# SELF-SUPERVISED LEARNING FOR CROP CLASSIFICATION USING PLANET FUSION

Caglar Senaras[1*], Piers Holden[2], Timothy Davis[2], Akhil Singh Rana[2], Maddie Grady[2], Annett Wania[2], Richard de Jeu[1]

[1] Planet Labs Haarlem, Netherlands - (caglar.senaras, ries )@planet.com
[2] Planet Labs Germany GmbH, Berlin, Germany - (piers, akhil.singh.rana, tim.davis, maddie.grady, annett )@planet.com

**KEY WORDS:** Self-supervised learning, Deep Learning, RapidAI4EO, Crop Classification, Planet Fusion

**ABSTRACT:**

Benefiting from the high cadence and spatial resolution of the new generation of Earth observation satellites, remote sensing technology is allowing us to derive more valuable information for the agricultural sector. Crop classification is one of the fundamental information derivatives from Earth observation data researchers used for food security, crop monitoring, and economic assessment. The robustness of a crop classification model to variations in environmental and management conditions due to time and location is one of the crucial requirements. To achieve this, we developed a novel self-supervised method using the advantage of unlabeled samples and transformer architectures. We used six different areas in Germany and four years to evaluate the robustness of the model. Our experiments showed that self-supervised deep learning methods could provide a significant advantage in handling these variations. In some cases, we observed around 30 percentage points improvements in F1-score performance compared to a Random Forest based model.

## 1. INTRODUCTION

Developing agricultural policies and sustainable farm management is essential for handling the increasing global demand for food. Accurate and timely information about the types and extent of crops present in an area is necessary for evaluating crop health and productivity, forecasting crop yields, and optimizing farming practices. For that reason, developing a robust machine learning-based crop classification method has become an important research topic.

Most state-of-the-art crop classification approaches use the advantages of supervised learning methods such as Random Forest (RF), Support Vector Machines (SVM) (Saini, 2018). As with other remote sensing applications, supervised deep learning-based methods have also become popular for crop analysis in recent years (He, 2019), (Zhao, 2021). These deep learning approaches have demonstrated strong performance on datasets with limited geographical and temporal extent (Metzger, 2021). However, large datasets with crop type labels covering multiple years and countries have not been readily available until recently (eg. EuroCrops (Schneider, 2021)). Due to the data-intensive nature of training deep methods, this historical lack of data has limited efforts to comprehensively evaluate the potential of deep learning methods to perform at continental scale, across the full range of climate conditions that require many years of data to be fully expressed.

Self-supervised learning (SSL) represents a paradigm-shift in the machine learning domain, enabling the pre-training of deep learning models using sets of unlabeled samples. Many studies demonstrated that models pre-trained with SSL have better classification performance than fully supervised models when the number of labelled samples is limited (Wang, 2022). They showed that self-supervision outperforms supervision when reducing the number of labels using BigEarthNet (Sumbul, 2019) land cover classification dataset. In one of the recent studies,

Marszalek pretrained a deep learning-based crop classifier with SSL on a limited number of unlabeled data (Marszalek, 2022). They trained the crop classifiers for a given region by using multiple years and measured the performance. According to them, it is essential to introduce a few labels for the predicted year during the training.

In this study, we adapt one of the state-of-the-art SSL architectures, called SimSiam (Chen, 2021), to pre-train our attention-based crop classifier. Since we need a large number of unlabeled samples to train our model in a self-supervised manner effectively, we combine the advantages of two datasets: RapidAI4EO (Marchisio, 2021) and EuroCrops. EuroCrops provides the boundary of the fields around Europe. On the other hand, RapidAI4EO satisfies the need for high-resolution and high-cadence time series for some of these fields. To provide a reliable crop classification approach, spatial and temporal robustness become essential, where spatial robustness refers to the ability of an algorithm to classify crops in different locations accurately, and temporal robustness refers to its ability to classify crops accurately in different years. Therefore, we design a set of experiments to compare the ability of a deep learning model pre-trained with SSL to generalize to unseen spatial and temporal contexts, against a baseline of deep learning models (without SSL pre-training) and Random Forest models.

The following Section 2 will describe the proposed method, including the sample selection strategy using an unlabeled SSL dataset and the deep learning architecture. Section 3 introduces the dataset used for the crop classification downstream task. Section 4 presents the results of our experiments, including data analysis to observe the impact of these (temporal and spatial) variability on crops and quantitative evaluations of the classifiers' performance. Finally, in Section 5, we conclude the paper by summarizing our main findings and discussing their implications for future work.

---

[*] Corresponding author

## 2. PROPOSED METHOD

The proposed method includes two important steps. In the first one, we combine publicly available datasets to prepare a large number of positive pairs, which will be used during the pretraining of the model in an SSL fashion without any crop type information. After that, we pretrain our model using SSL and apply supervised training to our model to handle crop classification downstream tasks.

### 2.1 Self-Supervised Learning Dataset

As mentioned in the introduction, we integrate the advantages of two datasets: RapidAI4EO and EuroCrops.

The RapidAI4EO corpus, which was produced under Horizon2020 Project RapidAI4EO (https://rapidai4eo.eu/), is one of the most extensive remote sensing datasets, and it is publicly available to the machine learning and remote sensing community. It contains two data sources, Sentinel-2 and Planet Fusion (Planet Fusion, 2021), and covers 500,000 patches of 600m x 600m, distributed across Europe (Figure 1).

In this study, we use Planet Fusion data, which provides a sensor-harmonized, gap-filled, and cloud-free time series. It represents the highest spatial and temporal resolution data in the corpus with a ground resolution of three meters, five-day cadence, and four spectral bands (VNIR) for 2018 and 2019. A major advantage of using Planet Fusion imagery was that it provided regular, complete and high-quality time series that required almost no pre-processing to prepare them for use in our machine learning pipelines, highlighting its suitability for these types of workflows.

The EuroCrops dataset, an initiative of the Technical University of Munich (https://www.eurocrops.tum.de/), aims to combine all publicly available self-declared crop reporting datasets from countries of the European Union. It provides the boundaries of the fields and the crop type for a given year.

We observed that 91,944 RapidAI4EO locations intersect with 862,691 EuroCrops fields in total. However, for most countries, EuroCrops includes the crop type information after 2019. Since the temporal range of the RapidAI4EO corpus does not cover this period, we use only the field boundary information. We assume that farmers planted only one type of crop at a given time. As a result, we can assume that randomly selected samples in a field will represent the same crop type.
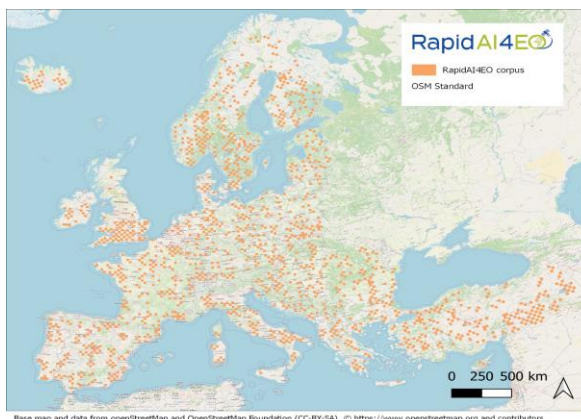


**Figure 1.** The approximate location of the 500,000 RapidAI4EO patches

To prepare our unlabeled dataset, we select 24,407 RapidAI4EO locations and use the largest two fields for each location. For each field, we randomly pick ten samples (Figure 2). During the training of the SSL model, two of these samples are randomly selected as positive pairs and fed to our model as described in Section 2.2.
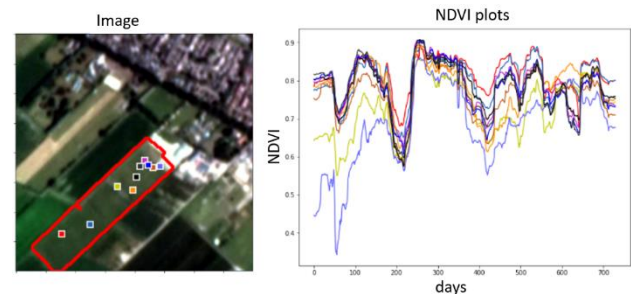


**Figure 2.** Left: An example RapidAI4EO patch, a field inside that patch (red polygon in the left image), and ten sample points inside the field. Right: 2-year NDVI time-series (right) of the collected ten samples for the given field.

### 2.2 SSL Crop Classification Architecture

In this study, we adapted SimSiam to introduce a simple and effective framework for unsupervised representation learning based on the Siamese network architecture without requiring any negative pairs. In SimSiam, the model is trained to predict the parameters of a fixed encoder network, given the inputs of an augmented view of the same data. Thus the model learns to identify essential characteristics in the data, creating a valuable representation that can be utilized for the downstream tasks.

After the positive pairs are generated, as explained in Section 2.1, we use them for self-supervised learning. For each pair, we obtain daily time series of the surface reflectance (SR) values and feed these time series to SimSiam to train our attention-based encoder. The encoder is inspired by (Garnot, 2020). The encoder combines the Pixel-Set Encoder (PSE) and the Transformer architecture to extract features from the input images. We adapt the PSE to use a single SR value for each field by using fully connected layers. After the pretraining, we modify our encoder to handle crop classification downstream by adding fully connected and soft-max layers.

## 3. CROP CLASSIFICATION DATASET

For the crop classification downstream task, we create a new dataset that will allow us to evaluate the temporal and spatial robustness of the models across a single-nation context. To achieve our goal, we selected six areas in Germany (Figure 3).

Area-5 and Area-6 were previously used in the AI4FoodSecurity[2] challenge. The selected 4 new areas were longitudinally stratified, to ensure variability in climate, planting practices and environmental conditions, each being part of a distinct region. Each area has a size of 24 km x 24 km. All areas except Area-5 and Area-6 have a daily Planet Fusion time series between 2019 and 2021. On the other hand, Area-5 and Area-6 have only the daily samples for 2019 and 2018, respectively.



**Figure 3.** Location of the six areas in Germany used during the crop classification experiments. Red and Purple tiles are used for testing and the remaining four are used for the training phase.

Crop planting information and field boundaries for these areas were sourced from publicly available user submitted datasets that form the basis of official EU reporting (Integrated Administration and Control System). The datasets were accessed through the region specific data portals[3]. After consideration, we decide to focus on six crop types: winter wheat, winter barley, winter rye, maize, winter oilseed rape, and sugar beet. These crops were well-represented across all tiles in the dataset and with adequate numbers to avoid wildly imbalanced classes. Figure 4 shows the NDVI time series of the selected crop types.
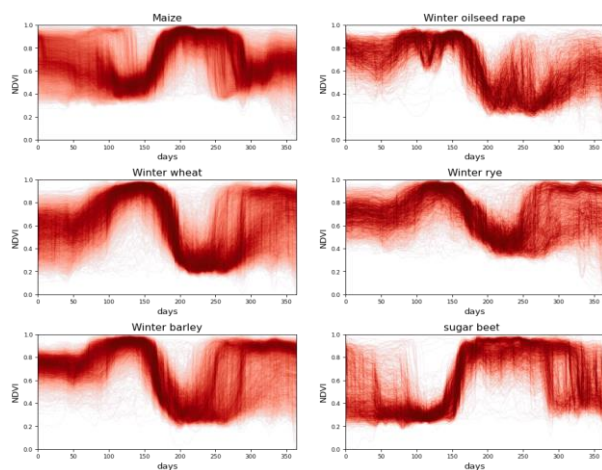


**Figure 4.** Extracted NDVI time series for the selected crop types by using our dataset.

Finally, we split the dataset into training and test sets by considering the class distributions and geolocations of the areas. The Area-3 and Area-6 are used for testing, and the rest are used for training. Ten percent of the training samples are randomly selected for validation purposes. Table 1 summarizes the number of samples for each crop class used for training, validation and testing purposes.

| Crop Type | Training | Validation | Testing |
|---|---|---|---|
| Maize | 11,618 | 1,295 | 1,190 |
| Winter barley | 5,158 | 602 | 1,072 |
| Winter oilseed rape | 1,418 | 155 | 1,218 |
| Winter rye | 2,304 | 275 | 946 |
| Winter wheat | 7,079 | 744 | 4,144 |
| Sugar beet | 1,766 | 190 | 1,160 |

**Table 1.** The number of samples for each crop class.



**Figure 5.** False-color images of a field in our dataset with 15 days cadence.

## 4. EXPERIMENTS

### 4.1 Data Analysis

Before starting to measure the classification performance of different models, we prefer to visually observe the impact of the temporal and spatial variances on crops. For that reason, we analyze the Normalized Difference Vegetation Index (NDVI) time series of selected crops under different scenarios.

First, we aim to focus on temporal impact. Therefore, we focus on a single area, Area4 and plot the NDVI time series of winter barley fields for three sequential years (Figure 6). Although the harvesting time for the three years are close to each other, the earlier phases show significant differences. Similar pattern is observed also for different areas.

---

[2] https://platform.ai4eo.eu/ai4food-security-germany
[3] Brandenburg: https://geobroker.geobasis-bb.de

Lower Saxony:
https://sla.niedersachsen.de/landentwicklung/LEA/
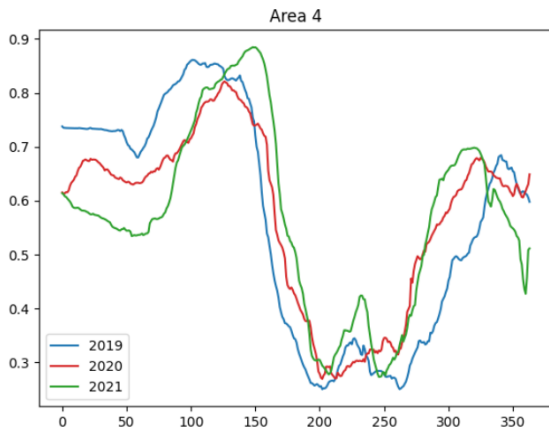North Rhine-Westphalia: https://ckan.open.nrw.de/

**Figure 6.** The average NDVI times series of winter barley in Area 4 for three years

Then we analyze the impact of spatial differences on crops. In Figure 7, we plot the NDVI time series of winter barley fields for a single year (2020) in four different areas. Since we focus on a single country, we believe that the spatial variance is not as problematic as the temporal variance for this dataset. However, when we need to solve temporal and spatial variability simultaneously, the intra-class dissimilarity and inter-class similarity may create some problems for some of the classifiers. Figure 8 displays the NDVI plots of two different cereals, which are obtained from four different areas and three years. A similar issue is observed between sugar beet and maize.
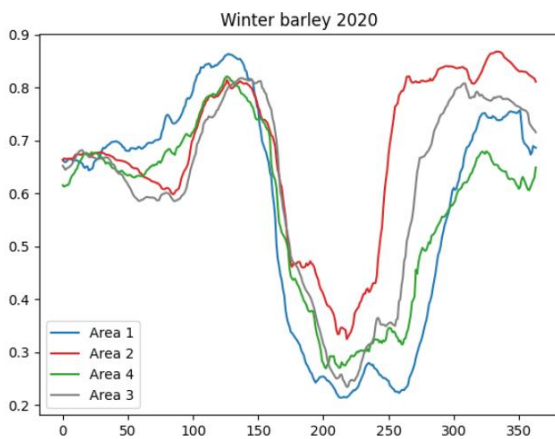


**Figure 8.** The average NDVI times series of winter barley and winter wheat for each different area and different years.



**Figure 7.** The average NDVI times series of winter barley for our 4 areas.

### 4.2 Experiment Methodology

In order to evaluate the robustness of our approach, we design two experimental setups. For each setup, we train our proposed model for ten epochs. Similarly, to observe the impact of SSL, we trained the same encoder without any SSL pretraining for ten epochs. In addition to deep learning-based methods, we train two different Random Forest (RF) classifiers.
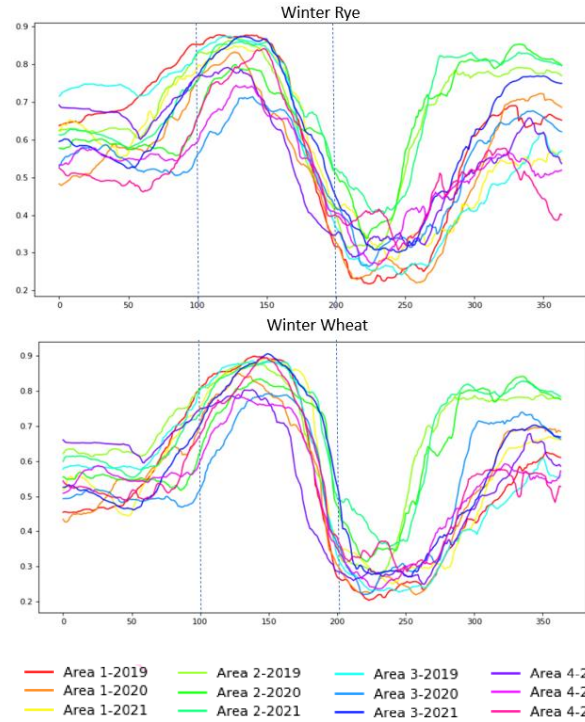
As a feature set of the first RF classifier, we use NDVI, Normalized Difference Water Index (NDWI), and Green-Blue NDVI (GBNDVI) indices. The second RF classifier uses only surface reflectance values obtained from Planet Fusion. We use grid search to optimize the hyper-parameters of both RF classifiers.

To evaluate the effectiveness of the classifiers, we employ three commonly used evaluation metrics: F1-Macro, F1-Micro, and F1-weighted average. F1-Macro calculates the F1 score for each class and takes the average of all the classes. F1-Micro calculates the F1 score by considering the total number of true positives, false negatives, and false positives across all classes. F1-weighted average calculates the F1 score for each class and takes a weighted average of all the classes.

### 4.3 Experiment 1- Evaluate Spatial Robustness

The spatial robustness of a crop classifier is a crucial aspect of its accuracy and reliability. To evaluate the spatial robustness of a crop classifier, it is essential to train and test it in different locations with varying environmental and management conditions. In the first experiment, we aim to understand the spatial robustness of the proposed model. Therefore we use the same years (2019, 2020 and 2021) for training and testing. Since Area-6 has only samples from 2018, we exclude it from our test set. Table 2 summarizes the selected areas for this experiment. In total, we used 29,343 fields during the training and tested on 8,668 fields.

| Tile | Selected Year(s) | Subset |
|------|------------------|--------|
| Area-1 | 2019, 2020, 2021 | Train |
| Area-2 | 2019, 2020, 2021 | Train |
| Area-4 | 2019, 2020, 2021 | Train |
| Area-5 | 2019 | Train |
| Area-3 | 2019, 2020, 2021 | Test |

**Table 2.** The selected areas for the Experiment-1

We compare four models' performances based on metrics described in section 4.2 (Table 3). The proposed model with SSL pretraining achieves the highest performance values. The second best model is the same encoder without any pretraining. We observe that SSL pretraining helped us to distinguish winter wheat and rye crop types in a better way for this experiment. The crop-based precision values of the winter rye are 63% and 91% without and with SSL, respectively.

Interestingly, we observe that RF performs significantly better when we use surface reflectance values instead of indices. The indices-based RF had a lower classification performance, especially for sugar beets and maize classes.

| Model Name | F1-Micro | F1-Macro | F1-Weight. |
|------------|----------|----------|------------|
| Random Forest (indices) | 87.64% | 83.99% | 87.43% |
| Random Forest (SR) | 92.52% | 90.39% | 92.38% |
| Our model without SSL | 93.62% | 92.58% | 93.95% |
| Our model | **96.09%** | **95.41%** | **96.06%** |

**Table 3.** Performance Comparison under spatial variance

### 4.4 Experiment 2- Evaluate Spatial and Temporal Robustness

The second experiment is an extension of the first, and in addition to spatial variability, it involves temporal variability by training and testing samples from different years. The goal is to evaluate the classifiers' ability to handle variations in environmental and management conditions over time. During the experiment, we trained the models using daily PF data of 20,012 fields captured in 2019 and 2020 from Areas 1, 2, 4, and 5. After the training, we evaluated the models on 3,865 fields captured in 2018 and 2021 from Area-6 and Area-3, respectively. Table 4 shows the classification performance of four different models.

In this experiment, we observe that deep-learning models have better performances. Again the proposed model with an SSL pretraining achieves the highest performance values. On the other hand, RF-based classifiers face a significant performance decrease under the spatial and temporal variability between train and test sets. After including the temporal variability, the performance decrease is more than 20 percent for the indices-based RF classifier. When we compare two different RF-based classifiers, we observe that the indices-based one has a lower capability to distinguish cereals from each other (i.e., wheat, barley, and rye). Similarly, it confuses sugar beets and maize classes.

| Model Name | F1-Micro | F1-Macro | F1-Weight. |
|------------|----------|----------|------------|
| Random Forest (indices) | 69.39% | 61.73% | 67.30% |
| Random Forest (SR) | 86.18% | 83.14% | 85.79% |
| Our model without SSL | 94.15% | 93.19% | 94.13% |
| Our model | **95.21%** | **94.77%** | **95.20%** |

**Table 4.** Performance Comparison under temporal and spatial variance

These results are consistent with our preliminary data analysis explained in Section 4.1. It is possible to observe the NDVI similarity within cereals and the NDVI similarity of maize-sugar beet pairs in Figure 4. The generated feature space for indices-based RF is not enough to handle intra-class dissimilarity and inter-class similarity with high accuracy.
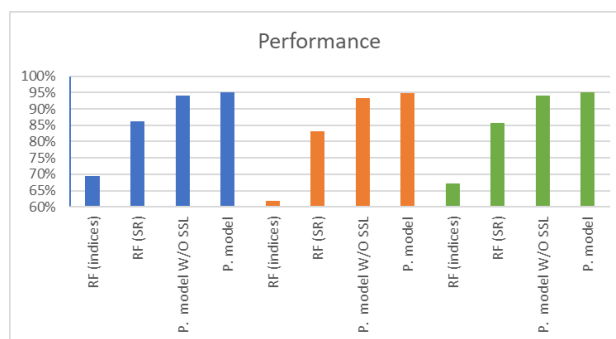


**Figure 8.** The performance Comparison under temporal and spatial variance. The blue, orange and green colors represent F1-Micro, F1-Macro and F1-Weighted respectively.

When we analyze the proposed model, we observe that it achieves very satisfying results for all crop types. Table 5 displays each crop type's class-based precision, recall, and f-score values. Although our dataset has some imbalance issues and the number of sugar beet fields used for training is 10x less than maize fields, the model was able to classify sugar beets with very high accuracy.

| Crop Type | Precision | Recall | F-score |
|-----------|-----------|--------|---------|
| Maize | 97.72% | 98.41% | 98.07% |
| Winter barley | 86.05% | 94.13% | 89.91% |
| Winter oilseed rape | 99.84% | 99.52% | 99.68% |
| Winter rye | 92.11% | 83.60% | 87.65% |
| Winter wheat | 95.16% | 95.29% | 95.23% |
| Sugar beet | 98.63% | 97.55% | 98.09% |

**Table 5.** Crop based classification performance for DL with SSL pretraining under temporal and spatial variance

Winter barley and rye had the lowest classification performances, with 89.91% and 87.65% F-scores, respectively. To understand the reason, we generated the confusion matrix (Table 6), and as expected, we observed that distinguishing between cereal types is relatively more challenging than classifying other crops.

| | | Prediction | | | | | |
|---|---|---|---|---|---|---|---|
| | | Maize | W. Barley | W. Oil. R. | W. Rye | W. Wheat | S. Beets |
| **A** | **Maize** | 558 | 1 | 0 | 3 | 2 | 3 |
| **c** | **W. Barley** | 1 | 401 | 0 | 5 | 18 | 1 |
| **t** | **W. Oil. R.** | 1 | 1 | 623 | 0 | 1 | 0 |
| **u** | **W. Rye** | 1 | 22 | 0 | 362 | 48 | 0 |
| **a** | **W. Wheat** | 4 | 40 | 1 | 22 | 1377 | 1 |
| **l** | **S. Beet** | 6 | 1 | 0 | 0 | 1 | 359 |

**Table 6.** Confusion matrix for DL with SSL pretraining under temporal and spatial variance

## 5. CONCLUSION

In this study, we focus on crop classification, an important research area for remote sensing in agriculture. We observe that the characteristics of crops vary significantly in terms of both time and location. For a reliable crop classification solution, it's essential to satisfy both temporal and spatial robustness. We successfully apply self-supervised learning methods to our transformer-based deep learning techniques to achieve our goal. We test our model under two different scenarios to see the impact of temporal as well as temporal and spatial variability. The results show that our method provides the best results in both scenarios, with a performance difference of over 30 points compared to a traditional random forest classifier based on indices.

Our findings show that:
- Our Deep Learning architecture outperforms RF on our single-nation, multi-year dataset up to 30 percentage points.
- SSL pre-training delivers a performance improvement over standard supervised training when testing on unseen spatial and temporal contexts
- The total performance gain from SSL pre-training is variable. The gain was largest when evaluating spatial generalization, which is also the dimension in which the pre-training dataset expressed the greatest variation. The gain was smaller when evaluating generalization to new years, which may be partly explained by the fact that the pre-training dataset only covered two years.

Based on our results, we would expect SSL pre-training to deliver a more substantial performance improvement over the other methods when:
- limited training labels are available and for only a limited spatial or temporal extent; or
- the unlabeled dataset covers larger regions of time and space that are not covered in the labelled data, but are relevant to the domain.

For our dataset, the deep learning-based method without SSL pretraining already has very high F1-micro and F1-weighted values above 94%. Therefore it is unclear whether SSL pre-training delivers enough of a performance advantage to warrant the additional computational effort required. We believe it is important to validate the performance impact of SSL on more challenging problems.

In future work, we plan to use more labelled samples from around Europe to increase the number of classes and to test for higher spatial variability. As another exciting topic, the high-frequency data from Planet Fusion may be beneficial for early-season crop classification which would be a valuable building-block for real-time food security and supply monitoring across field, regional and national scales .

## 7. REFERENCES

Chen, X., & He, K. (2021). Exploring simple siamese representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15750-15758).

He, T., Xie, C., Liu, Q., Guan, S., & Liu, G. (2019). Evaluation and comparison of random forest and A-LSTM networks for large-scale winter wheat identification. Remote Sensing, 11(14), 1665.

Garnot, V. S. F., & Landrieu, L. (2020, September). Lightweight temporal self-attention for classifying satellite images time series. In International Workshop on Advanced Analytics and Learning on Temporal Data (pp. 171-181). Springer, Cham.

Guo, D., Xia, Y., & Luo, X. (2021). Self-supervised GANs with similarity loss for remote sensing image scene classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 2508-2521.

Marchisio, G., Helber, P., Bischke, B., Davis, T., Senaras, C., Zanaga, D., ... & Wania, A. (2021, July). RapidAI4EO: A Corpus for Higher Spatial and Temporal Reasoning. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 1161-1164). IEEE.

Marszalek, M. L., Saux, B. L., Mathieu, P. P., Nowakowski, A., & Springer, D. (2022). Self-supervised learning--A way to minimize time and effort for precision agriculture?. arXiv preprint arXiv:2204.02100.

Metzger, N., Turkoglu, M.O., D'Aronco, S., Wegner, J.D., Schindler, K. (2021). Crop Classification under Varying Cloud Cover with Neural Ordinary Differential Equations. arXiv preprint arXiv:2012.02542

Saini, R., & Ghosh, S. K. (2018). Crop classification on single date sentinel-2 imagery using random forest and support vector machine. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 683-688.

Seydi, S. T., Amani, M., & Ghorbanian, A. (2022). A Dual Attention Convolutional Neural Network for Crop Classification Using Time-Series Sentinel-2 Imagery. Remote Sensing, 14(3), 498.

Schneider, M., Broszeit, A., & Körner, M. (2021). Eurocrops: A pan-european dataset for time series crop type classification. arXiv preprint arXiv:2106.08151.

Sumbul, G., Charfuelan, M., Demir, B., & Markl, V. (2019, July). Bigearthnet: A large-scale benchmark archive for remote sensing

image understanding. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 5901-5904). IEEE.

Wang, Y., Albrecht, C. M., Braham, N. A. A., Mou, L., & Zhu, X. X. (2022). Self-supervised learning in remote sensing: A review. arXiv preprint arXiv:2206.13188.

Zhao, H., Duan, S., Liu, J., Sun, L., & Reymondin, L. (2021). Evaluation of five deep learning models for crop type mapping using sentinel-2 time series images with missing information. Remote Sensing, 13(14), 2790.

Planet Fusion  2022, https://assets.planet.com/docs/Fusion-Tech-Spec_v1.0.0.pdf