What is self-supervised learning?

Devis Tuia

ECEO laboratory

EPFL

RepreSent / F-TEP,
04.05.2023

EPFL

École polytechnique fédérale de Lausanne

EO data processing is better with supervision!

- Several methods exist for unsupervised classification and change detection
- They are usually outperformed by supervised methods [Volpi et al., 2013]
- This is because the desired categories of interest are defined explicitly

- With supervision you can
  - Discriminate types of changes
  - Easily separate them from application irrelevant spectral variations (illumination, phenology, …)

M. Volpi, D. Tuia, F. Bovolo, M. Kanevski, and L. Bruzzone. Supervised change detection in VHR images using contextual information and support vector machines. *Int. J. Appl. Earth Obs. Geoinf.*, 20:77–85, 2013.

But gathering labeled datasets is HARD

- Changes can be of different type
- Perturbation factors can be of many kinds
- Each problem calls for its own data collection

- For specific tasks, we often have only small datasets
- e.g. urban changes, OCSD dataset (Daudt et al., 2018):
  - 14 labeled image pairs from Sentinel 2 available (+ 10 with labels undisclosed)
  - Binary changes, mainly urban areas

- Deep learning models struggle to learn and generalise.

R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, “Urban change detection for multispectral earth observation using convolutional neural networks,” in IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 2115– 2118, 2018.

Image, t_1

Image, t_2

Ground truth

The dream: how can we pre-train a CNN without labels?

In a dream world, we would like to have

- the benfits of supervised learning
- without having to have any labels

Also, in a dream world, we would like the representations learned to generalise well across tasks and geographies

The dream: how can we pre-train a CNN without labels?

In a dream world, we would like to have

- the benfits of supervised learning
- without having to have any labels

Also, in a dream world, we would like the representations learned to generalise well across tasks and geographies

We might not live in a dream world,
but we can get close enough!

Self supervised learning

In a nutshell: let's replace labels with artificial labels that we can generate using the data only

~~Labels~~ → Artificial labels

The artificial labels need to help us build features that are robust to variations we are not interested in.

We call the task predicting those artificial labels a pretext task.

Examples:

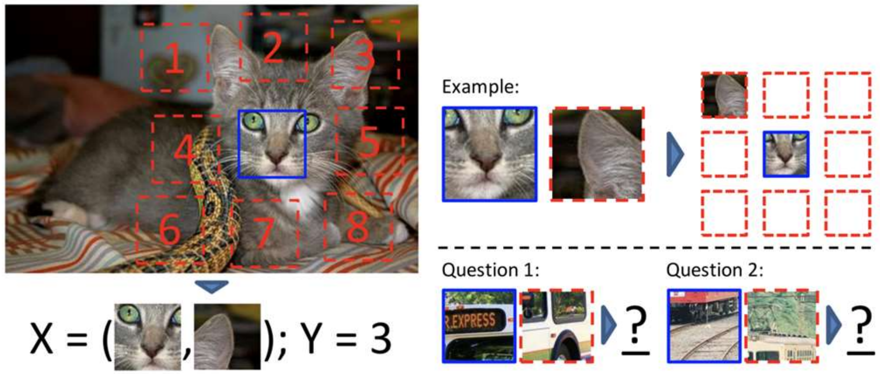
Examples of pretext tasks

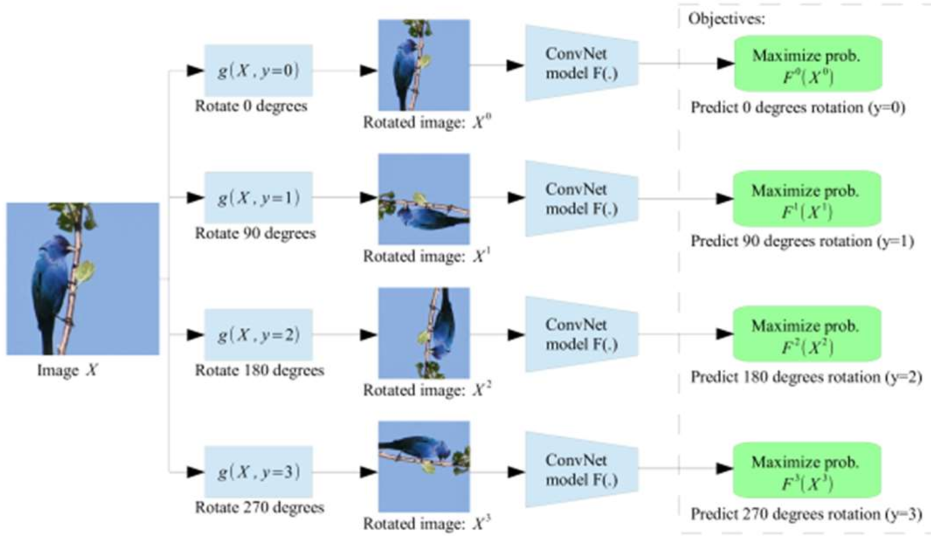
- Learning relative positions of patches helps understanding spatial structures

  [Doersch et al. ICCV 2015]

- Learning rotation angles teaches rotation invariance

  [Gidaris et al. ICLR 2018]

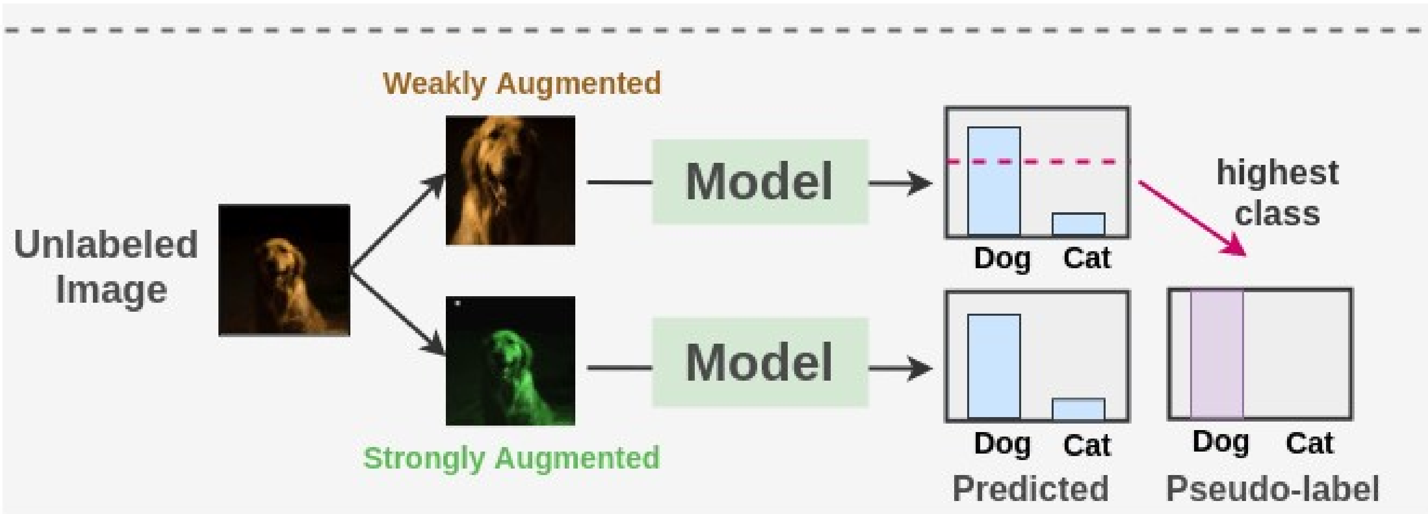




C. Doersch, A. Gupta, and A. A. Efros, “Unsupervised visual representation learning by context prediction,” in IEEE International Conference on Computer Vision (ICCV), pp. 1422–1430, 2015.
S. Gidaris, P. Singh, and N. Komodakis, “Unsupervised representation learning by predicting image rotations,” in International Conference on Learning Representations (ICLR), 2018.

Examples of pretext tasks

- Learning consistency across augmentations makes your classifier robust to small shifts, jitter, etc.



Xie et al., Unsupervised Data Augmentation for Consistency Training, Neurips 2020.

Self supervised learning (2)

- In all these examples, you don't have to label anything!

1. split the image in parts and predict relative positions
2. rotate the image
3. apply augmentations and make sure that predictions do not deviate

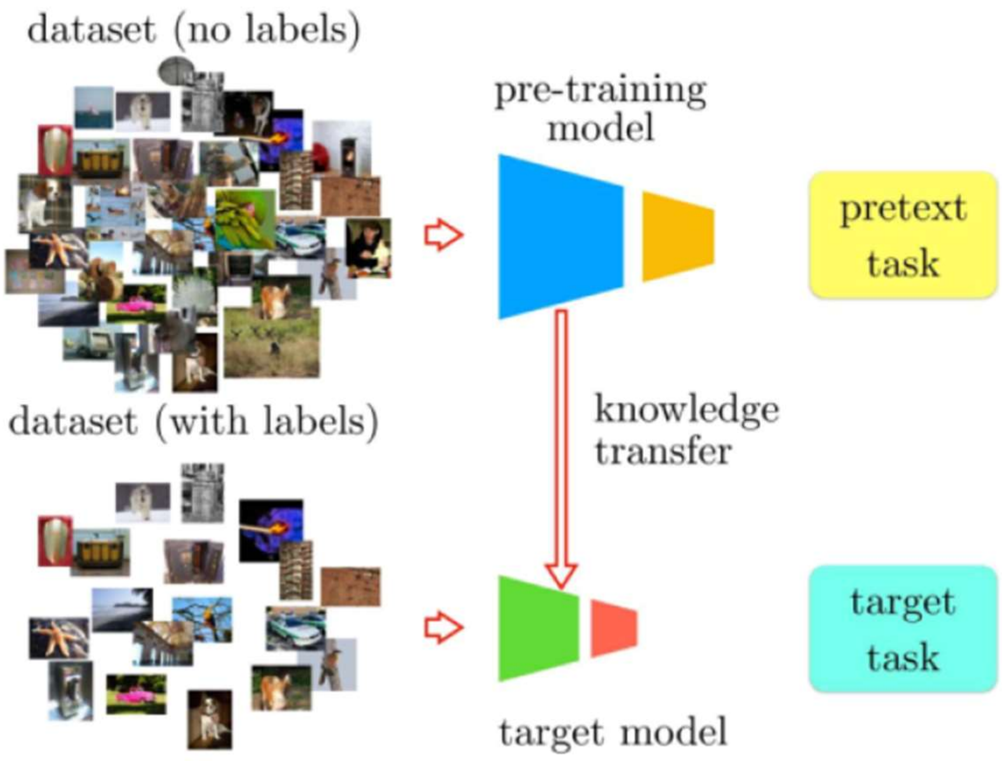
- The labels are an inherent part of your data

EPFL

Self supervised learning is usually a 2 steps business

1. you learn the pretext task without labels
   - A task for which the labels can be extracted automatically from the data;
   - A task that is connected to the main one. Learning it helps the main one.

2. you use the few labels you have to finetune to your true problem.



Groundai.com

SSL concepts in remote sensing (1)

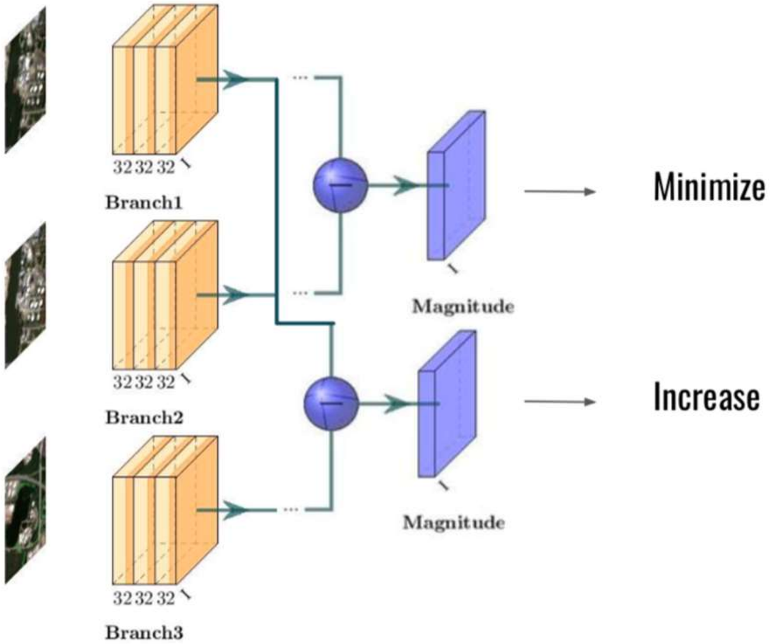
Leenstra et al., 2021: change detection in an unsupervised way

- Discounts application-irrelevant changes with SSL
- Assumes no change is present in randomly sampled image pairs
- Uses contrastive learning to make features of pairs co-located more similar and push away pairs located in different places.

Leenstra, M., Marcos, D., Bovolo, F. and Tuia, D., 2021. Self-supervised pre-training enhances change detection in Sentinel-2 imagery. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VII* (pp. 578-590). Springer International Publishing.

SSL concepts in remote sensing (1)

- Leenstra et al.: change detection in an unsupervised way



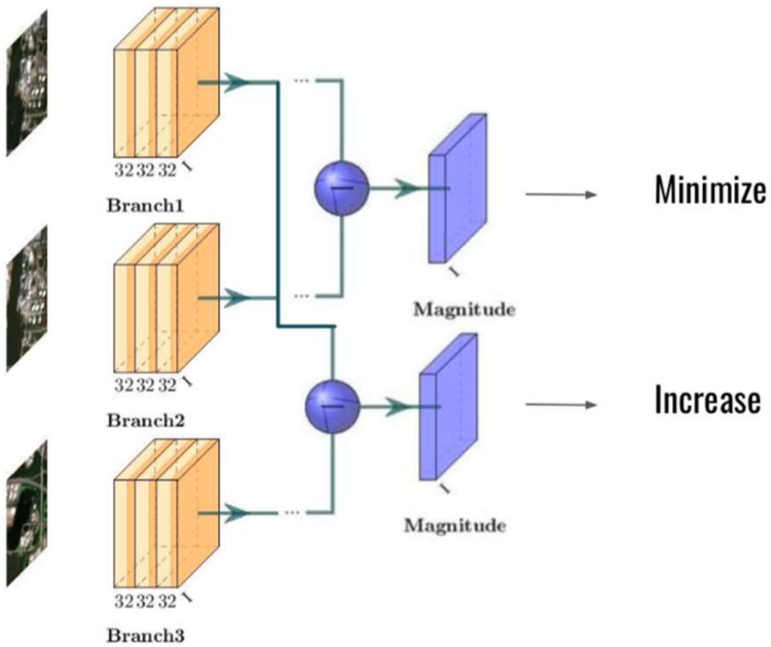
Contrastive learning objective (triplet loss):

$$L = max(||\mathbf{f}^1 - \mathbf{f}^2||_2 - ||\mathbf{f}^1 - \mathbf{f}^3||_2 + m, 0) + \gamma \cdot |\mathbf{f^1} - \mathbf{f^2}|$$

Leenstra, M., Marcos, D., Bovolo, F. and Tuia, D., 2021. Self-supervised pre-training enhances change detection in Sentinel-2 imagery. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VII* (pp. 578-590). Springer International Publishing.

SSL concepts in remote sensing (1)

- Leenstra et al.: change detection in an unsupervised way



Contrastive learning objective (triplet loss):

Most entries of f¹ and f² vectors must be the same (makes CVA easier)

$$L = max(\boxed{||\mathbf{f}^1 - \mathbf{f}^2||_2} - \boxed{||\mathbf{f}^1 - \mathbf{f}^3||_2} + m, 0) + \boxed{\gamma \cdot |\mathbf{f}^1 - \mathbf{f}^2|}$$

Features of co-located patches must be similar

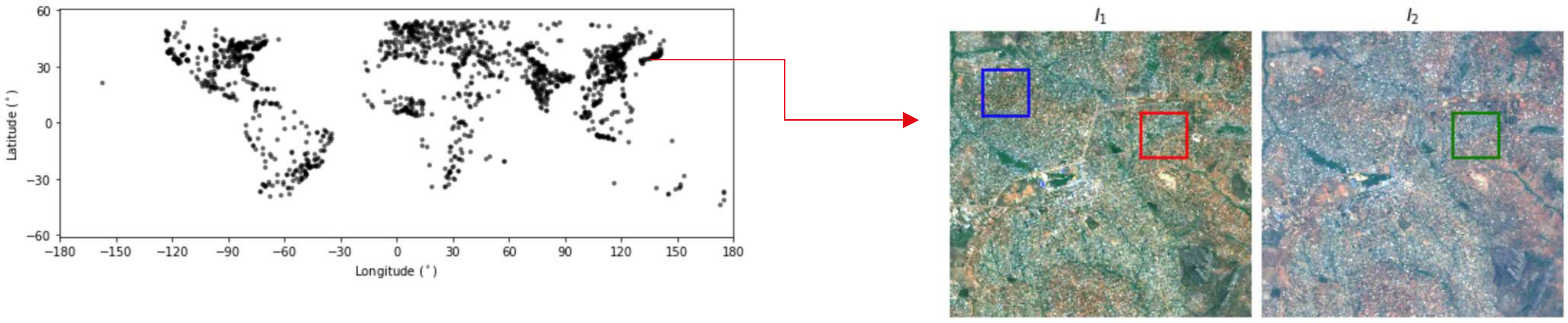
Features of <u>not</u> co-located patches must <u>not</u> be similar

Leenstra, M., Marcos, D., Bovolo, F. and Tuia, D., 2021. Self-supervised pre-training enhances change detection in Sentinel-2 imagery. In *Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10-15, 2021, Proceedings, Part VII* (pp. 578-590). Springer International Publishing.

EPFL

Data for pre-training: the Sentinel-2 multitemporal cities pairs (S2MTCP) dataset

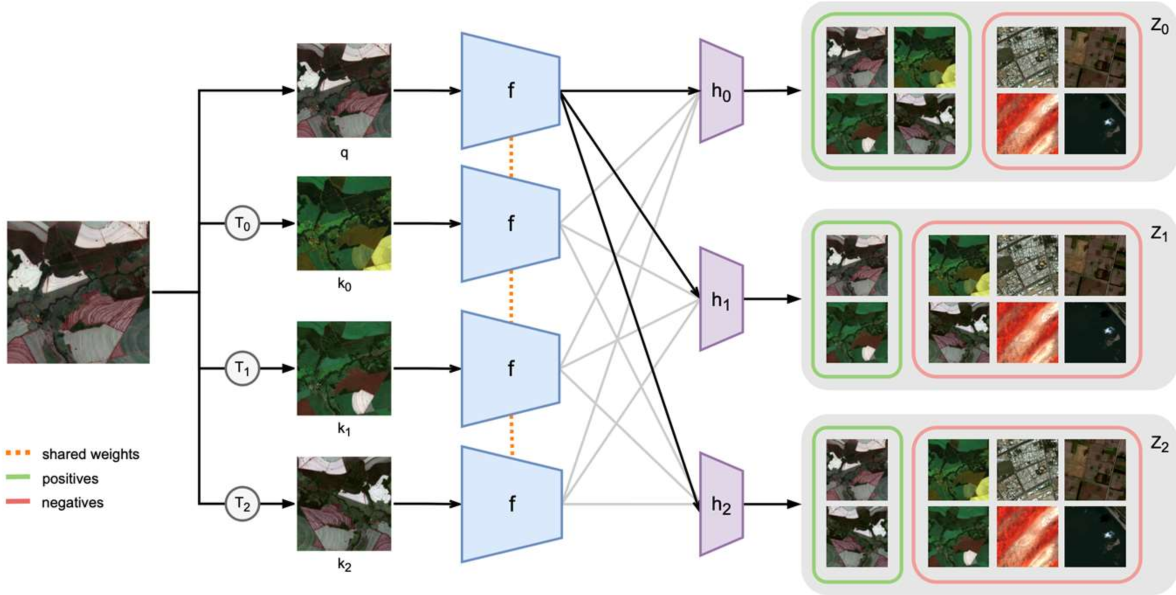
- 1'520 image pairs
- From 1'520 different cities in the world
- Each pair is a 600 x 600 pixels pair of tiles
- Available on https://zenodo.org/record/4280482#.X9I2Ri-ZOUk



SSL concepts in remote sensing (2)

Manas et al.: Seasonal contrast: discounting for seasonal effects in time series

- Discounting application irrelevant (e.g. illumination/seasonal) changes with SSL

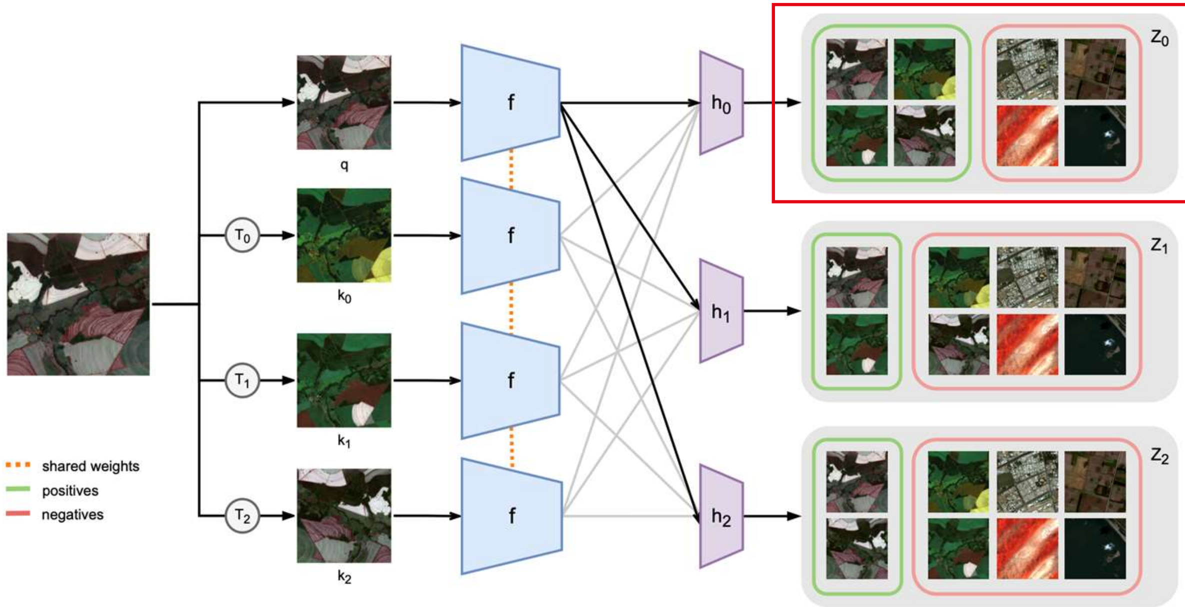


Manas, O., Lacoste, A., Giró-i-Nieto, X., Vazquez, D. and Rodriguez, P., 2021. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*(pp. 9414-9423).

SSL concepts in remote sensing (2)

Manas et al.: Seasonal contrast: discounting for seasonal effects in time series

- Discounting application irrelevant (e.g. illumination/seasonal) changes with SSL



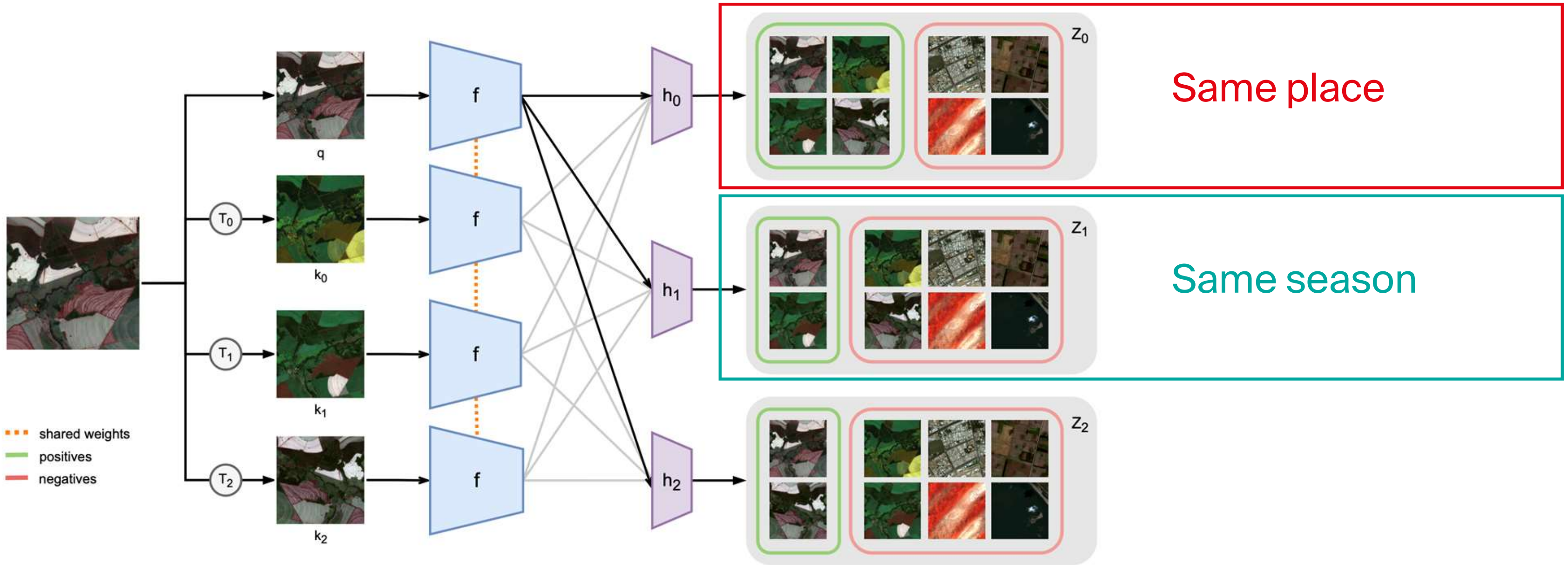
Same place

Manas, O., Lacoste, A., Giró-i-Nieto, X., Vazquez, D. and Rodriguez, P., 2021. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*(pp. 9414-9423).

SSL concepts in remote sensing (2)

Manas et al.: Seasonal contrast: discounting for seasonal effects in time series

- Discounting application irrelevant (e.g. illumination/seasonal) changes with SSL

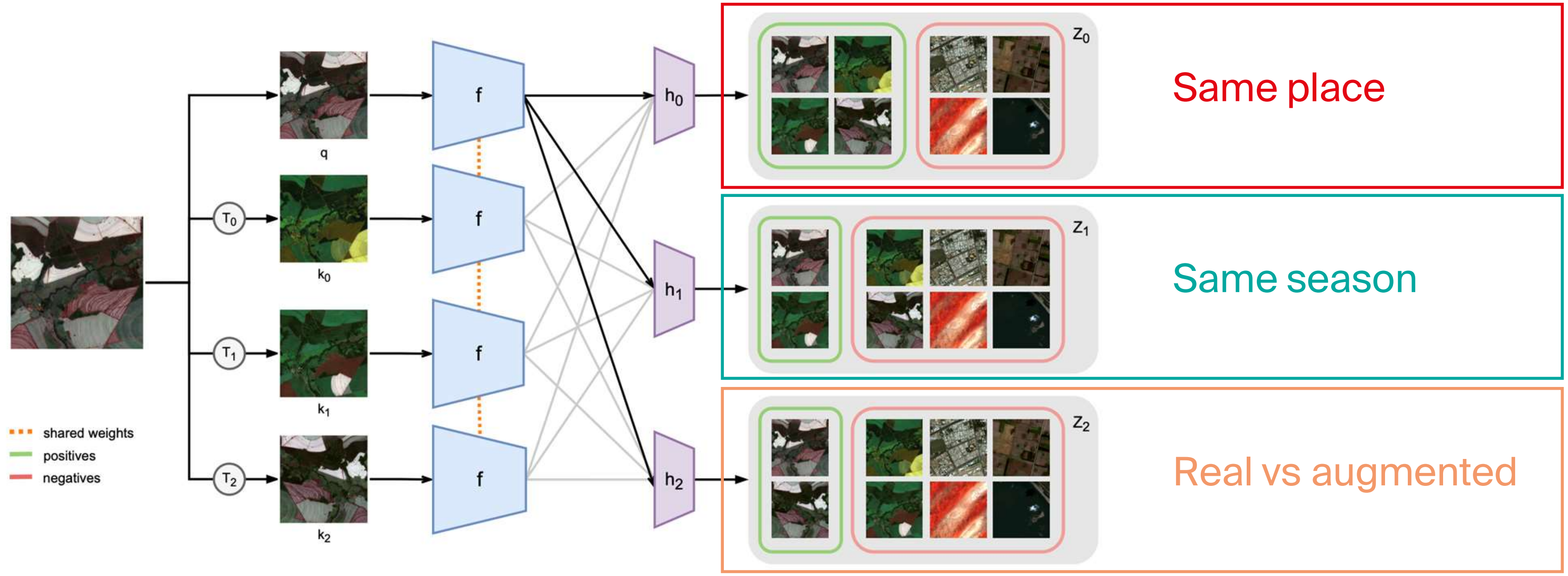


Manas, O., Lacoste, A., Giró-i-Nieto, X., Vazquez, D. and Rodriguez, P., 2021. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*(pp. 9414-9423).

SSL concepts in remote sensing (2)

Manas et al.: Seasonal contrast: discounting for seasonal effects in time series

- Discounting application irrelevant (e.g. illumination/seasonal) changes with SSL



Manas, O., Lacoste, A., Giró-i-Nieto, X., Vazquez, D. and Rodriguez, P., 2021. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*(pp. 9414-9423).

SSL concepts in remote sensing (2)

Manas et al.: Seasonal contrast: discounting for seasonal effects in time series

- Again: it works because large dataset to learn from.

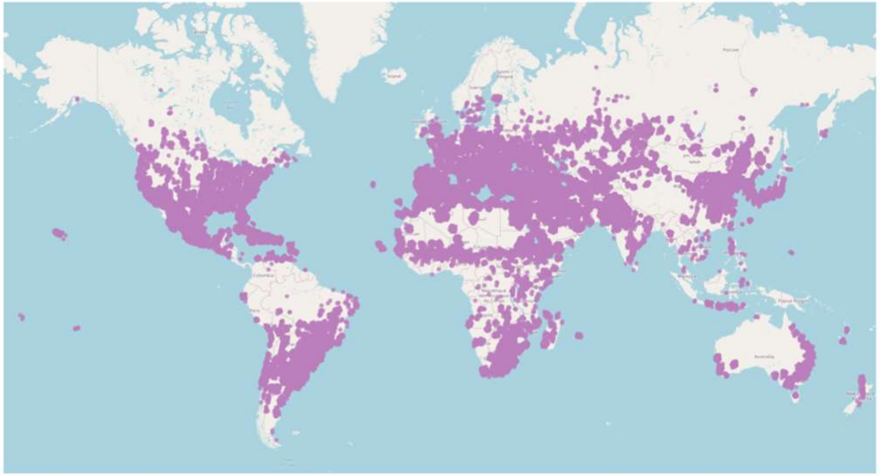


Figure 1. **Distribution of the Seasonal Contrast (SeCo) dataset.**

Manas, O., Lacoste, A., Giró-i-Nieto, X., Vazquez, D. and Rodriguez, P., 2021. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*(pp. 9414-9423).

To conclude

- SSL helps you learning robust representations for deep learning
- SSL does not solve your final task!
- In SSL you must be able to craft your pretext task directly from the data
- The pretext data must make the final task easier to solve