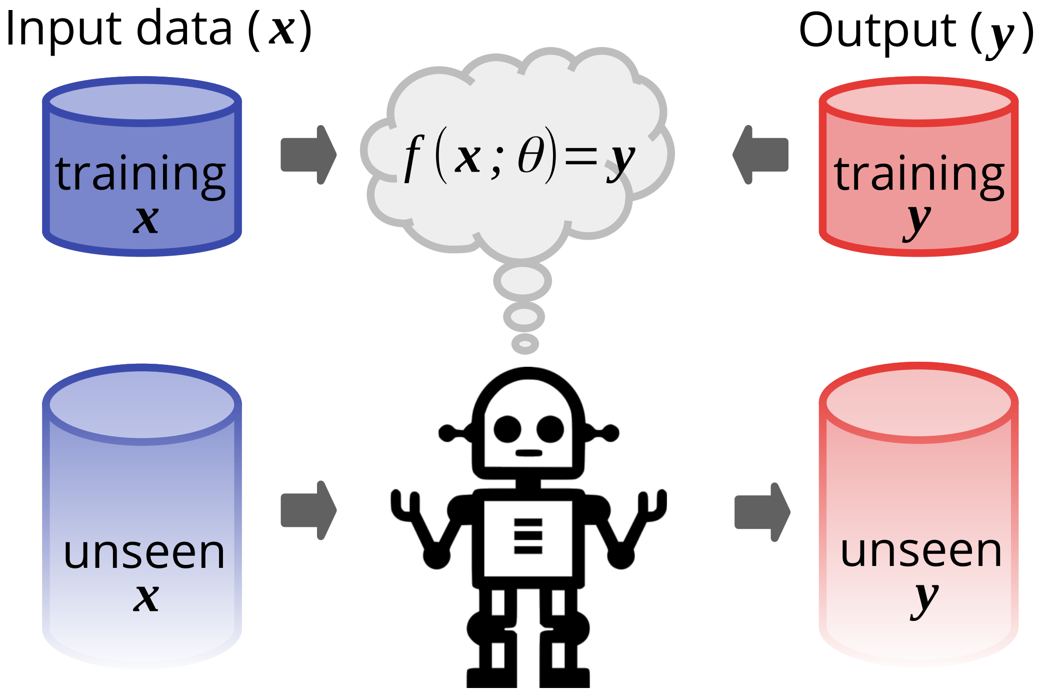
Supervised learning with Neural Networks



A machine learns a task from **annotated examples**.

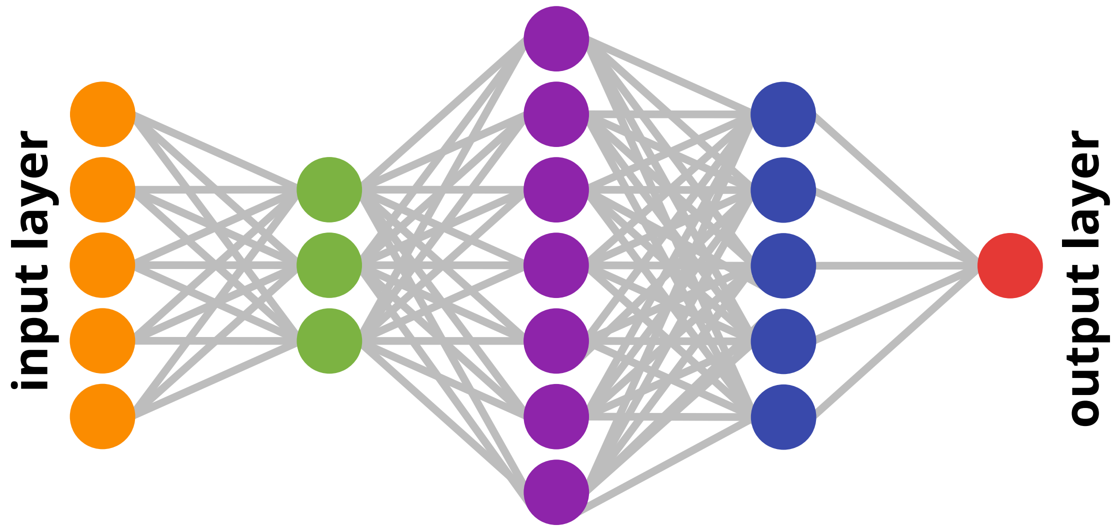
Mathematically, it learns a function, f, that maps input data, x, to the output, y.

Supervised learning with Neural Networks



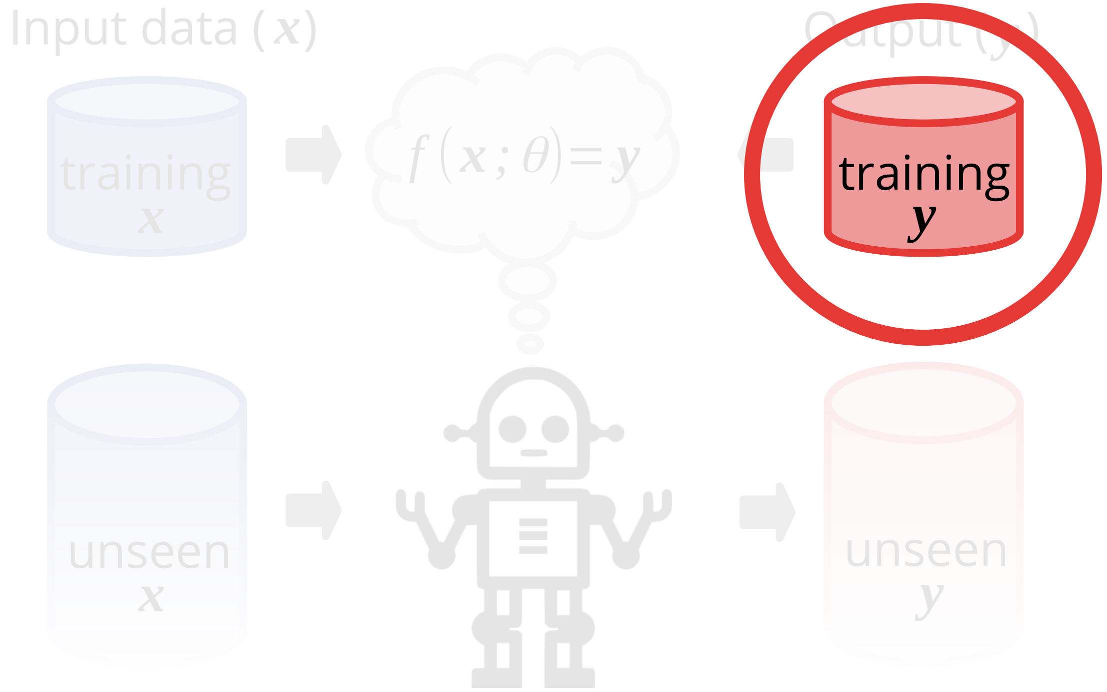
A machine learns a task from **annotated examples**.

Mathematically, it learns a function, f, that maps input data, x, to the output, y.



A Neural Network is a cascade of mathematical functions; each neuron contains learnable weights that represent the learned knowledge.

Supervised learning with Neural Networks



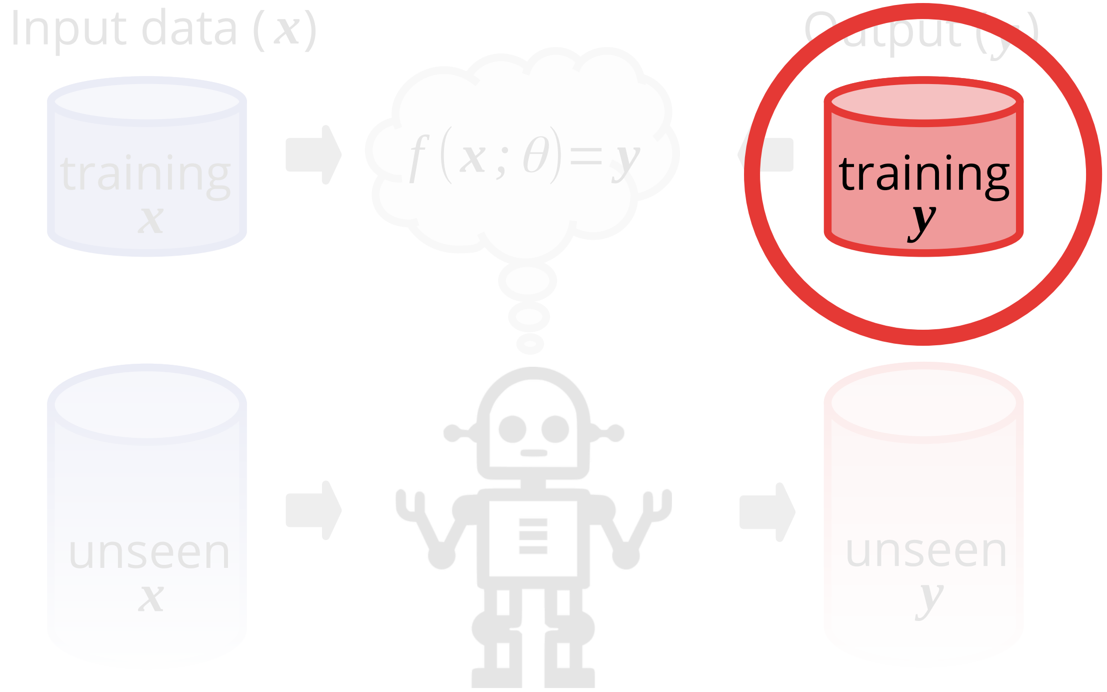
The availability of annotations typically represents the most important **bottleneck** in supervised learning.

A machine learns a task from **annotated examples**.

Mathematically, it learns a function, f, that maps input data, x, to the output, y.

A Neural Network is a cascade of mathematical functions; each neuron contains learnable weights that represent the learned knowledge.

Supervised learning with Neural Networks



The availability of annotations typically represents the most important **bottleneck** in supervised learning.

Can we force the model to use the available annotations more **efficiently**?

Can we take advantage of the vast amounts of **unannotated data**?

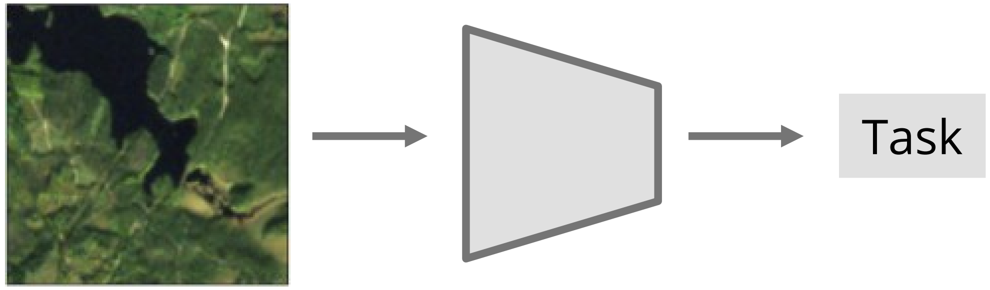
A machine learns a task from **annotated examples**.

A Neural Network is a cascade of mathematical functions; each neuron contains learnable weights that represent the learned knowledge.

Mathematically, it learns a function, f, that maps input data, x, to the output, y.

How can we use annotated data more efficiently?

- Data augmentations

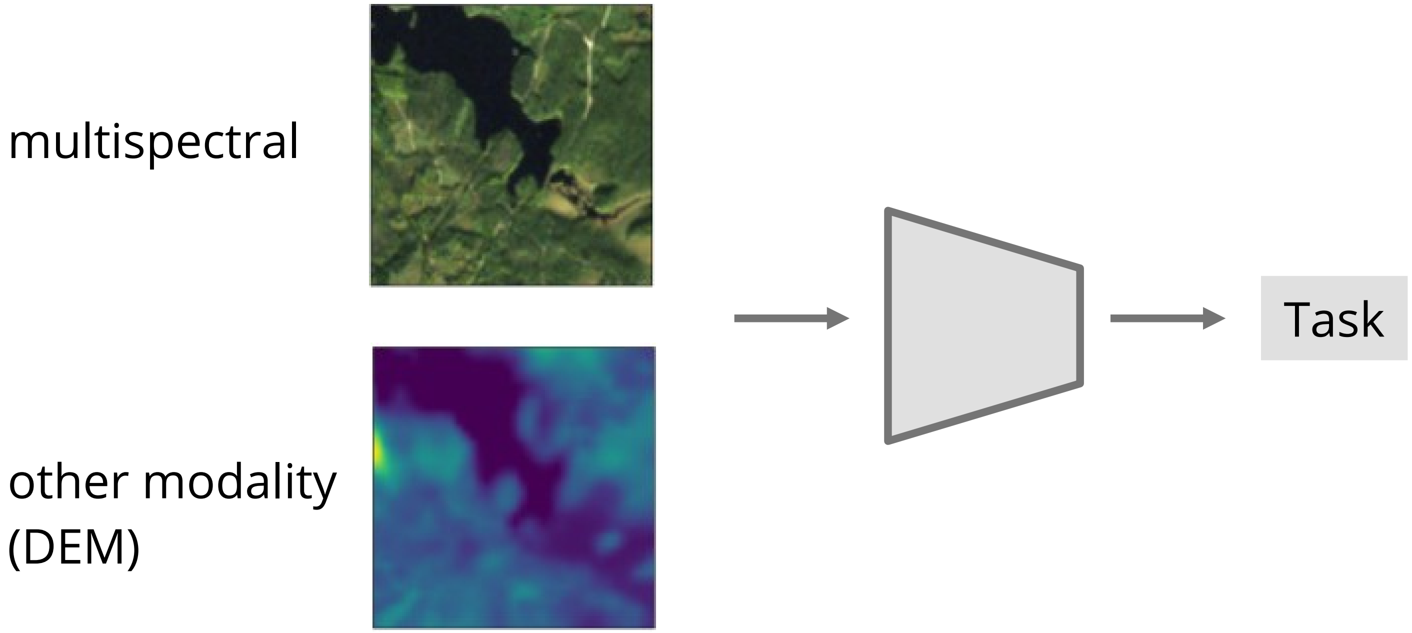


How can we use annotated data more efficiently?

- Data augmentations

How can we use annotated data more efficiently?

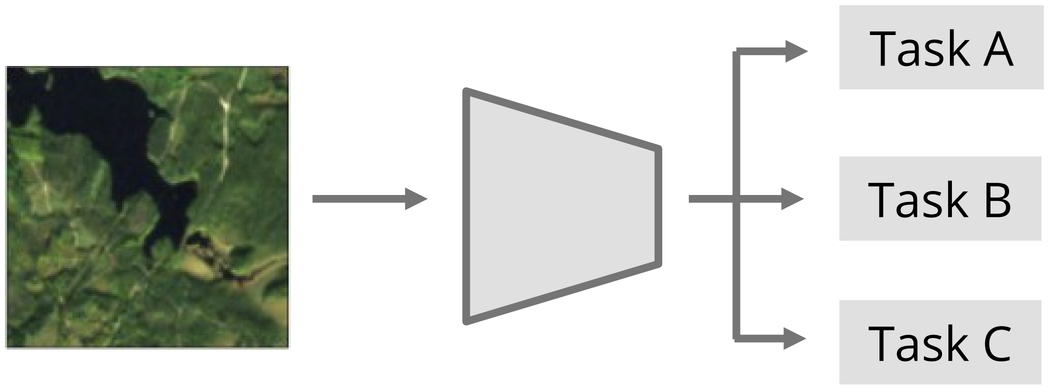
- Data augmentations
- Data Fusion



Scheibenreif, L., Mommert, M., Borth, D., "Towards Global Estimation of Ground-Level NO2 Pollution with Deep Learning and Remote Sensing", IEEE Transactions on Geoscience and Remote Sensing, 2022.

How can we use annotated data more efficiently?

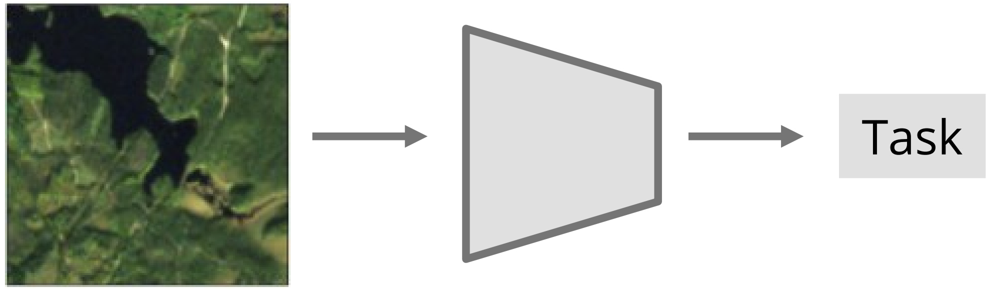
- Data augmentations
- Data Fusion
- Multi-task Learning



Hanna, J., Mommert, M., Scheibenreif, L. Borth, D., “Multitask Learning for Estimating Power Plant Greenhouse Gas Emissions from Satellite Imagery”, NeurIPS 2021 Tackling Climate Change with Machine Learning Workshop.

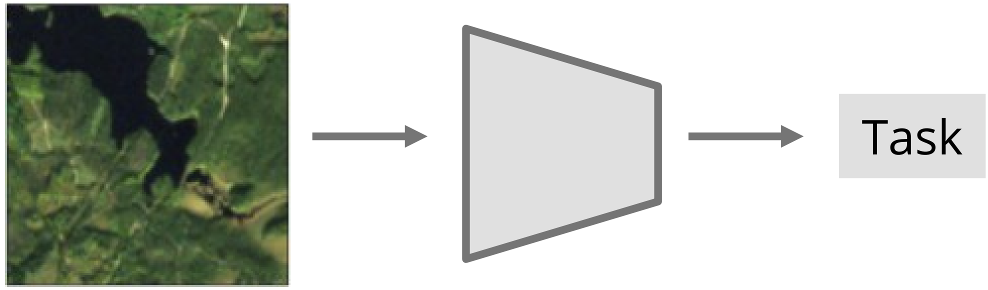
How can we use annotated data more efficiently?

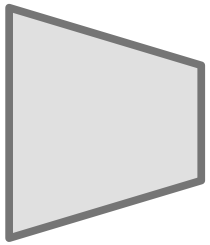
- Data augmentations
- Data Fusion
- Multi-task Learning
- Transfer Learning



How can we use annotated data more efficiently?

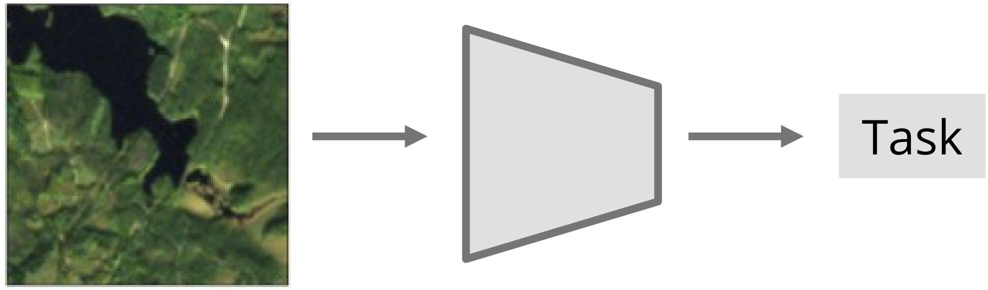
- Data augmentations
- Data Fusion
- Multi-task Learning
- Transfer Learning



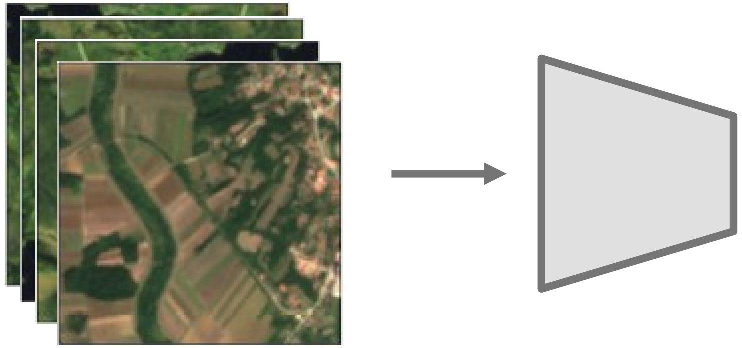


How can we use annotated data more efficiently?

- Data augmentations
- Data Fusion
- Multi-task Learning
- Transfer Learning

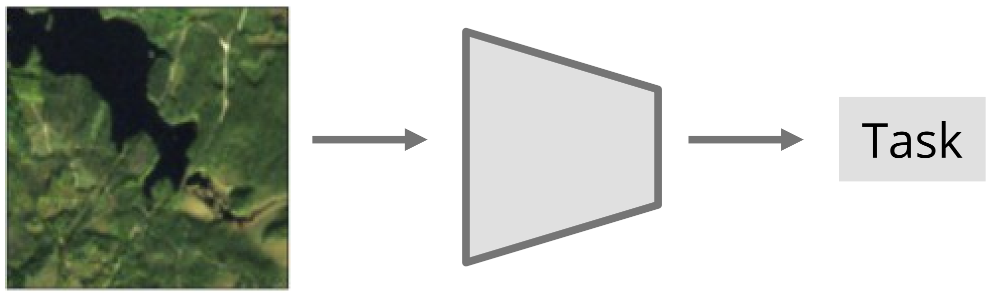


other dataset

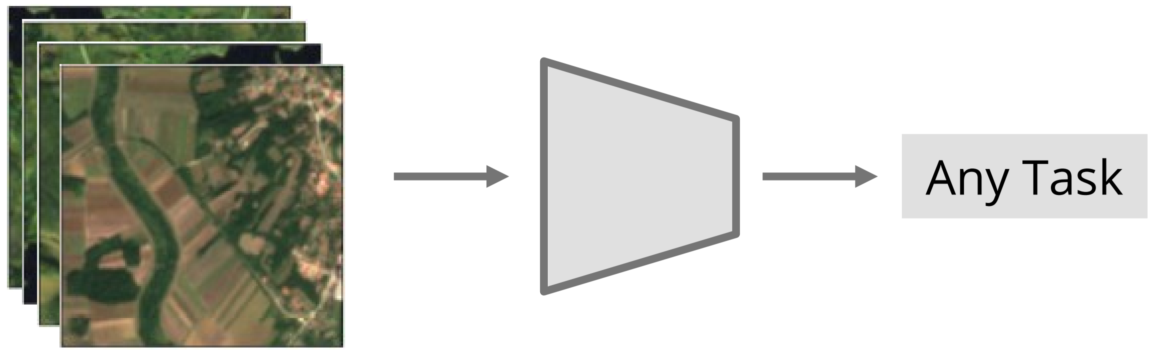


How can we use annotated data more efficiently?

- Data augmentations
- Data Fusion
- Multi-task Learning
- Transfer Learning

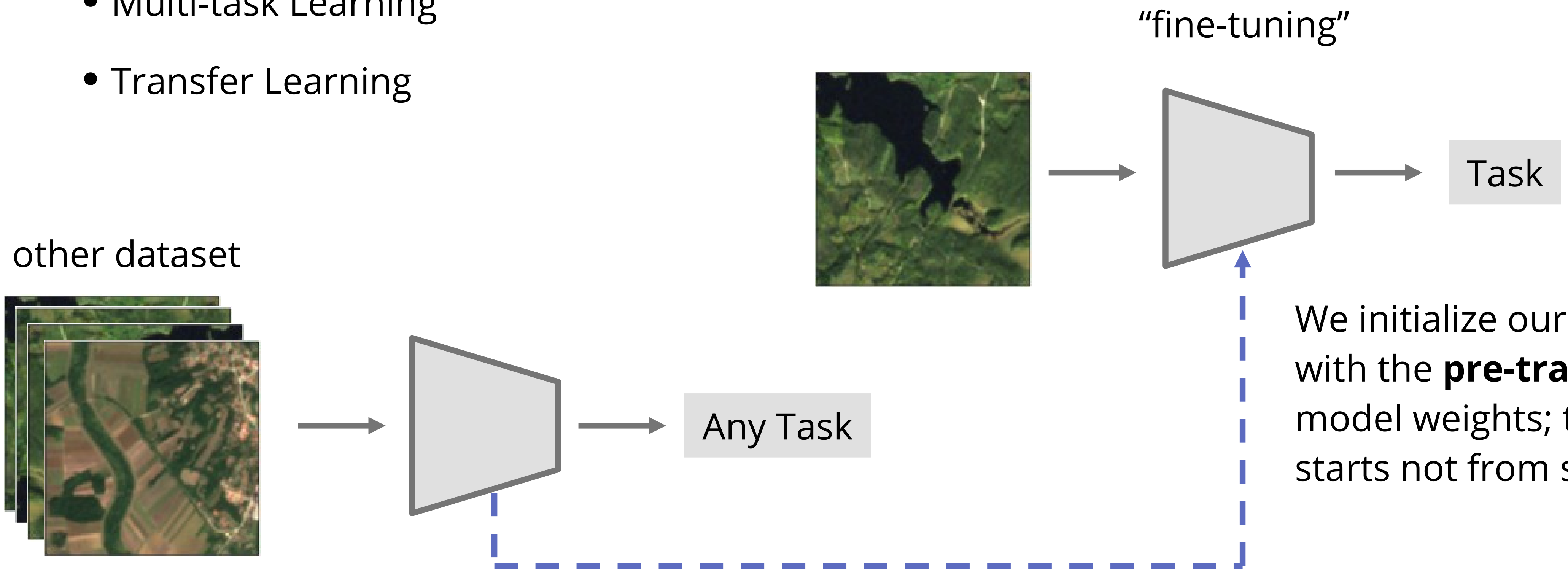


other dataset



How can we use annotated data more efficiently?

- Data augmentations
- Data Fusion
- Multi-task Learning
- Transfer Learning

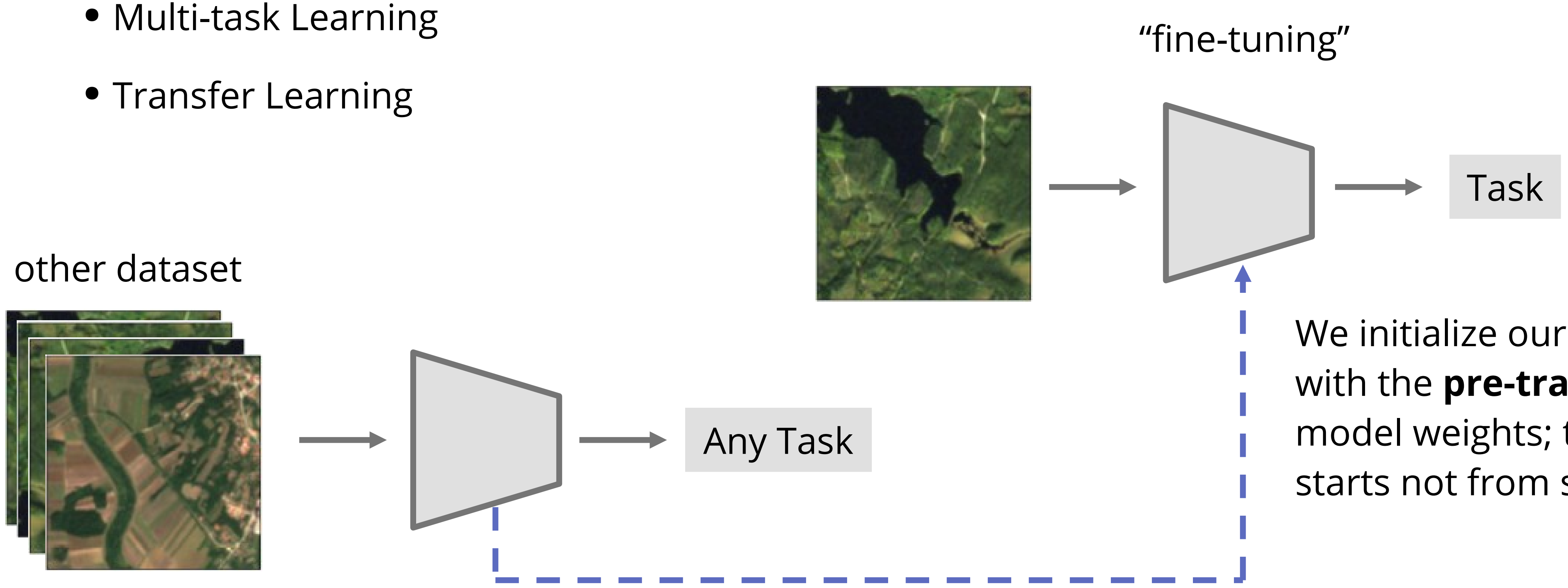


We initialize our model with the **pre-trained** model weights; training starts not from scratch!

How can we use annotated data more efficiently?

- Data augmentations
- Data Fusion
- Multi-task Learning
- Transfer Learning

Can we pretrain a model from unannotated data?

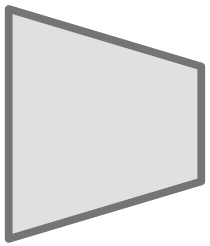


We initialize our model with the **pre-trained** model weights; training starts not from scratch!

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

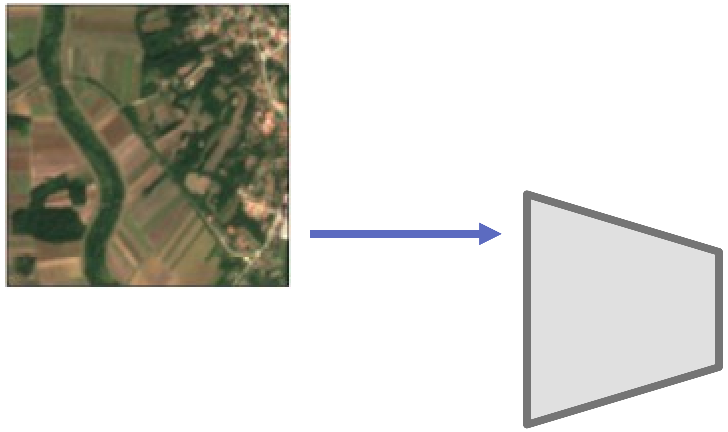


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. “A Simple Framework for Contrastive Learning of Visual Representations.” ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

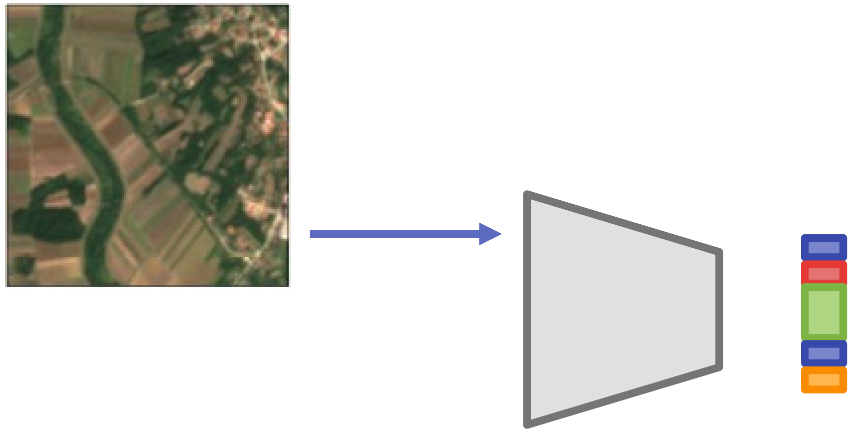


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. "A Simple Framework for Contrastive Learning of Visual Representations." ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

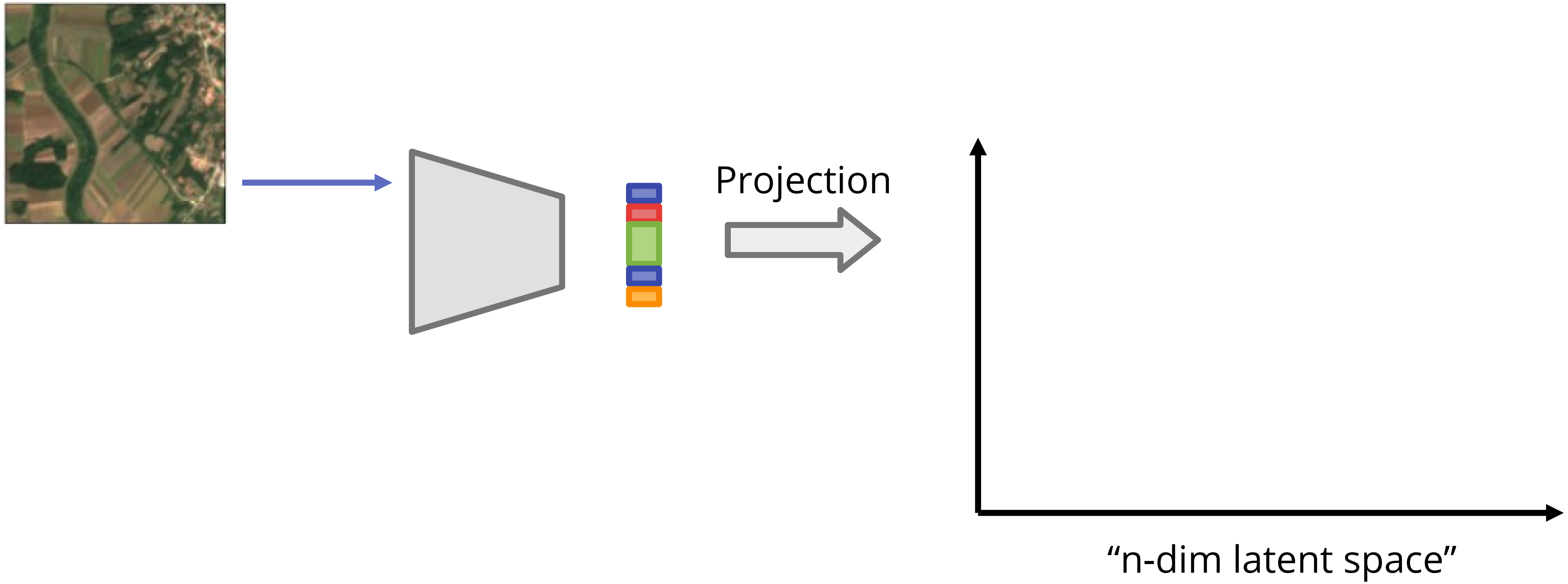


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. "A Simple Framework for Contrastive Learning of Visual Representations." ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

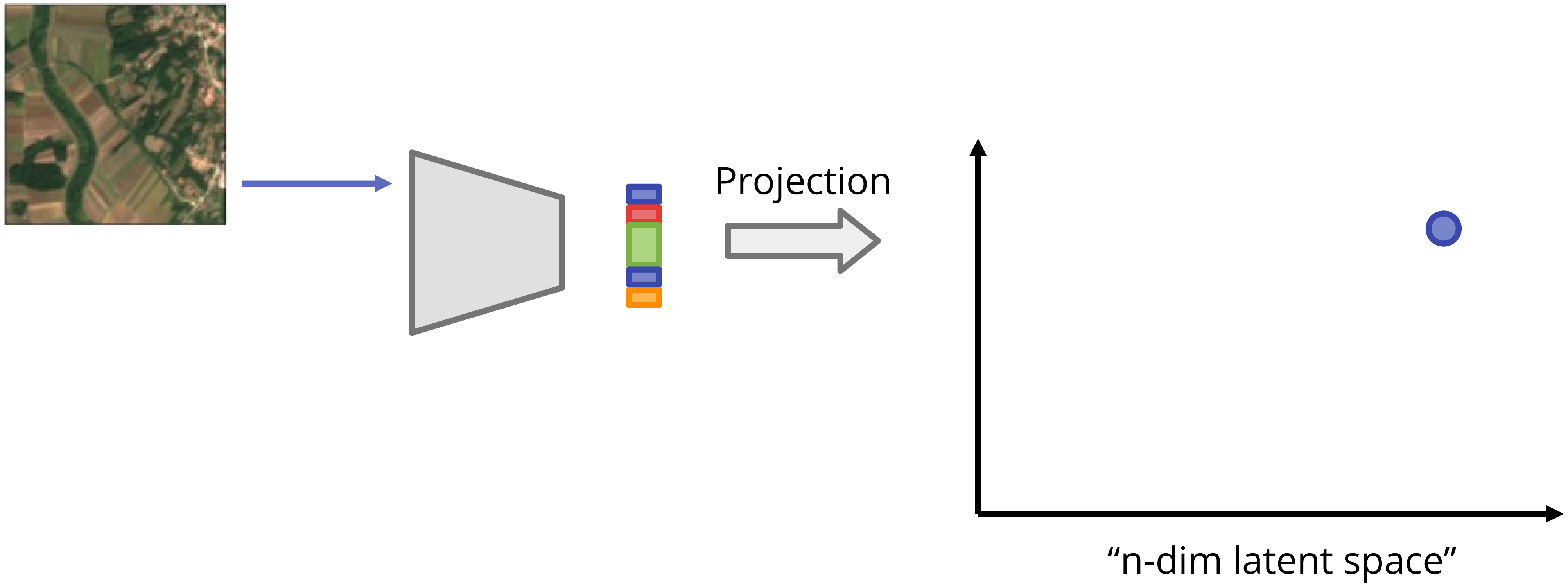


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. "A Simple Framework for Contrastive Learning of Visual Representations." ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

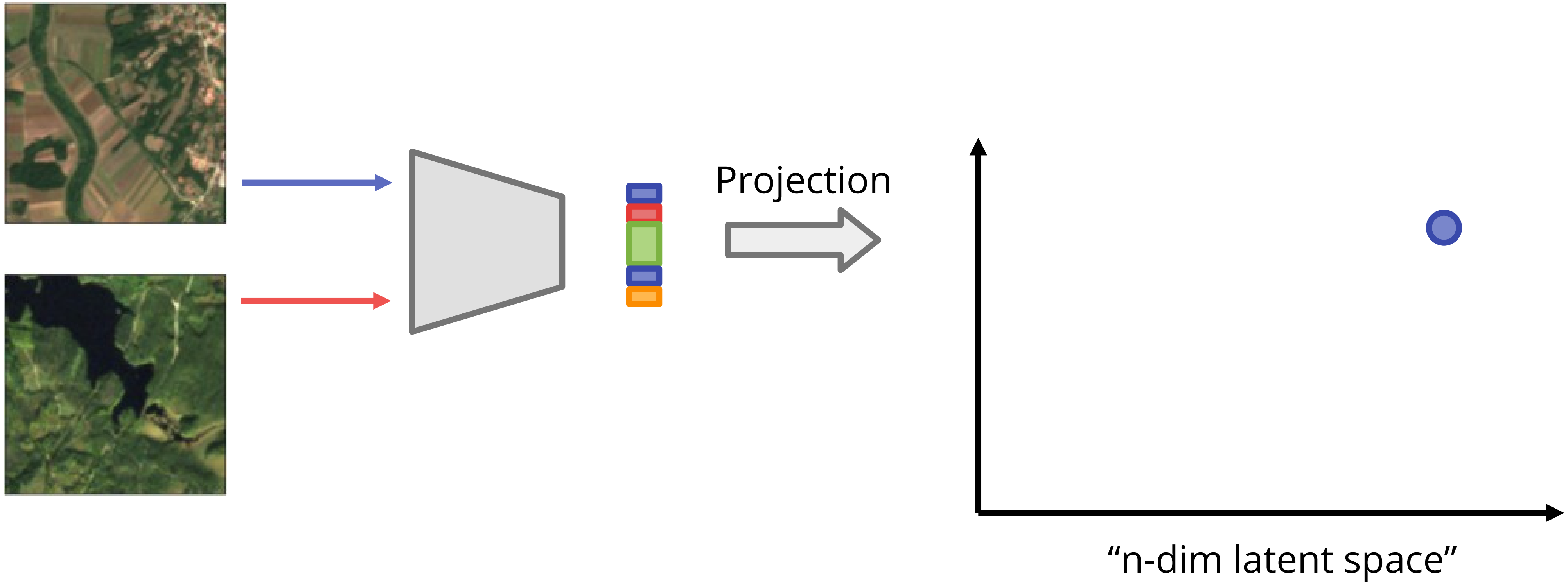


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. “A Simple Framework for Contrastive Learning of Visual Representations.” ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

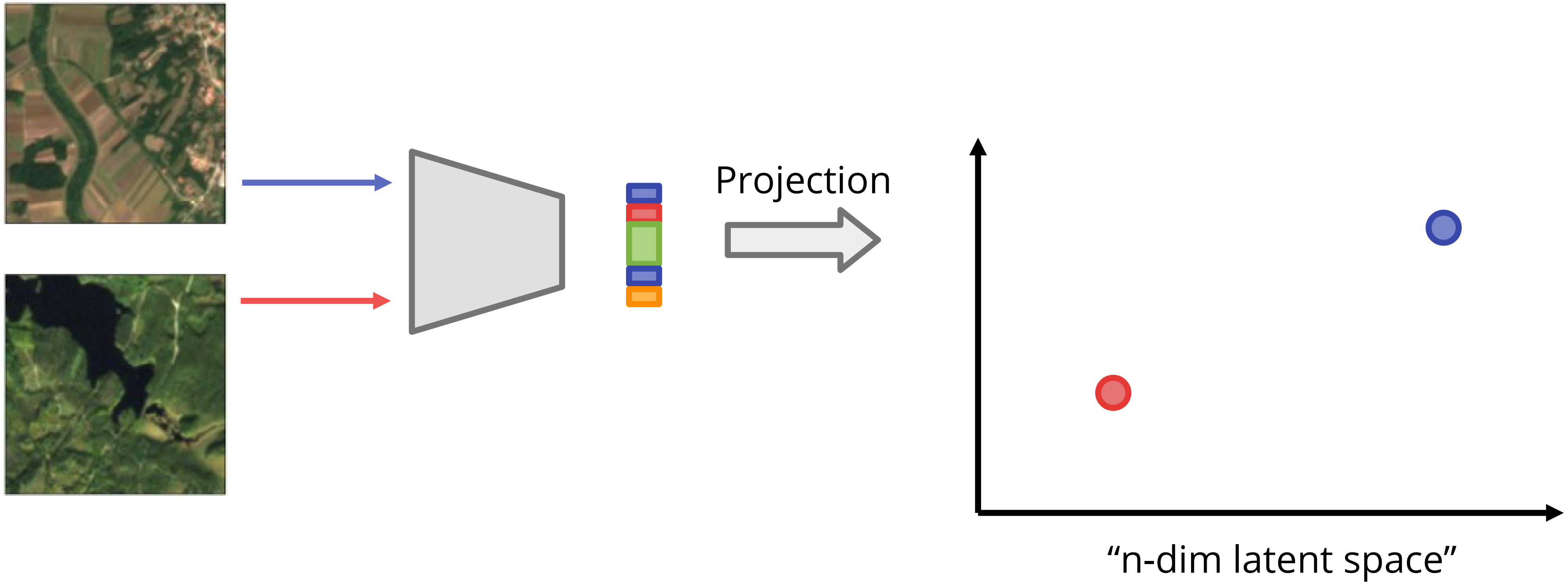


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. “A Simple Framework for Contrastive Learning of Visual Representations.” ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

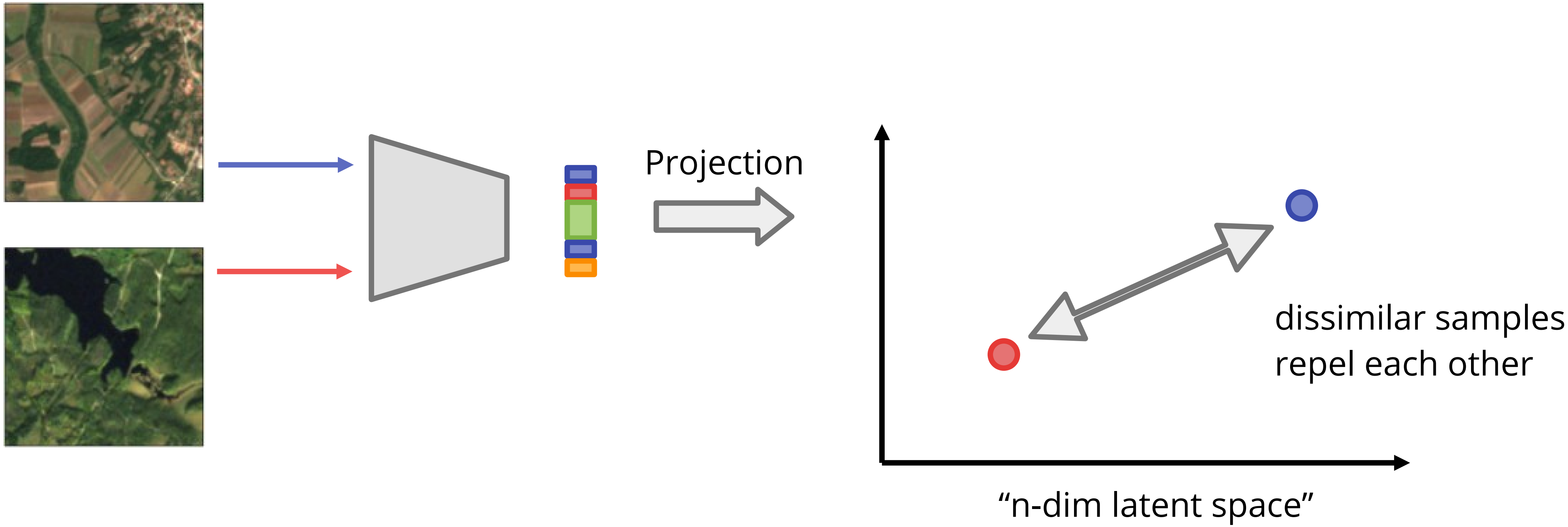


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. “A Simple Framework for Contrastive Learning of Visual Representations.” ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

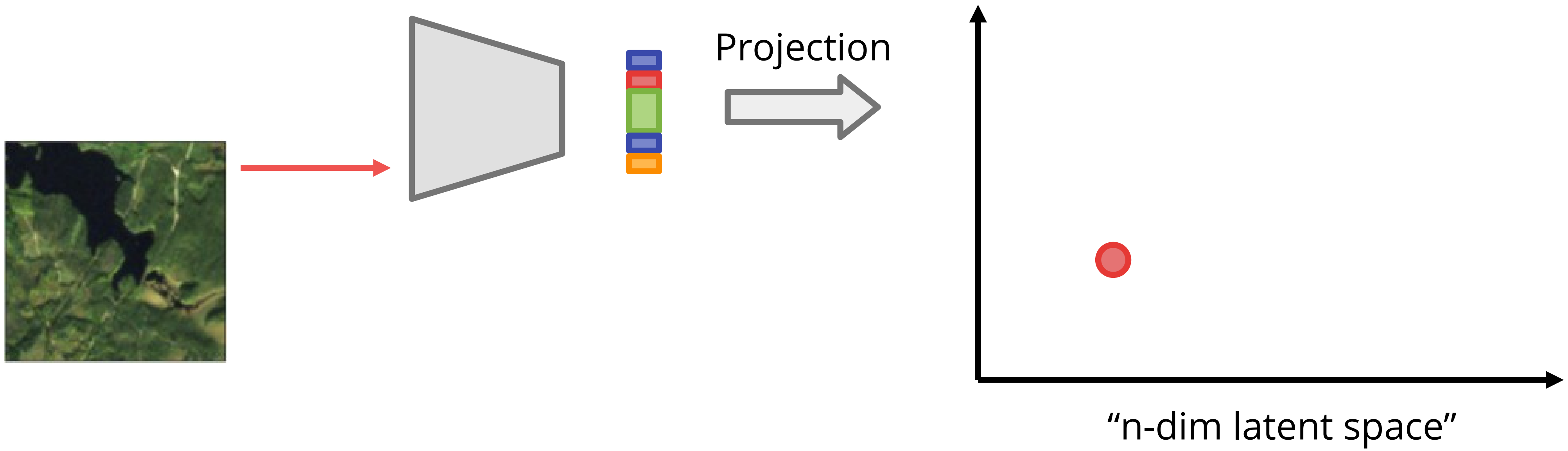


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. “A Simple Framework for Contrastive Learning of Visual Representations.” ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

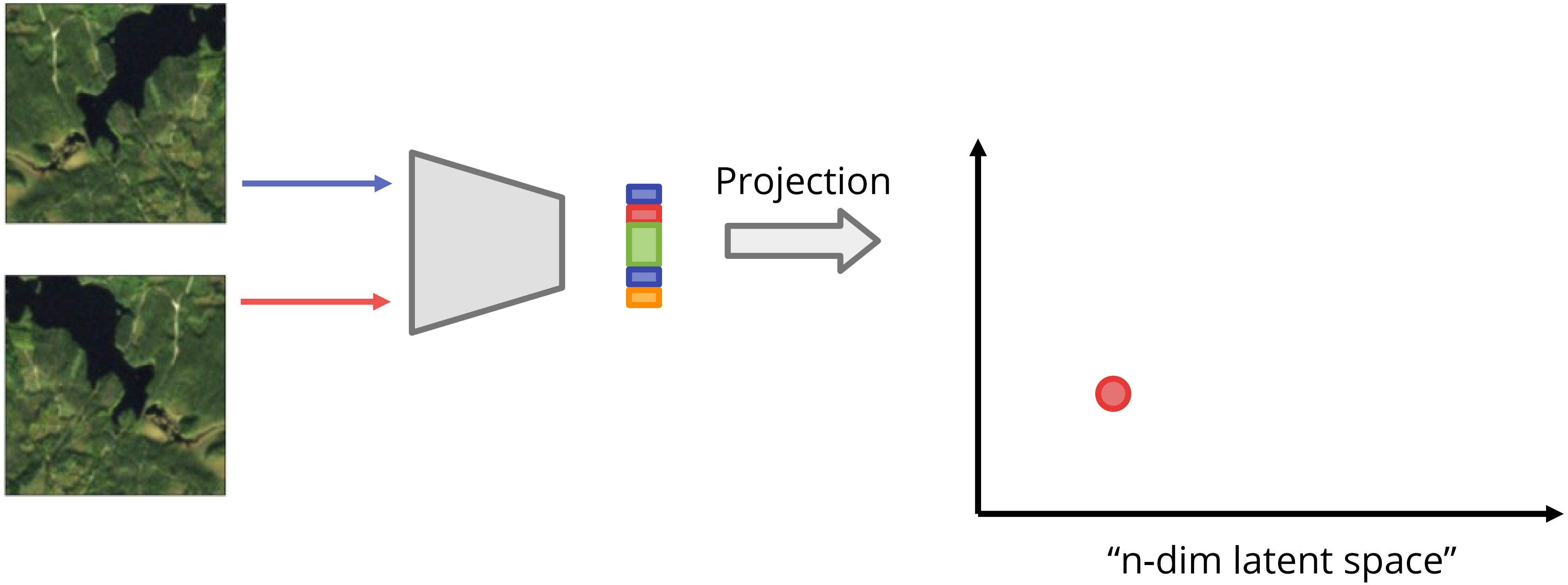


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. “A Simple Framework for Contrastive Learning of Visual Representations.” ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

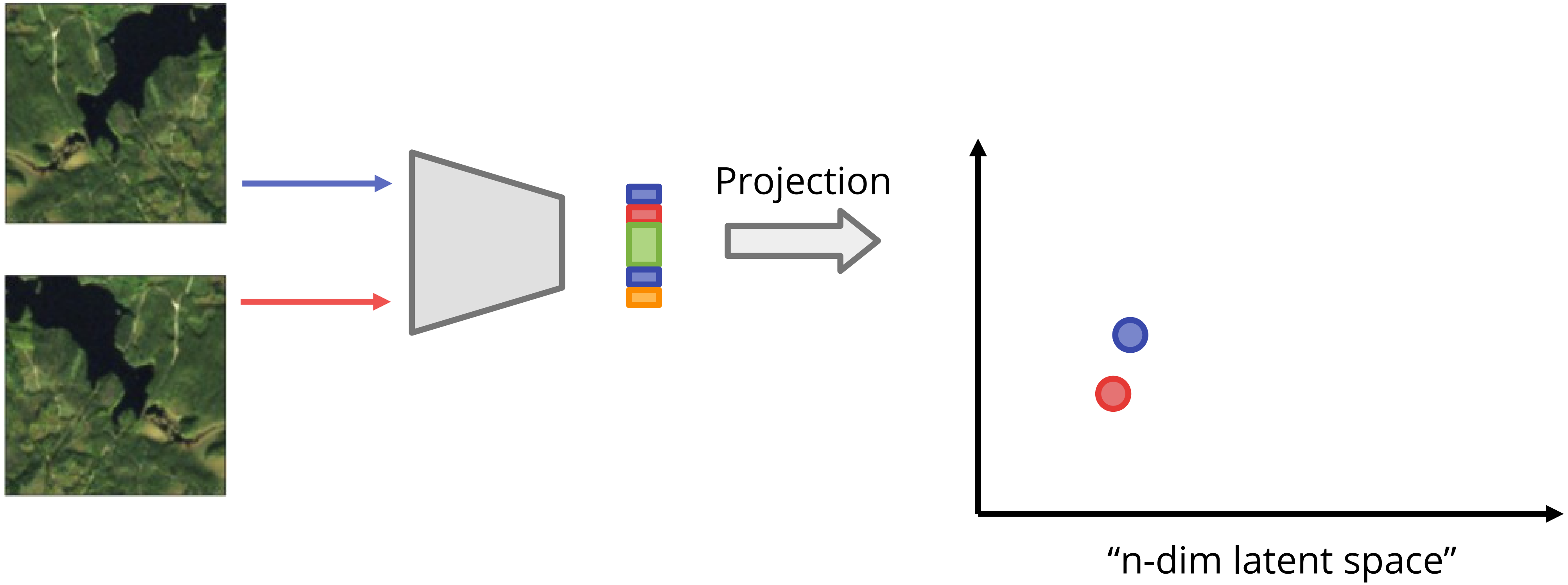


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. "A Simple Framework for Contrastive Learning of Visual Representations." ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):

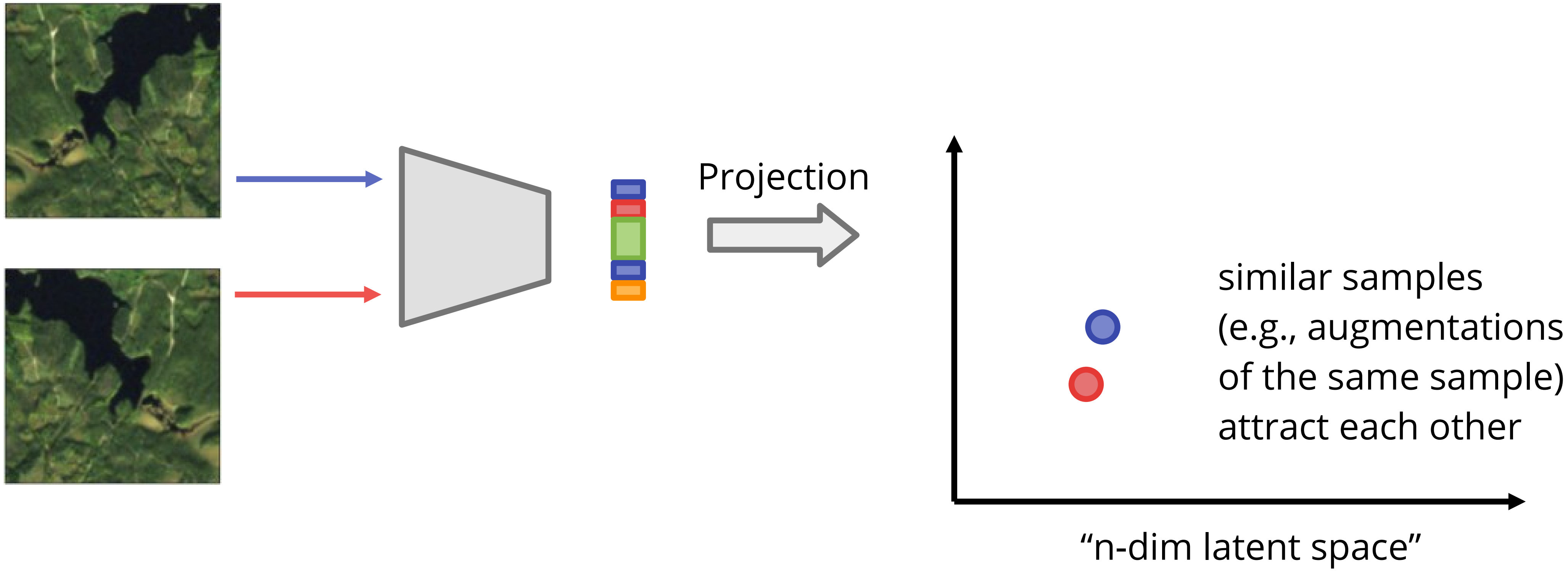


Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. "A Simple Framework for Contrastive Learning of Visual Representations." ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning

Yes we can! SSL is able to learn **rich representations** from large amounts of unannotated data.

Contrastive learning setup (following SimCLR):



Chen, Ting, Simon Kornblith, Mohammad Norouzi and Geoffrey E. Hinton. "A Simple Framework for Contrastive Learning of Visual Representations." ArXiv abs/2002.05709 (2020)

Contrastive self-supervised learning for Earth observation

Scheibenreif, L., Mommert, M., Borth, D., “Contrastive Self-Supervised Data Fusion for Satellite Imagery”, ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 1: Pre-training

Scheibenreif, L., Mommert, M., Borth, D., “Contrastive Self-Supervised Data Fusion for Satellite Imagery”, ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 1: Pre-training

- Multimodal dataset (SEN12MS, ~181k Sentinel-1/2 patch pairs)

Scheibenreif, L., Mommert, M., Borth, D., "Contrastive Self-Supervised Data Fusion for Satellite Imagery", ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 1: Pre-training

- Multimodal dataset (SEN12MS, ~181k Sentinel-1/2 patch pairs)
- Separate backbones for each modality

Scheibenreif, L., Mommert, M., Borth, D., “Contrastive Self-Supervised Data Fusion for Satellite Imagery”, ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 1: Pre-training

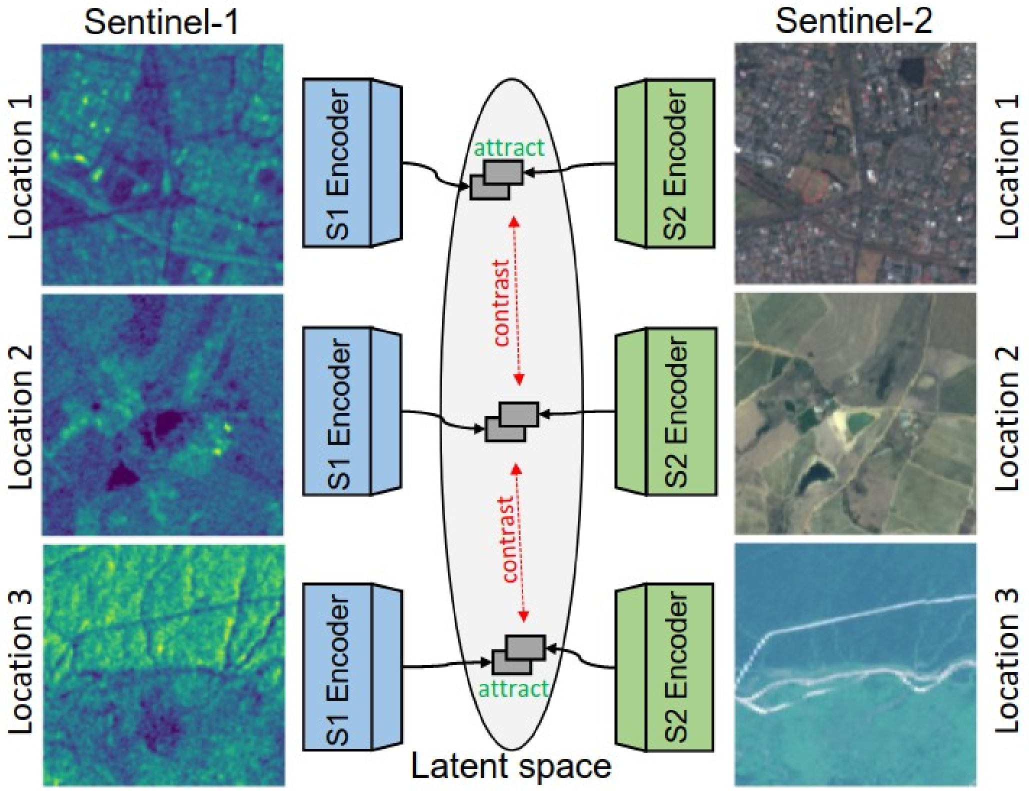
- Multimodal dataset (SEN12MS, ~181k Sentinel-1/2 patch pairs)
- Separate backbones for each modality
- Augmentation-free, contrastive setup

Scheibenreif, L., Mommert, M., Borth, D., “Contrastive Self-Supervised Data Fusion for Satellite Imagery”, ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 1: Pre-training

- Multimodal dataset (SEN12MS, ~181k Sentinel-1/2 patch pairs)
- Separate backbones for each modality
- Augmentation-free, contrastive setup



Scheibenreif, L., Mommert, M., Borth, D., "Contrastive Self-Supervised Data Fusion for Satellite Imagery", ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Scheibenreif, L., Mommert, M., Borth, D., "Contrastive Self-Supervised Data Fusion for Satellite Imagery", ISPRS 2022.

Contrastive self-supervised learning for Earth observation

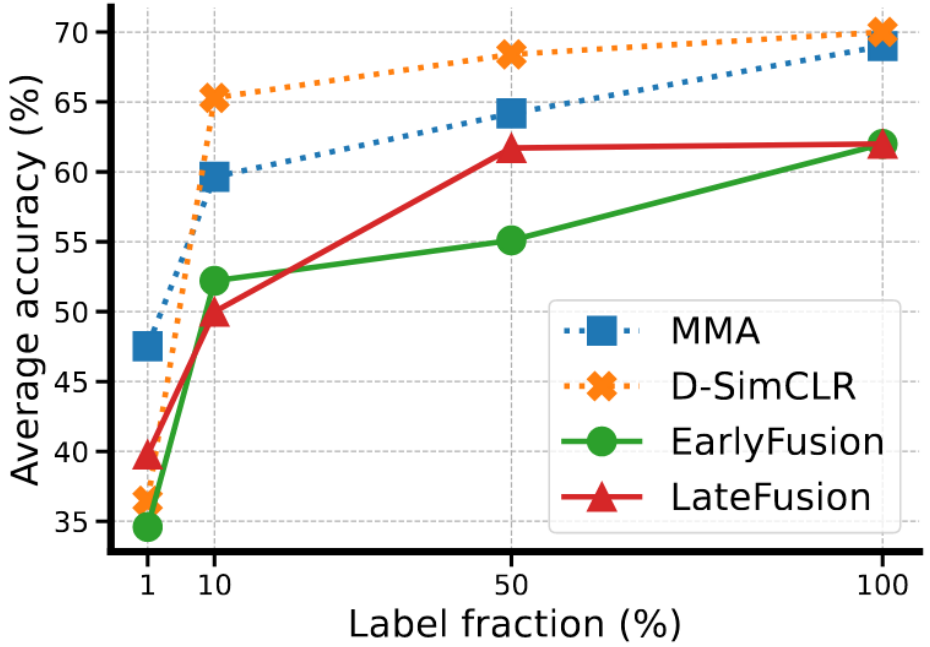
Step 2: Fine-tuning on classification task

Scheibenreif, L., Mommert, M., Borth, D., “Contrastive Self-Supervised Data Fusion for Satellite Imagery”, ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 2: Fine-tuning on classification task

- Annotations from DFC2020 high-res (10m) land use/land cover maps, ~5k patches, 8 classes

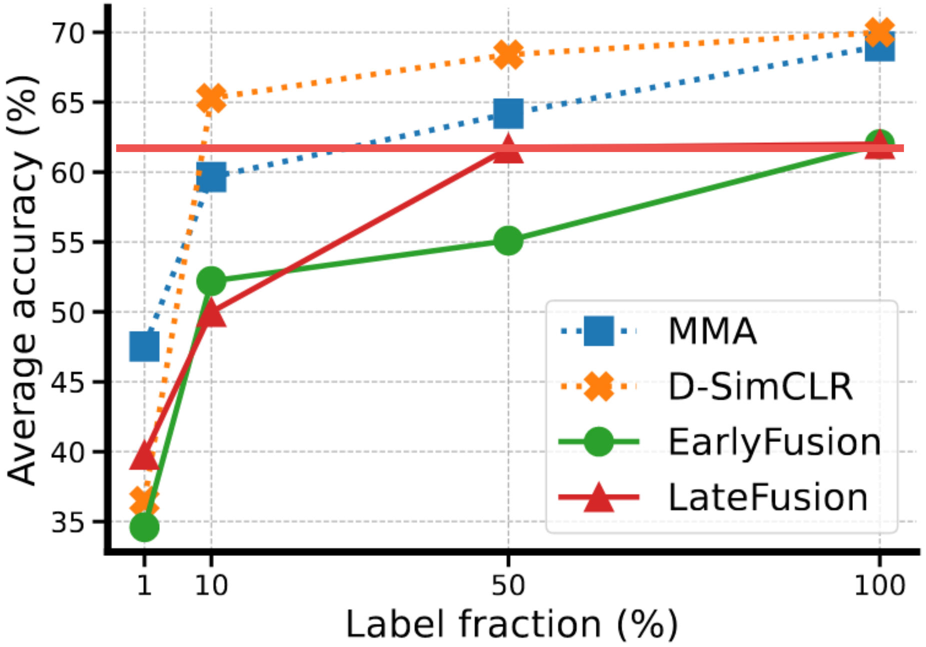


Scheibenreif, L., Mommert, M., Borth, D., “Contrastive Self-Supervised Data Fusion for Satellite Imagery”, ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 2: Fine-tuning on classification task

- Annotations from DFC2020 high-res (10m) land use/land cover maps, ~5k patches, 8 classes

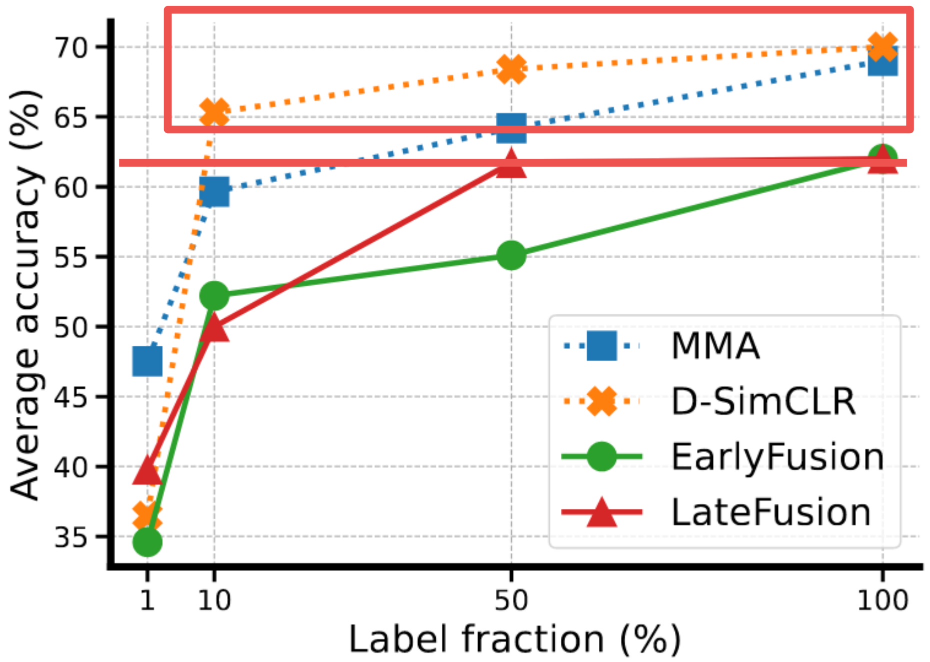


Scheibenreif, L., Mommert, M., Borth, D., “Contrastive Self-Supervised Data Fusion for Satellite Imagery”, ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 2: Fine-tuning on classification task

- Annotations from DFC2020 high-res (10m) land use/land cover maps, ~5k patches, 8 classes

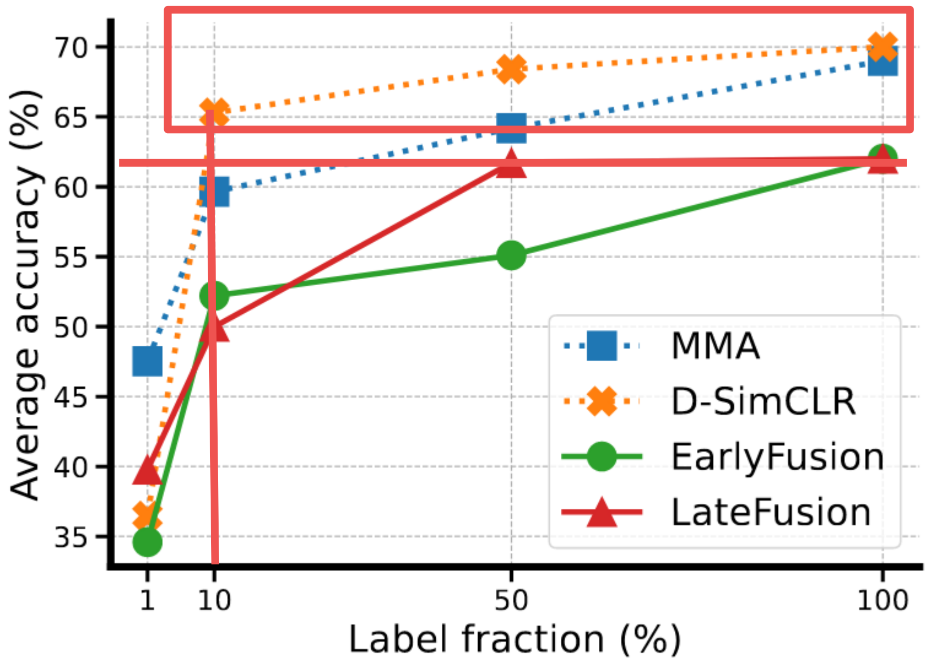


Scheibenreif, L., Mommert, M., Borth, D., "Contrastive Self-Supervised Data Fusion for Satellite Imagery", ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 2: Fine-tuning on classification task

- Annotations from DFC2020 high-res (10m) land use/land cover maps, ~5k patches, 8 classes

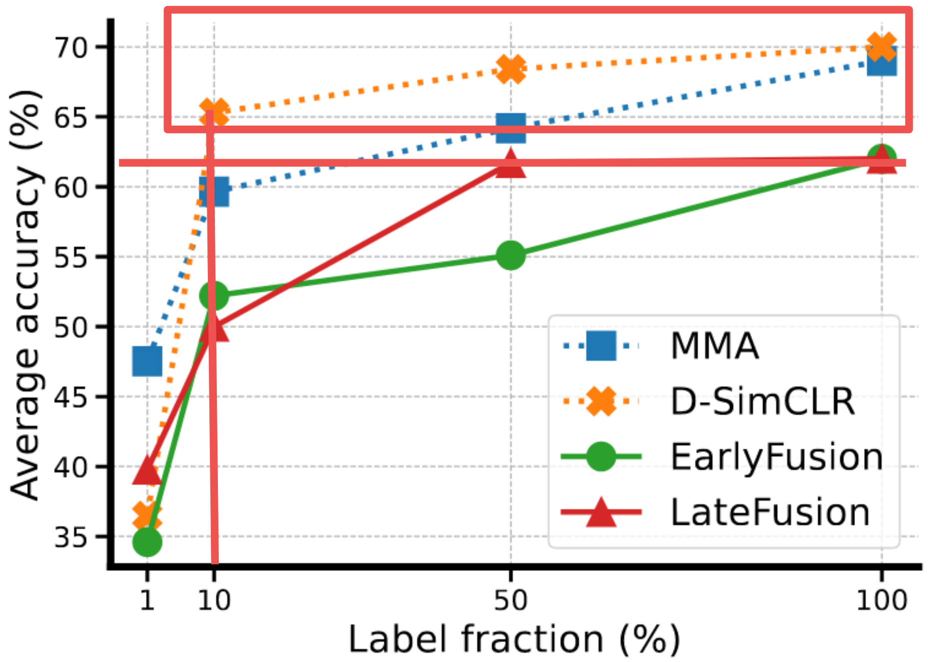


Scheibenreif, L., Mommert, M., Borth, D., "Contrastive Self-Supervised Data Fusion for Satellite Imagery", ISPRS 2022.

Contrastive self-supervised learning for Earth observation

Step 2: Fine-tuning on classification task

- Annotations from DFC2020 high-res (10m) land use/land cover maps, ~5k patches, 8 classes
- Main result: pretrained models outperform supervised baselines with only **10% of training data**

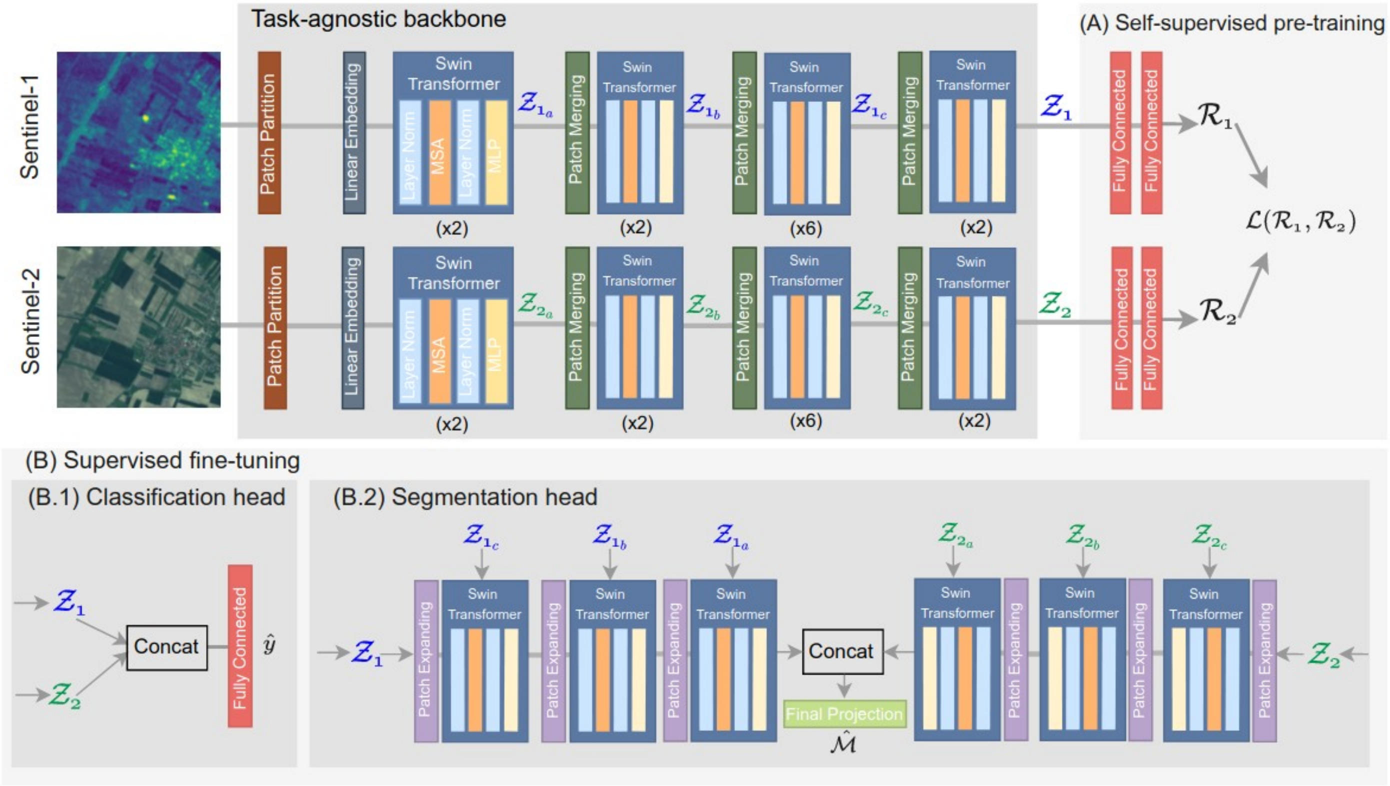


Scheibenreif, L., Mommert, M., Borth, D., "Contrastive Self-Supervised Data Fusion for Satellite Imagery", ISPRS 2022.

Self-supervised Vision Transformers for Earth observation

Scheibenreif, L., Hanna, J., Mommert, M., Borth, D., “Self-Supervised Vision Transformers for Land-cover Segmentation and Classification”, Earthvision Workshop at IEEE/CVPR 2022, publication (open access). This work was awarded the **best student paper award** of the Earthvision 2022 workshop.

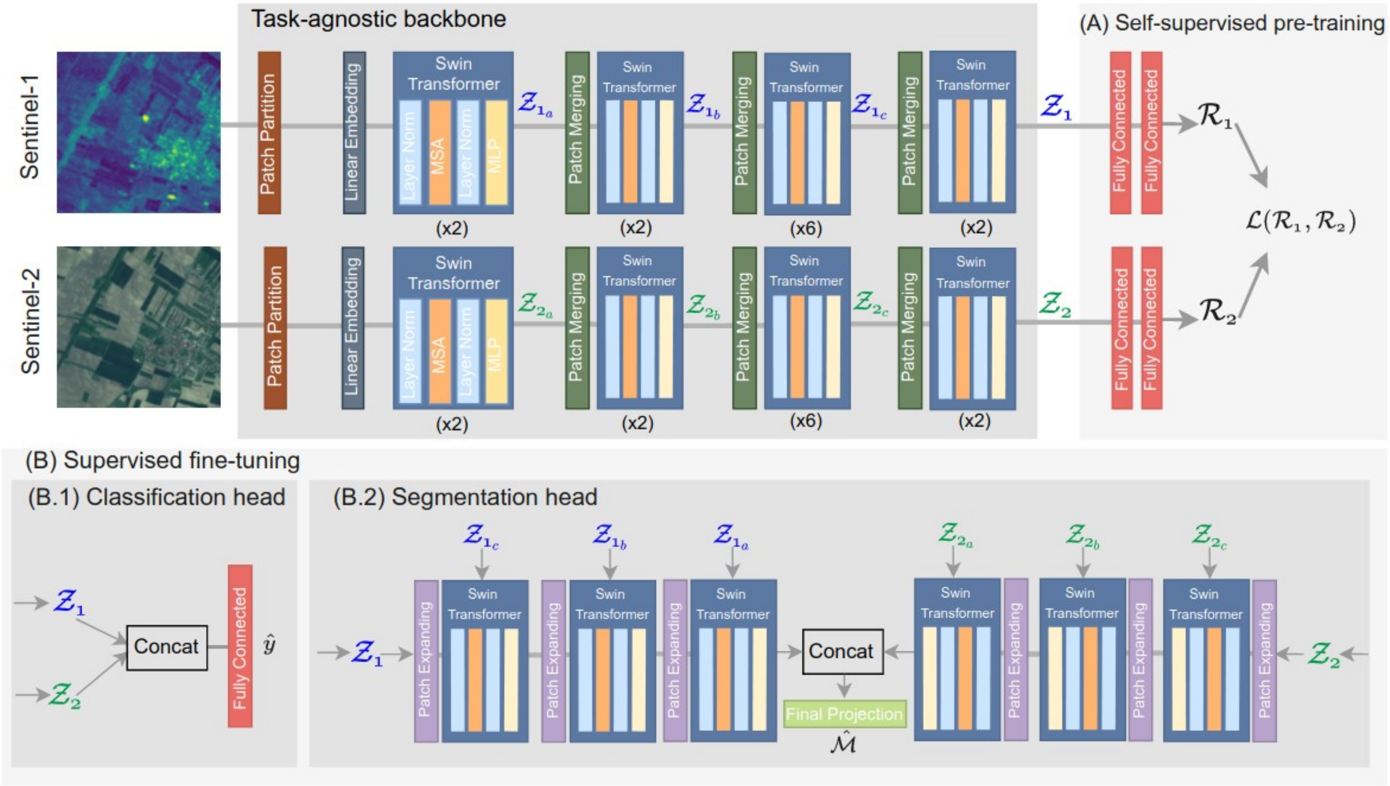
Self-supervised Vision Transformers for Earth observation



Scheibenreif, L., Hanna, J., Mommert, M., Borth, D., “Self-Supervised Vision Transformers for Land-cover Segmentation and Classification”, Earthvision Workshop at IEEE/CVPR 2022, publication (open access). This work was awarded the **best student paper award** of the Earthvision 2022 workshop.

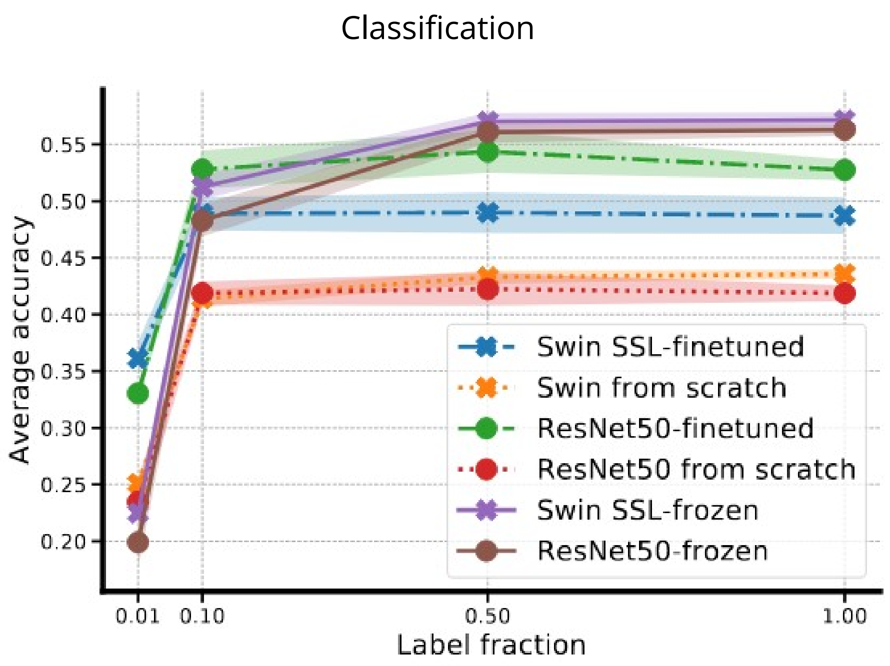
Self-supervised Vision Transformers for Earth observation

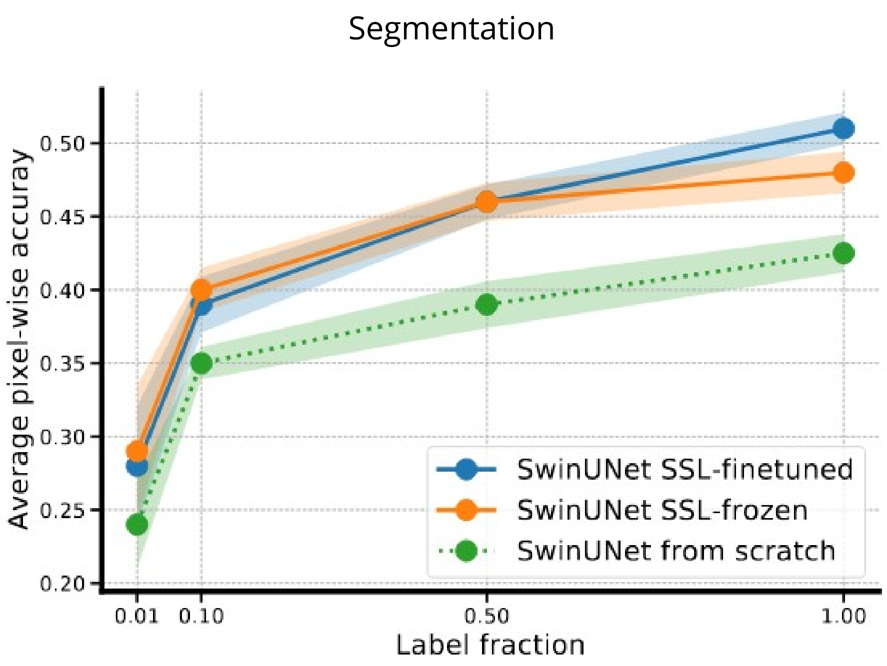
We retry our experiment with a **Transformer architecture**.



Scheibenreif, L., Hanna, J., Mommert, M., Borth, D., “Self-Supervised Vision Transformers for Land-cover Segmentation and Classification”, Earthvision Workshop at IEEE/CVPR 2022, publication (open access). This work was awarded the **best student paper award** of the Earthvision 2022 workshop.

Self-supervised Vision Transformers for Earth observation





The effect of SSL is **task agnostic** and **architecture agnostic**.

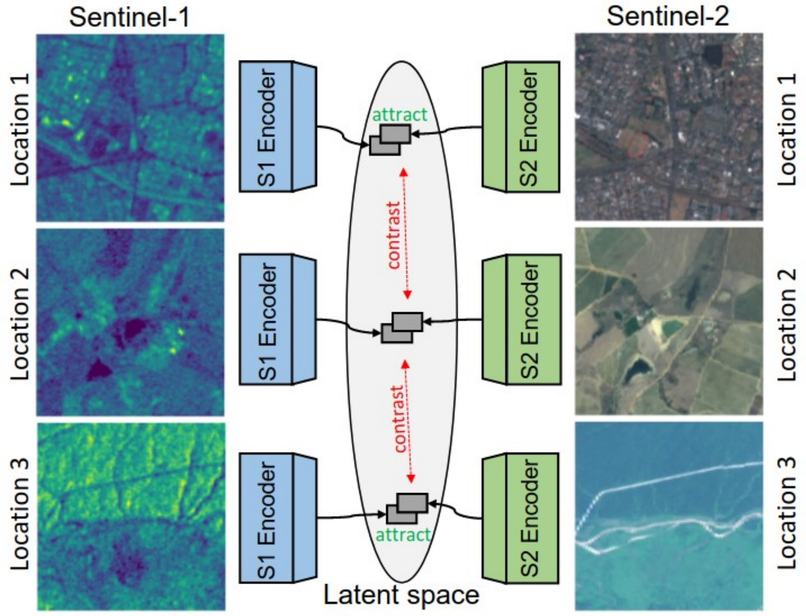
→ SSL pretraining can boost the performance significantly, especially for small datasets!

Scheibenreif, L., Hanna, J., Mommert, M., Borth, D., “Self-Supervised Vision Transformers for Land-cover Segmentation and Classification”, Earthvision Workshop at IEEE/CVPR 2022, publication (open access). This work was awarded the **best student paper award** of the Earthvision 2022 workshop.

Summary

Self-supervised learning (SSL) improves the data-efficiency of model training by leveraging vast amounts of unannotated data:

- We show that models pre-trained with SSL and fine-tuned on **10% of annotated data** outperform models trained on 100% of annotated data in a fully supervised approach.
- This performance boost is **task-agnostic** and **architecture-agnostic**.



Michael Mommert
University of St.Gallen, School of Computer Science
Rosenbergstrasse 30, 9000 St.Gallen, Switzerland

michael.mommert@unisg.ch
mommermi.github.io

Self-Supervised Learning for Earth Observation: Leveraging a wealth of multi-modal data

SWISS NATIONAL SCIENCE FOUNDATION

submitted to IGARSS 2023

BEN-GE: EXTENDING BIGEARTHNET WITH GEOGRAPHICAL AND ENVIRONMENTAL DATA

Michael Mommert[1], Nicolas Kesseli[1], Joëlle Hanna[1], Linus Scheibenreif[1], Damian Borth[1], Begüm Demir[2]

(1): School of Computer Science, Universität St. Gallen; (2): Technische Universität Berlin

submitted to IGARSS 2023

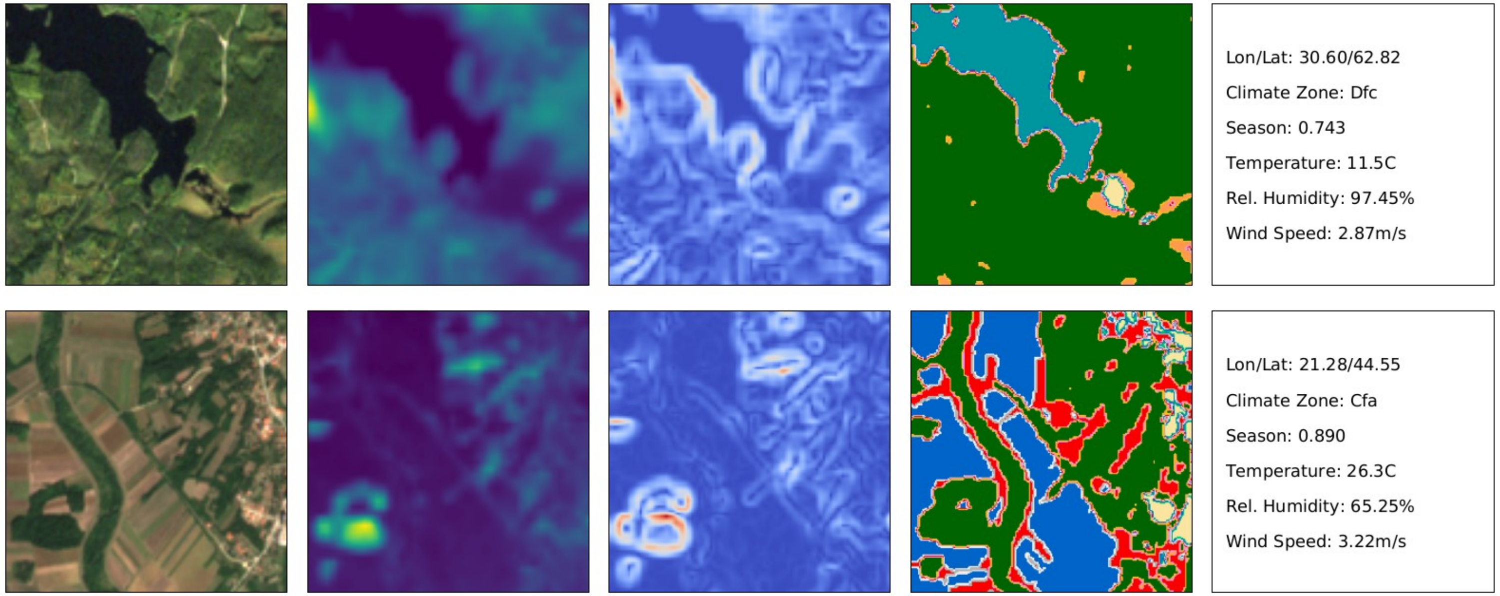
Self-Supervised Learning for Earth Observation: Leveraging a wealth of multi-modal data

SWISS NATIONAL SCIENCE FOUNDATION

BEN-GE: EXTENDING BIGEARTHNET WITH GEOGRAPHICAL AND ENVIRONMENTAL DATA

Michael Mommert[1], *Nicolas Kesseli*[1], *Joëlle Hanna*[1], *Linus Scheibenreif*[1], *Damian Borth*[1], *Begüm Demir*[2]

(1): School of Computer Science, Universität St. Gallen; (2): Technische Universität Berlin



Elevation Slope + Az Land-use Environmental x **590,326** tiles

submitted to IGARSS 2023

Swiss National Science Foundation

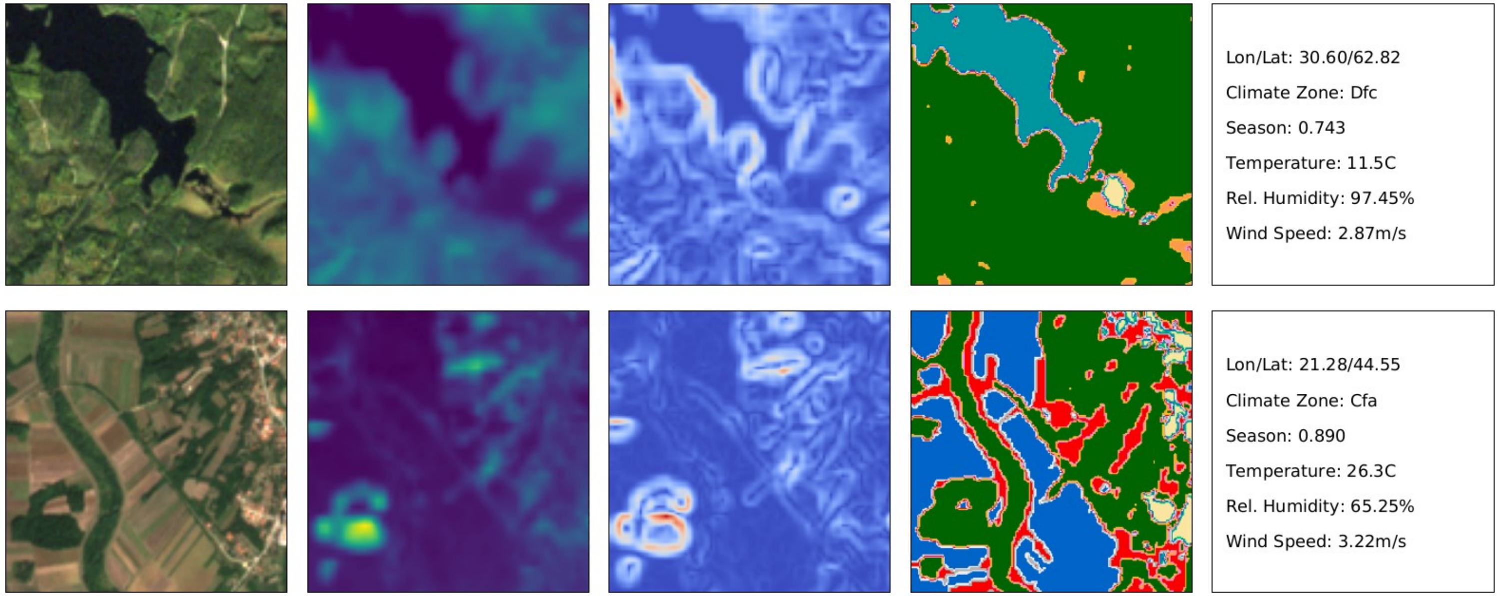
Self-Supervised Learning for Earth Observation: Leveraging a wealth of multi-modal data

BEN-GE: EXTENDING BIGEARTHNET WITH GEOGRAPHICAL AND ENVIRONMENTAL DATA

Michael Mommert[1], *Nicolas Kesseli*[1], *Joëlle Hanna*[1], *Linus Scheibenreif*[1], *Damian Borth*[1], *Begüm Demir*[2]

(1): School of Computer Science, Universität St. Gallen; (2): Technische Universität Berlin

Stay tuned!



Elevation | Slope + Az | Land-use | Environmental | x **590,326** tiles

submitted to IGARSS 2023