Self-Supervised Learning on Small In-Domain Datasets Can Overcome Supervised Learning in Remote Sensing

Andres J. Sanchez-Fernandez, Sergio Moreno-Álvarez, *Graduate Student Member, IEEE,* Juan A. Rico-Gallego, *Member, IEEE,* and Siham Tabik

***Abstract*—The availability of high-resolution satellite images has accelerated the creation of new datasets designed to tackle broader remote sensing (RS) problems. Although popular tasks like scene classification have received significant attention, the recent release of the Land-1.0 RS dataset (https://doi.org/10.5281/zenodo.7858952) marks the initiation of endeavors to estimate land-use and land-cover (LULC) fraction values per RGB satellite image. This challenging problem involves estimating LULC composition, i.e., the proportion of different LULC classes from satellite imagery, with major applications in environmental monitoring, agricultural/urban planning, and climate change studies. Currently, supervised deep learning models—the state-of-the-art in image classification—require large volumes of labeled training data to provide good generalization. To face the challenges posed by the scarcity of labeled RS data, self-supervised learning (SSL) models have recently emerged, learning directly from unlabeled data by leveraging the underlying structure.**

This is the first paper to investigate the performance of SSL in LULC fraction estimation on RGB satellite patches using in-domain knowledge. We also performed a complementary analysis on LULC scene classification. Specifically, we pretrained Barlow Twins, MoCov2, SimCLR, and SimSiam SSL models with ResNet-18 using the Sentinel2GlobalLULC small RS dataset (https://doi.org/10.5281/zenodo.6941662) and then performed transfer learning to downstream tasks on Land-1.0. Our experiments demonstrate that SSL achieves competitive or slightly better results when trained on a smaller high-quality in-domain dataset of 194,877 samples compared to the supervised model trained on ImageNet-1k with 1,281,167 samples. This outcome highlights the effectiveness of SSL using in-distribution datasets, demonstrating efficient learning with fewer but more relevant data.

***Index Terms*—Remote sensing (RS), deep learning, self-supervised learning (SSL), land use and land cover (LULC), fraction estimation, multispectral data, RGB satellite images.**

I. Introduction

REMOTE sensing (RS) has emerged as an increasingly popular research field within Earth observation [1]. This discipline focuses on acquiring and analyzing data from sensors on platforms such as satellites, airplanes, and drones to study the Earth's surface changes. These sensors remotely measure the electromagnetic radiation reflected and emitted, which is then processed, analyzed, and transformed into different forms of imagery and information. RS offers numerous advantages over traditional ground-based methods by (i) covering large areas at reasonable time intervals, (ii) collecting data in hard-to-reach or hazardous locations, and (iii) providing consistent and repeatable measurements over time, among others [2]. Consequently, RS has become a valuable tool in various fields, including agriculture [3], [4], forestry [5], urban planning [6], natural resource management [7], [8], and environmental monitoring [9].

One of the key applications of RS lies in the classification and description of different types and uses of land surfaces, commonly known as land use and land cover (LULC) [10]. LULC maps are crucial, providing detailed insights into human activities and land characteristics, including residential, agricultural, industrial, and commercial areas, along with natural features such as forests, water bodies, wetlands, grasslands, and urban regions. Accurate and up-to-date LULC information is essential for a broad spectrum of RS applications and decision-making processes [11]–[13].

Supervised deep learning models have demonstrated exceptional performance in processing RS data and performing fundamental computer vision tasks [14]. However, achieving these

Manuscript received April 19, 2021; revised August 16, 2021.

This work was partially supported by the project SmartFoRest (Ref. TED2021-129690B-I00, funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/ PRTR), the Spanish Ministry of Science and Innovation through project PID2019-110315RB-I00 (APRISA), and the European Regional Development Fund (ERDF/FEDER) and by *Junta de Extremadura* (Ref. IB20040). Additionally, this work received funding from the project *Thematic Center on Mountain Ecosystem & Remote sensing, Deep learning-AI e-Services University of Granada-Sierra Nevada* (LifeWatch-2019-10-UGR-01), which has been co-funded by the Ministry of Science and Innovation through the FEDER funds from the Spanish Pluriregional Operational Program 2014-2020 (POPE), LifeWatch-ERIC action line, within the Workpackages LifeWatch-2019-10-UGR-01_WP-8, LifeWatch-2019-10-UGR-01_WP-7 and LifeWatch-2019-10-UGR-01_WP-4.

Authors thankfully acknowledge the computer resources at (i) Turgalium, managed by the Extremadura Research Centre for Advanced Technologies (CETA-CIEMAT); and (ii) Lusitania and the technical support provided by Cénits-COMPUTAEX and Red Española de Supercomputación, RES (QS-2021-1-0027 and QHS-2021-2-0022).

Andres J. Sanchez-Fernandez is with the Department of Computer Systems Engineering and Telematics, School of Technology, University of Extremadura, Av. de la Universidad S/N, 10003, Cáceres, Spain (email: sfandres@unex.es).

Sergio Moreno-Álvarez is with the Department of Computer Languages and Systems, National University of Distance Education, UNED, C. Juan del Rosal 16, 28040, Madrid, Spain (email: smoreno@lsi.uned.es).

Juan A. Rico-Gallego is with CénitS-COMPUTAEX: Extremadura Center for Research, Technological Innovation, and Supercomputing, Carretera Nacional 521 Km 41.8, 10071, Cáceres, Spain (email: juanantonio.rico@cenits.es).

S. Tabik is with the Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence, DaSCI, University of Granada, 18071, Granada, Spain (email: siham@ugr.es).

Code: https://github.com/sfandres/ssl-overcome-supervised-rs/

remarkable results requires extensive and high-quality labeled datasets, which are resource-intensive and time-consuming to create. To mitigate these challenges, self-supervised learning (SSL) has emerged as a promising alternative. In SSL, a deep learning model first learns valuable representations from a large unlabeled dataset by completing an assigned pretext task—a specific learning objective or problem designed to be solved in this initial pretraining phase. These acquired representations then serve as a solid foundation for subsequent supervised tasks [15]. This innovative learning approach has gained considerable traction in computer vision, often producing results that compete with or at times surpass traditional supervised methods [16]. Despite its potential, the adoption of SSL in RS is still limited, primarily focusing on a few datasets and tasks, such as scene classification [16].

The application of SSL approaches for estimating LULC fractions per RGB satellite image remains unexplored. This complex task involves quantifying the composition of LULC classes (e.g., in percentages) within a given RGB satellite patch. Recently, a significant development in this field was the release of the first dataset specifically tailored for this purpose, known as Land-1.0[1]. The availability of this dataset, which provides LULC fraction estimates for individual satellite patches, holds immense potential to advance applications in diverse areas such as environmental monitoring, natural resource management, and land change analysis.

In light of the above discussion, this paper attempts to answer the following question: *Can SSL models pretrained on a small in-distribution dataset outperform popular techniques such as fully supervised learning with random and ImageNet-1k weight initialization in LULC fraction estimation?* To this aim, we assess the transferability of features learned by pretrained SSL models for LULC fraction estimation on RGB satellite patches, particularly under scenarios with limited training sample availability.

Our experimentation is carried out in two phases. Initially, four state-of-the-art SSL models—Barlow Twins, MoCov2, SimCLR, and SimSiam—were pretrained on the Sentinel2GlobalLULC dataset [17] using 175,381 pure-pixel RGB satellite images without labels to acquire transferable representations. Subsequently, the Barlow Twins model, which demonstrated superior performance in the preliminary phase, underwent supervised fine-tuning on the Land-1.0 dataset. This smaller dataset consists of 21,489 mixed-pixel RGB images. In this phase, we systematically tested the model's performance across different subsets of training samples, varying from 5% to 100% of the dataset, to simulate conditions of limited sample availability. Then, Barlow Twins is compared against traditional deep learning approaches, including fully-supervised models initialized with ImageNet-1k or random weights, focusing on LULC fraction estimation. Additionally, we assessed its applicability to scene classification tasks.

The findings from our study underscore the competitiveness of SSL models in both downstream tasks compared to their fully-supervised counterparts. It is worth noting that the in-domain RS dataset used for pre-training the SSL models is approximately seven times smaller than ImageNet-1k, significantly reducing pretraining time. Moreover, when the full dataset is utilized for scene classification, the SSL model exhibits a distinct advantage, outperforming traditional ImageNet-based approaches. These results provide compelling evidence of the potential of SSL in RS applications. The main contributions of this study are summarized below.

1) This study conducts a detailed evaluation of SSL models for LULC fraction estimation on RGB satellite images. It includes pretraining the SSL models using a small in-domain dataset of LULC RGB patches that closely mirrors the distribution of the Land-1.0 downstream dataset, ensuring relevance and applicability.
2) The article provides a thorough comparative analysis between SSL pretrained models and fully-supervised models initialized with ImageNet-1k and random weights. The comparison per training percentages showcases the superiority of SSL models over random initialization and demonstrates their competitiveness with models initialized with ImageNet-1k weights. These results underscore the promising potential of SSL models for LULC fraction estimation and scene classification as efficient learners.
3) The findings offer valuable insights and practical recommendations for practitioners in the RS field, particularly benefiting researchers who are less familiar with SSL approaches. This guidance aims to foster the integration of SSL into existing RS methodologies.

The remainder of the paper is organized as follows: Section II reviews recent publications on the use of SSL models in RS, focusing mainly on their application to RGB satellite images. Section III introduces the core concepts and characteristics of the SSL models considered in this research. Section IV details the methodology used in this study. Section V presents the results derived from the analyses. Section VI offers a comprehensive discussion of these results. Finally, Section VII summarizes the main findings.

II. Related Work

In the domain of satellite data-based applications, SSL has been gaining increasing attention, a trend further amplified by significant achievements in computer vision. Recent advancements and explorations underscore the growing interest in leveraging this convergence. In this context, three predominant research directions have emerged: (i) the creation of extensive RS datasets specifically designed for SSL pretraining, (ii) the evaluation of state-of-the-art SSL models on popular RS datasets, and (iii) the creation of RS-specific SSL architectures. Below, we discuss the most relevant works within each trend.

A. RS-Specific Pretraining Datasets

The work presented in [18] introduces the SSL4EO-S12 dataset, marking a pioneering contribution to the realm of unlabeled RS pretraining datasets. This dataset comprises diverse images from different geographical regions worldwide, leveraging data from the Sentinel-2 L1C, Sentinel-2 L2A, and Sentinel-1 platforms. Notably, it incorporates overlap filtering

[1]Official repository: https://doi.org/10.5281/zenodo.7858952

for image patches, a feature that sets it apart from earlier datasets like SeCo [19]. Purposefully designed to meet the demands of multiple RS applications, the SSL4EO-S12 dataset is tailored to bridge the domain gap that typically challenges conventional ImageNet pretraining practices.

Rigorous empirical evaluations demonstrated the efficacy of SSL4EO-S12, highlighting its superiority over ImageNet and other smaller RS datasets. Specifically, DINO, MAE, and data2vec exhibited superior performance when pretrained on SSL4EO-S12 and tested under linear probing across three distinct downstream tasks: scene classification, semantic segmentation, and change detection. However, several notable limitations of SSL4EO-S12 were identified, including its lack of representation of certain land-cover regions, the inclusion of only medium-resolution images, and the presence of instances with ambiguous geospatial overlap.

B. Evaluating SSL in RS

Within the second research context, the study in [20] conducts a thorough investigation into SSL models, complemented by a series of insightful benchmarking experiments. This study explores the application of generative, predictive, and contrastive methodologies to analyze RS data. A significant strength of this work is its meticulous focus on elucidating the inherent challenges and distinct properties associated with multispectral, hyperspectral, synthetic aperture radar (SAR), and multimodal data. These findings encompass the physical significance of data measurements, the representation of multiple objects within scenes—especially in low-resolution scenarios where a single pixel may encompass multiple object categories, and the temporal variability of the data. Specifically, the study employed pretrained SSL models such as MoCov2, SwAV, SimSiam, and Barlow Twins, using popular RS datasets including BigEarthNet, SEN12MS, and So2Sat-LCZ42. Their evaluation demonstrates that the contrastive negative sampling model, MoCov2, outperforms alternative methods for the classification task on EuroSAT [21]. The study provided valuable insights into the issue of model collapsing, thereby enhancing the understanding of SSL in RS contexts.

Pursuing a similar objective, the authors in [16] conducted a comprehensive evaluation of diverse SSL methods across public scene classification datasets, including RESISC-45 [22], EuroSAT, Aerial Image Dataset (AID) [23], and UC-Merced [24]. Their results showcased the performance of SSL models through a benchmarking process that involved training on incrementally larger portions of the same training dataset.

In the evolving research landscape, the study presented in [25] highlights the capability of SSL methods to capture visual features with higher discriminative power than conventional supervised approaches with ImageNet initialization. This work effectively leverages contrastive strategies to transfer meaningful representations extracted from aerial imagery to RS datasets, significantly improving classification performance for downstream tasks. Additionally, notable improvements are demonstrated through hierarchical pretraining, which involves initial training on natural images followed by subsequent training on RS images [26].

The work in [27] introduces a comprehensive analysis of future trends and challenges associated with SSL methods in RS scenes. By employing SimSiam or MoCov2 models, this research underscores the effectiveness of SSL models in extracting highly informative features for downstream tasks. A major contribution of this study is its emphasis on considering spatial resolution when creating pretraining datasets.

C. Specialized SSL Architectures for RS

RS datasets contrast sharply with popular object-centric datasets like ImageNet, where individual images are independent and lack spatial or temporal relationships. Adjacent satellite images maintain temporal and spatial continuity, providing a unique advantage. These inherent characteristics have been leveraged to enhance accuracy in various RS studies.

A notable example is given in [28], where the authors developed a novel SSL architecture based on the SimCLR framework, specifically tailored for RS applications. Their approach utilizes the spatial continuity in neighboring images by treating them as positive pairs of the anchor image. By averaging the embeddings of these neighboring images, they achieve smoother and more accurate representations. The authors in [29] presented IndexNet, a novel SSL framework designed for RS semantic segmentation, capable of capturing both image-level and pixel-level spatio-temporal representations from large amounts of unlabeled data. Another work proposes the use of MoCov2 and geography-aware contrastive learning, leveraging the spatio-temporal structure of remote sensing data by using temporal positive pairs from spatially aligned images over time and incorporating geo-location information into pretext tasks [30].

Recent research contributions collectively offer valuable and growing evidence supporting the efficacy of SSL methodologies in enriching RS applications, particularly for classification purposes. However, LULC fraction estimation has not been previously explored in this context. By applying SSL methodologies to LULC fraction estimation using small in-domain pretraining datasets, our work pioneers a direct comparison with traditional fully-supervised methods, providing valuable insights for practitioners in the RS field.

III. Background on SSL

SSL is recognized as one of the most promising methodologies for acquiring data representations that exhibit robust generalization across downstream tasks [16], [20], [31]. Unlike supervised learning, SSL eliminates the need for explicit labels but requires understanding the semantic proximity between data samples [32].

In the literature, SSL methods are typically categorized into four groups based on their learning processes [16]: (i) generative, (ii) predictive, (iii) contrastive, and (iv) non-contrastive. The first two categories fall outside the scope of this research and are not addressed here[2]. Instead, this study revisits popular SSL models based on convolutional architectures that belong to the latter two categories (Fig. 1): Barlow Twins [33], MoCov2 [34], SimCLR [15], and SimSiam [35].

[2]A more extensive review on these topics can be found in [31]

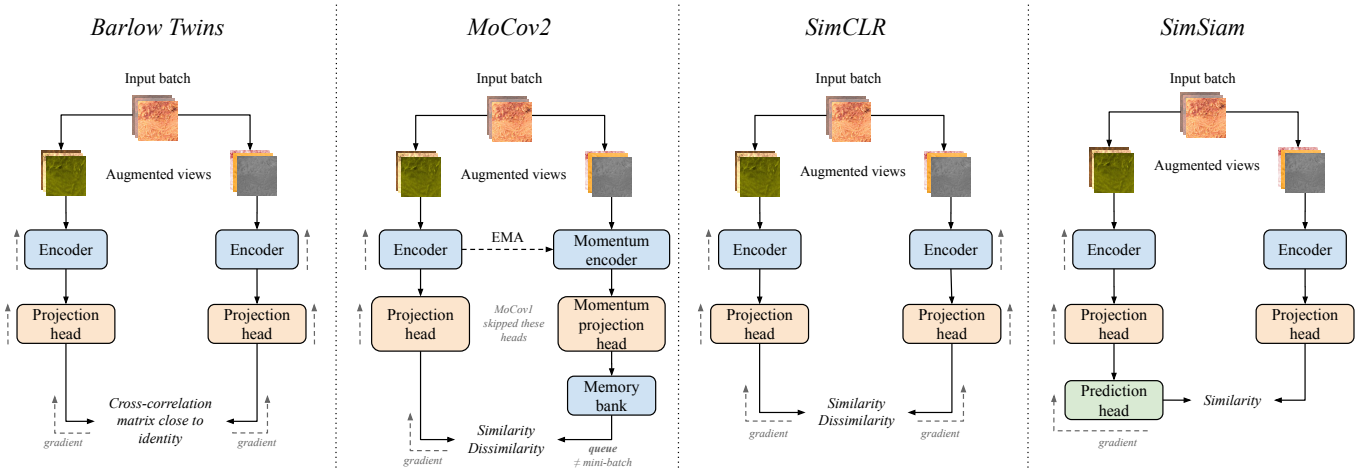


Fig. 1. Main SSL models based on convolutional architectures assessed in this research: Barlow Twins [33], MoCov2 [36], SimCLR [15], and SimSiam [35]. In cases where gradients are not explicitly indicated, gradient stopping is enabled. EMA stands for exponential moving average.

A. Contrastive

The core principle underlying contrastive learning involves training a model to discern similarities and dissimilarities among samples. Essentially, the objective is to maximize the similarity between representations of positive pairs—two augmented versions of the same image—while minimizing the similarity between representations of negative pairs—two variations from different images. This approach aims to improve the ability of the SSL model to encapsulate meaningful and discriminative features within the learned representations.

*1) Simple Framework for Contrastive Learning of Visual Representations (**SimCLR**):* Developed by Google Research's Brain Team in 2020, SimCLR [15] is a widely recognized SSL algorithm that effectively maps semantically similar images to proximal points within the representation space, while pushing apart semantically dissimilar images. This effectiveness derives from four key elements: data augmentation techniques, the base encoder, the projection head, and the contrastive loss function. Among these components, data augmentation plays a pivotal role by exposing the model to diverse perspectives of the same image, thereby facilitating the learning of invariant and informative features. The transformations applied include random cropping, color jittering, Gaussian blurring, and others. During training, a set of data augmentation transforms is randomly sampled from a predefined pool and applied to each image. This process generates two distinct batches that contain different views of the same images, crucial for training the model to recognize semantic similarities under varied conditions. The augmented batches are then fed into the base encoder, which typically employs a ResNet architecture [37], to extract high-level features and generate representation vectors for the images. Following this, a projection head, designed as a compact feedforward neural network, maps the high-dimensional representations produced by the encoder to a lower-dimensional embedding space. Training the projection head to optimize the contrastive loss function enables SimCLR to develop robust representations of input images that are useful for downstream tasks. Notably, the original research implemented the normalized temperature-scaled cross-entropy loss (NT-Xent loss).

*2) Momentum Contrast (**MoCo**):* Introduced by Facebook AI Research (FAIR) in 2020, MoCo is an SSL model that leverages contrastive learning to train deep neural networks and extract visual representations from unlabeled data [36].

Unlike traditional contrastive learning methods limited by the fixed number of negative samples that can be included in each mini-batch, MoCo introduces a more flexible approach: it constructs a large, dynamic dictionary that maintains consistent queries and keys. Positive query-key pairs are formed by aligning keys with augmented views of the query images (dictated by the pretext task), whereas negative pairs represent the opposite scenario. The dictionary operates as a queue that continuously accommodates encoded representations from the most recent mini-batches while removing the oldest ones. This setup allows for a significantly larger pool of samples beyond the mini-batch size, offering a richer diversity of negative samples. To learn effective features, MoCo employs the noise-contrastive estimation (InfoNCE) loss.

Another novel aspect of MoCo is the implementation of a momentum-based update process for the key encoder. This method involves updating the key encoder gradually, using a fraction of its previous state, which ensures consistency across different mini-batches. Such a momentum-based update fosters a smoother evolution of the key encoder, resulting in higher-quality feature representations and mitigating potential inconsistencies.

The optimized version of MoCo, ***MoCov2*** [34], represents a refined release built upon the foundation laid by its predecessor. MoCov2 integrates two key features from SimCLR to enhance its functionality: a 2-layer MLP projection head including a hidden layer of 2048 units, and more robust data augmentation techniques such as Gaussian blur. These enhancements significantly strengthen MoCov2's stability and efficacy within the domain of contrastive learning.

*3) **Barlow Twins**:* This SSL model, released by the FAIR group in 2021, is rooted in the redundancy-reduction principle, originally proposed by H. Barlow [33]. In contrast to con-

ventional SSL methods that depend on negative sample pairs, Barlow Twins adopts a unique approach similar to contrastive learning [31], but without requiring negative samples. As detailed in [38], this unique learning methodology is termed *negative-sample-free contrastive learning*. The essence of this model lies in measuring the cross-correlation matrix between the outputs of two identical networks—each consisting of an encoder and projector head—fed with different augmented versions of the same samples within a batch.

Barlow Twins seeks to minimize the difference between the cross-correlation and identity matrices. This optimization drives the model to increase the similarity of the embedding vectors derived from augmented samples while minimizing redundancy among their components. This strategy allows the model to effectively mitigate the well-known *dimensional collapse* problem without requiring large batches or introducing asymmetry between network twins. This critical problem occurs when the model identifies a trivial solution—producing the same constant value for all input samples—that maximizes similarity but impairs learning. Moreover, this is achieved without resorting to elements such as prediction networks, gradient stopping, or weighted updates in one of the branches.

B. Non-Contrastive

Non-contrastive models are specifically designed to acquire representations from unlabeled data without the need for negative sample pairs. To achieve this, these models leverage pretext tasks such as data augmentation. The primary objective is to enforce the proximity between augmented views of the same image within the latent space. These augmented views offer diverse perspectives, enabling SSL models to capture robust and invariant features. However, it is crucial to highlight that non-contrastive models are particularly susceptible to the problem of dimensional collapse.

*1) Simple Siamese Network (**SimSiam**):* This network, developed by the Facebook AI Research (FAIR) group in 2021, adopts a Siamese network architecture, utilizing two parallel networks, one of which includes a prediction head [35]. SimSiam aims to maximize the similarity between the augmented views of the same image, thus offering an advantage by eliminating the need for negative sample pairs. This feature is particularly beneficial as it addresses hardware constraints and challenges often associated with managing large batches.

Notably, SimSiam operates without the use of momentum encoders, which simplifies its architecture. However, it is susceptible to the collapse problem. To overcome this, SimSiam employs two mitigation approaches. First, a stop-gradient approach is applied to one of the networks, serving as a primary strategy to prevent collapse. Second, a prediction head is added to the other branch to introduce diversity and disrupt symmetry, further stabilizing the learning process.

IV. Methodology

LULC fraction estimation (Table I) from RGB satellite patches plays a critical role across various RS applications due to its extensive utility and practical significance. This process involves determining the proportions of the different

TABLE I
EXAMPLE SHOWCASING THE FRACTION ESTIMATION OF LAND-USE AND LAND-COVER (LULC) CLASSES FROM A SINGLE RGB SATELLITE IMAGE.

Ground-truth fraction (%)	LULC class
68.370	- **Artificial**
9.082	- Annual Croplands
0.000	- Greenhouses
18.411	- Woody Croplands
0.000	- Combinations of Croplands and Natural Vegetation
0.168	- Grasslands and Grasslands with Trees
2.224	- Shrubland and Shrublands with Trees
0.000	- Forests
1.486	- Barelands
0.256	- Wetlands

LULC categories present in the satellite image of a specific area, providing essential insights into environmental changes, ecosystem health, and human activities. For example, in agricultural monitoring, accurately identifying the distribution of various crop types is crucial for forecasting yields and optimizing resource distribution. In the context of urban planning, discerning the ratio between built-up areas and green spaces is key to informed decision-making in infrastructure development and environmental preservation initiatives.

This research assesses the effectiveness of SSL models in performing LULC fraction estimation and, additionally, scene classification. The methodology followed in this study consists of two stages, as illustrated in Fig. 2. Initially, Barlow Twins, MoCov2, SimCLR, and SimSiam undergo pretraining on a small pure-pixel dataset of LULC scenes of RGB satellite patches. The knowledge acquired by the top-performing SSL model is then transferred to a smaller mixed-pixel dataset, where it is compared to traditional supervised approaches. Both phases involve extensive hyperparameter tuning.

A. First Stage: Self-Supervised Pretraining

1) Dataset: The pure-pixel dataset used to pretrain the SSL models is known as *Sentinel2GlobalLULC* [17]. This dataset includes all common LULC classes that cover the entire Earth's surface. It was meticulously crafted to facilitate the creation of regional and global LULC maps to train deep learning models. Sentinel2GlobalLULC contains RGB satellite patches, with each image uniquely representing a single LULC class—commonly referred to as pure pixels. These images are generated by integrating up to 15 global LULC products available in Google Earth Engine, which ensures a robust spatial-temporal consensus across the dataset.

The latest release of this dataset, Sentinel2GlobalLULC v2.1[3], comprises 194,877 single-class RGB image patches categorized into 29 different LULC classes, such as *Barren Lands*, *Moss and Lichen Lands*, and *Grasslands* (as detailed in the left portion of Table II). Each image is a cloud-free composite of 224×224 pixels with a resolution of 10×10 m, derived from the Sentinel-2 satellite imagery captured from June 2015 to October 2020. The dataset provides detailed

[3]Available at https://doi.org/10.5281/zenodo.6941662.

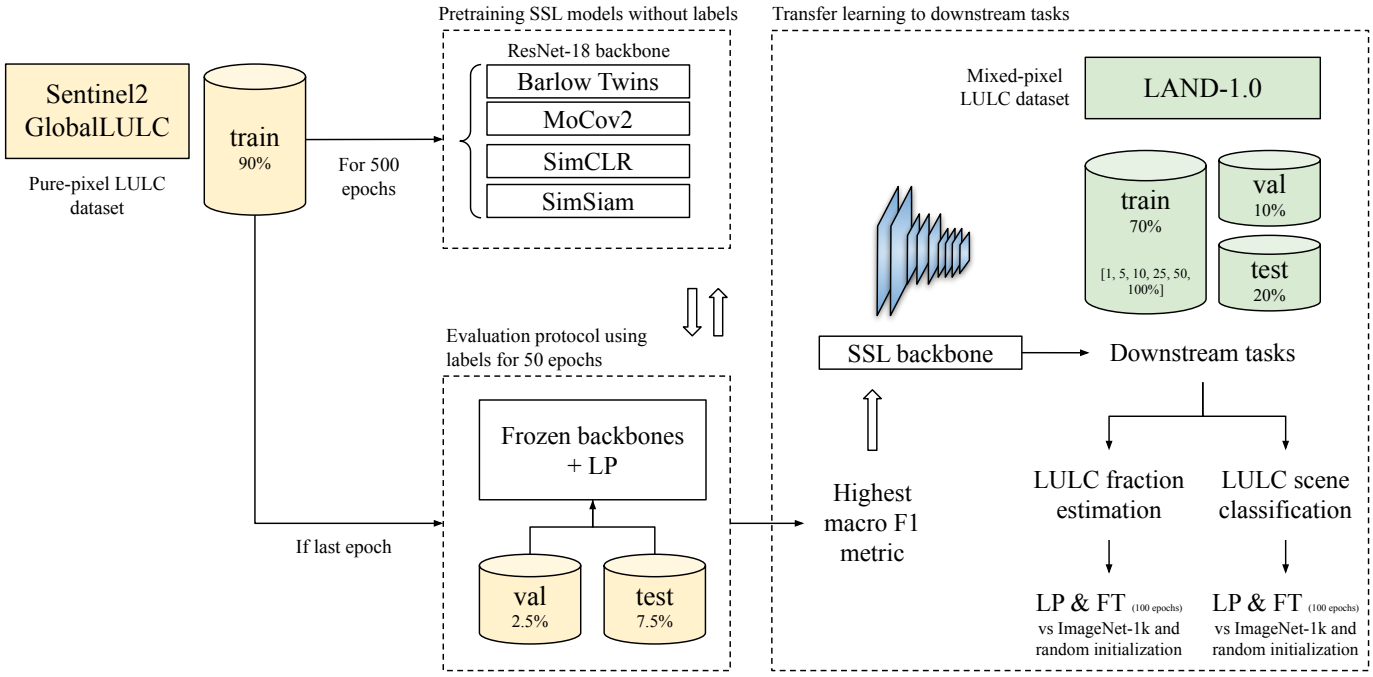


Fig. 2. Workflow of the proposed methodology that initiates with a pretraining phase for each SSL model, lasting 500 epochs, on the training subset of the Sentinel2GlobalLULC pure-pixel dataset [17]. Following pretraining, the quality of feature learning is assessed using linear probing (LP) on validation and testing subsets, with the model backbones' weights frozen. The model that exhibits the highest macro F1 evaluation metric during this phase is selected for knowledge transfer to specific downstream tasks: land-use and land-cover (LULC) fraction estimation and scene classification. Subsequently, different training partitions of the Land-1.0 mixed-pixel dataset [39] are considered to evaluate the SSL models. Evaluation encompasses both linear probing and fine-tuning (FT) techniques, which facilitates the comparison of SSL performance against fully-supervised models initialized with ImageNet-1k and random weights.

TABLE II
OVERVIEW OF THE LAND-USE AND LAND-COVER (LULC) CLASSES AND THEIR RESPECTIVE SAMPLE COUNTS IN THE DATASETS: THE PURE-PIXEL SENTINEL2GLOBALLULC DATASET [17] USED FOR PRETRAINING THE SSL MODELS (LEFT SIDE) AND THE MIXED-PIXEL LAND-1.0 DATASET [39] CONSIDERED IN THE TARGET DOWNSTREAM TASKS (RIGHT SIDE).

Sentinel2GlobalLULC (pure pixels)								LAND-1.0 (mixed pixels)		
ID	**LULC class**	**Samples**	**ID**	**LULC class**	**Samples**	**ID**	**LULC class**	**Samples**	**ID**	**LULC class**	**Samples**
01	Barren Lands	14,000	11	Dense Deciduous Needleleaf Forsts	2,880	21	Marine Water Bodies	14,000	01	Artificial	437
02	Moss and Lichen Lands	4,656	12	Open Evergreen Broadleaf Forests	567	22	Continental Water Bodies	14,000	02	Annual Croplands	3,690
03	Grasslands	8,869	13	Close Evergreen Broadleaf Forests	1,258	23	Permanent Snow	14,000	03	Greenhouses	123
04	Open Shrublands	14,000	14	Dense Evergreen Broadleaf Forests	14,000	24	Croplands Flooded with Seasonal Water	2,004	04	Woody Croplands	5,979
05	Close Shrublands	11,937	15	Open Evergreen Needleleaf Forests	3,914	25	Irrigated Cereal Croplands	842	05	Combinations of Croplands and Natural Vegetation	159
06	Open Deciduous Broadleaf Forests	4,437	16	Close Evergreen Needleleaf Forests	3,872	26	Rainfed Cereal Croplands	1,020	06	Grasslands and Grasslands with Trees	2,448
07	Close Deciduous Broadleaf Forests	1,348	17	Dense Evergreen Needleleaf Forests	13,991	27	Irrigated Broadleaf Croplands	353	07	Shrubland and Shrublands with Trees	6,383
08	Dense Deciduous Broadleaf Forests	14,000	18	Mangrove Wetlands	416	28	Rainfed Broadleaf Croplands	413	08	Forests	1,237
09	Open Deciduous Needleleaf Forests	10,438	19	Swamp Wetlands	487	29	Urban and Built-up Areas	12,590	09	Barelands	702
10	Close Deciduous Needleleaf Forests	6,380	20	Marshland Wetlands	4,205				10	Wetlands	331

LULC annotations for each image and additional metadata. This information includes details about the level of consensus reached in the annotation process, reverse geo-referencing information, the global human modification index, and the number of dates used to create the composite dataset.

2) Hyperparameter Tuning: Multiple configurations are explored for each SSL model, including different learning rates, hidden and output layer sizes, weight decays, and other parameters. Due to hardware limitations, the training set is reduced to 40% using a stratified sampling approach. This approach maintains the original dataset distribution while addressing hardware and time constraints. As a result, extensive experimentation is conducted without compromising the representativeness of the dataset.

Regarding hyperparameter tuning, we use Ray Tune [40], a widely recognized Python library for scalable hyperparameter optimization. This library provides an efficient API that supports tuning deep learning models across a broad spectrum of hyperparameters, algorithms, and frameworks. In this study, Ray Tune was seamlessly integrated with PyTorch to systematically explore the hyperparameter space, enabling the identification of a custom, near-optimal set of hyperparameters for each targeted SSL model.

3) Data Augmentations: In the pretraining process of SSL models, data augmentation proves crucial due to the reliance on unlabeled data. By generating diverse perspectives of the same input data, the model is exposed to a broad spectrum of examples. This exposure aids in acquiring meaningful and robust features that can adapt to various changes.

This research employs the data augmentations used by MoCov2 and SimCLR: random resized cropping, random horizontal and vertical flipping, color jittering (with constant hue), Gaussian blurring, and normalization. Maintaining the hue constant in color transformations is strategic, helping the

model discern color-related discriminative features critical in satellite imagery analysis.

4) Pretraining: Each SSL model is pretrained for 500 epochs using its pseudo-optimal hyperparameter configuration. This strategy aims to balance training effectiveness and the development of meaningful representations, a critical aspect given that most SSL methods require extensive training—typically hundreds of epochs—to achieve effective convergence [41], [42].

The training set comprises 90% of the Sentinel2GlobalLULC dataset, totaling 175,381 unlabeled samples. The remaining parts of the dataset are allocated for offline validation and testing, with 2.5% (4,861 samples) and 7.5% (14,635 samples), respectively. These subsets are used to assess the quality of the backbone weights upon completing pretraining.

During the evaluation phase, the backbone weights are duplicated and frozen, with a classification head attached on top. Subsequently, linear probing is conducted over 50 epochs, using the macro F1 metric to evaluate the effectiveness of the learned features, particularly important given the imbalanced nature of the dataset. The top-performing SSL model is selected based on achieving the highest macro F1 score.

B. Second Stage: Transfer Learning to Downstream Tasks

1) Downstream Tasks: The weights obtained from the pretraining phase are transferred to evaluate their efficacy in real-world RS applications. Specifically, the primary task is LULC fraction estimation using RGB satellite images. Additionally, scene classification is conducted as a supplementary analysis, where the most prevalent class in each image is designated as its label.

2) Dataset: The target dataset, known as *Land-1.0*[4] [39], comprises 21,489 mixed-pixel RGB satellite patches from Andalusia—the southern region of Spain. Each sample in the dataset is associated with fraction values that represent its composition, encompassing ten distinct classes such as *Artificial*, *Annual Croplands*, and *Greenhouses* (as detailed in the right portion of Table II). For practical application, the Land-1.0 dataset is split into three subsets: train, validation, and test, containing 15,038 (70%), 2,153 (10%), and 4,298 (20%) samples, respectively. To simulate a scenario with limited samples, the training subset is further subdivided into varying percentages, ranging from 5 to 100. This subdivision preserves the original class distribution and retains the inherent imbalance found in the dataset. The main objective of this segmentation is to assess the model's generalization capabilities relative to the proportion of labeled training samples available.

3) Weight Initialization Techniques: For each training data subset, we evaluate the performance of three fully-supervised learning methods, distinguished by different weight initialization of the backbone network; the attached classification head is always randomly initialized. The initialization methods include (i) *random weights*, (ii) *ImageNet-1k weights*, and (iii) *pretrained weights* from the top-performing SSL model on the Sentinel2GlobalLULC dataset.

[4]Available at https://doi.org/10.5281/zenodo.7858952.

The use of random weights serves as the foundational initialization method, providing a baseline for comparison and enabling the assessment of the impact of alternative strategies. Following this, we explore the use of ImageNet-1k weights, which are well-established and widely available, having been trained in a supervised manner on a comprehensive image classification dataset [43]. These weights encapsulate general features and patterns, making them suitable for transfer learning scenarios. Finally, we investigate the benefits of using SSL weights pretrained on Sentinel2GlobalLULC. This dataset, specifically tailored for LULC classification of pure-pixel RGB images, offers domain-specific knowledge that could enhance performance in LULC fraction estimation tasks, as tested using the Land-1.0 dataset.

By evaluating these different weight initialization methods across varying amounts of labeled data, we derive insights into their respective impacts and the advantages of integrating domain-specific pretraining for LULC fraction estimation.

4) Transfer Learning Methods: To identify the most suitable model for the specified downstream task, we explored two distinct transfer learning strategies: linear probing/evaluation and fine-tuning. In linear probing, the initialized weights of the layers are kept fixed (i.e., frozen), and only the classification head is trained. This approach makes it possible to evaluate the discriminative capability of the initialized features. In contrast, fine-tuning involves updating the weights throughout the entire network using task-specific data. This comprehensive approach is designed to refine the network's ability to extract features directly relevant to the target downstream task.

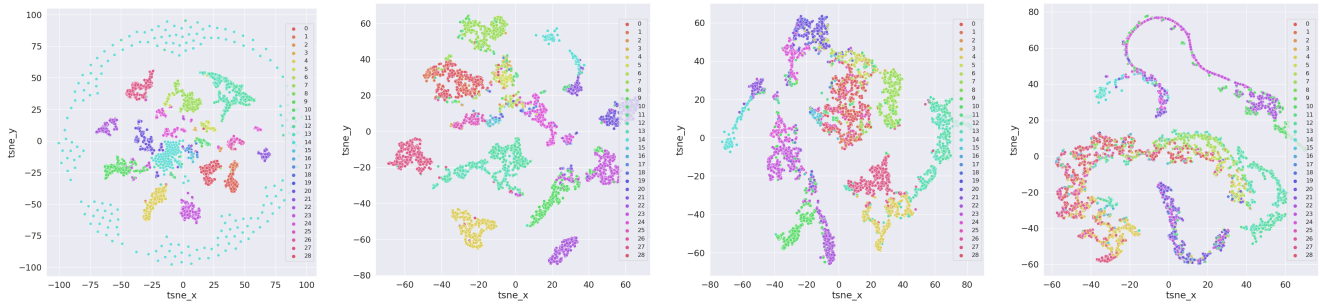
V. Experimental Results

A. Experimental Setup

1) Initial Settings: Training and hyperparameter tuning for each model takes place on the computing nodes of the Turgalium cluster, part of the CETA-Ciemat high-performance infrastructure. Each node within this cluster has two to four NVIDIA Tesla V100 GPUs. Every experiment shows the results after averaging five independent runs, each with a different seed. This approach ensures the results are reliable and not dependent on the randomness of a single seed, providing a more robust evaluation.

The SSL models detailed in Section III were implemented using the *Lightly* open-source computer vision framework, which is specifically designed for SSL applications [44]. This Python library offers a comprehensive and user-friendly framework that streamlines the implementation and evaluation of SSL models. To optimize memory usage and reduce training time, the ResNet-18 architecture [37] was employed.

In terms of the training setup, each model is trained on a single GPU, with configurations optimized as detailed below in Section V-B. Despite variations in specific settings, several hyperparameters remain consistent across all models. During pretraining, a batch size of 512 is used, along with the SGD optimizer. The learning rate is initially increased following a linear warm-up strategy for the first 10 epochs, then transitions to a cosine decay schedule, continuing without restarts. Additionally, MoCov2 implements a momentum value



(a) Barlow Twins (b) MoCov2 (c) SimCLR (d) SimSiam

Fig. 3. t-SNE visualization comparing embeddings generated by the studied SSL methods using the ResNet-18 backbone on the Sentinel2GlobalLULC dataset. These embeddings illustrate the feature representations learned by the models. Notably, Barlow Twins showcases superior clustering performance. Each color in the visualization represents one of the 29 distinct classes found in the Sentinel2GlobalLULC dataset.

TABLE III
PSEUDO-OPTIMAL HYPERPARAMETER VALUES OBTAINED FROM THE BEST TRIAL CONDUCTED WITH THE RAY TUNE TOOL AFTER EXPLORING THE HYPERPARAMETER SPACE. RESNET-18 WAS CHOSEN AS THE BACKBONE ARCHITECTURE.

Model	Hidden dimension	Learning rate	Momentum	Output dimension	Weight decay
Barlow Twins	256	0.0001	0.9	128	10^{-5}
MoCov2	128	0.01	0.9	256	10^{-5}
SimCLR	256	0.1	0.9	512	10^{-4}
SimSiam	128	0.1	0.9	128	10^{-5}

of 0.999, consistent with the practices outlined in the original research [36]. For downstream tasks, the batch size is reduced to 64, while the same optimizer is used.

2) Evaluation Metrics: The quality of predictions in LULC fraction estimation is assessed using two main metrics: overall Root Mean Squared Error (RMSE), which measures the average variance between predicted and actual values across all classes, and per-class RMSE, which calculates the RMSE for each class. In both cases, lower values indicate better performance. For scene classification, the primary evaluation metric is the macro-averaged F1 score, which averages the F1 scores across all classes and is particularly useful for unbalanced datasets. Additional metrics include micro F1, weighted F1, and per-class F1 scores, where higher scores reflect a better balance between precision and recall.

B. Self-Supervised Pretraining: Model Selection

Before pretraining the SSL models, we used the grid-search algorithm provided by Ray Tune to test hyperparameter combinations. This analysis covered different learning rates ranging from 10^{-4} to $0.1 \cdot \sqrt{bsz}$, where bsz stands for batch size; hidden layer sizes and output dimensions (128, 256, 512); momentum values (0.9, 0.99); and weight decays (0, 10^{-4}, 10^{-5}). The ASHA scheduler was used to determine the optimal configurations, aiming to minimize within a constrained timeframe of 12 epochs. Moreover, early stopping was implemented with a grace period of four epochs. The optimal hyperparameters for each model are detailed in Table III.

The selected SSL models were pretrained on the Sentinel2GlobalLULC training subset for 500 epochs. Subsequent evaluations were performed through linear probing on the test subset over 50 epochs, where the backbones were kept frozen to assess the robustness of learned features.

TABLE IV
F1 SCORES OBTAINED FOR EACH SSL MODEL OBTAINED FOLLOWING THE EVALUATION PROTOCOL CONDUCTED AFTER THE FINAL PRETRAINING EPOCH ON THE TEST SUBSET OF THE SENTINEL2GLOBALLULC DATASET. THE HIGHEST SCORES ARE HIGHLIGHTED IN BOLD.

Model	Micro F1	Macro F1	Weighted F1
Barlow Twins	**0.913**	**0.784**	**0.909**
MoCov2	0.715	0.356	0.650
SimCLR	0.521	0.205	0.419
SimSiam	0.204	0.037	0.077

For illustrative purposes, Fig. 3 shows the embeddings from the pretrained SSL models on the Sentinel2GlobalLULC test subset. The embeddings demonstrate that Barlow Twins produces the most structured and discriminative feature space. The well-defined and compact clusters suggest that Barlow Twins effectively reduces redundancy and learns meaningful features. MoCov2 and SimCLR perform moderately, with MoCov2 showing relatively better clustering than SimCLR, but neither achieves the same level of quality as Barlow Twins. Lastly, SimSiam's embeddings are widely scattered and overlapping, indicating its difficulty in learning useful representations. The embeddings suggest that the model might have learned collapsed or redundant representations.

The above results are also evidenced in Table IV, where Barlow Twins achieves the highest accuracy scores. Consequently, the backbone of the Barlow Twins model was selected for further transferring learning to downstream tasks.

C. Performance Analysis

The pseudo-optimal hyperparameters for the fully-supervised models—whether initialized with weights from pretrained SSL, ImageNet-1k, or random—are determined

TABLE V
PER-CLASS AND OVERALL RMSE RESULTS OBTAINED FOR LULC FRACTION ESTIMATION ON THE LAND-1.0 DATASET, AVERAGED OVER FIVE RUNS. THE EVALUATION WAS PERFORMED AFTER FINE-TUNING (FT) THE SSL BARLOW TWINS MODEL AND THE FULLY-SUPERVISED (FS) MODELS INITIALIZED WITH IMAGENET-1K (IN) AND RANDOM (R) WEIGHTS. ALL MODELS WERE TRAINED FOR 100 EPOCHS. TR STANDS FOR TRAIN RATIO AND THE CLASSES RANGE FROM 0 TO 9. THE BEST RMSE VALUES FOR EACH TRAIN RATIO ARE HIGHLIGHTED IN BOLD.

TR (%)	Model	RMSE	RMSE per class									
			0	1	2	3	4	5	6	7	8	9
5	*FS-IN-FT*	**0.091±0.001**	**0.052±0.002**	**0.124±0.006**	0.030±0.005	**0.160±0.003**	**0.040±0.001**	**0.115±0.004**	**0.147±0.003**	**0.085±0.002**	**0.094±0.002**	0.064±0.003
	FS-R-FT	0.114±0.001	0.069±0.002	0.165±0.004	0.030±0.001	0.232±0.003	0.047±0.001	0.134±0.002	0.182±0.005	0.092±0.002	0.112±0.006	0.076±0.003
	SSL-BT-FT	0.096±0.001	0.055±0.001	0.135±0.003	**0.028±0.002**	0.177±0.003	0.045±0.001	0.122±0.001	0.151±0.001	0.088±0.002	0.099±0.002	**0.063±0.002**
10	*FS-IN-FT*	**0.084±0.001**	**0.050±0.002**	**0.107±0.002**	**0.023±0.003**	**0.141±0.003**	**0.038±0.002**	**0.110±0.002**	**0.140±0.003**	0.083±0.003	**0.091±0.002**	**0.055±0.003**
	FS-R-FT	0.109±0.002	0.068±0.009	0.155±0.005	0.028±0.000	0.223±0.005	0.048±0.002	0.129±0.002	0.179±0.011	0.089±0.002	0.105±0.008	0.069±0.003
	SSL-BT-FT	0.090±0.002	0.053±0.002	0.124±0.006	0.025±0.001	0.164±0.004	0.043±0.001	0.117±0.004	0.144±0.002	**0.083±0.002**	0.093±0.001	0.056±0.003
25	*FS-IN-FT*	**0.076±0.001**	**0.043±0.001**	**0.095±0.001**	**0.019±0.002**	**0.126±0.004**	**0.037±0.001**	**0.103±0.004**	**0.126±0.000**	**0.076±0.001**	**0.087±0.002**	**0.049±0.002**
	FS-R-FT	0.100±0.002	0.055±0.002	0.144±0.013	0.026±0.002	0.200±0.006	0.043±0.001	0.122±0.002	0.163±0.004	0.086±0.000	0.097±0.003	0.062±0.002
	SSL-BT-FT	0.082±0.001	0.047±0.001	0.112±0.002	0.023±0.003	0.143±0.002	0.041±0.001	0.109±0.001	0.133±0.004	0.078±0.001	0.089±0.002	0.050±0.002
50	*FS-IN-FT*	**0.071±0.001**	**0.041±0.001**	**0.089±0.002**	**0.017±0.003**	**0.112±0.002**	**0.035±0.001**	0.101±0.005	**0.119±0.002**	0.074±0.002	**0.083±0.001**	**0.041±0.003**
	FS-R-FT	0.091±0.001	0.051±0.002	0.125±0.003	0.022±0.002	0.174±0.003	0.042±0.001	0.117±0.002	0.150±0.003	0.081±0.001	0.091±0.002	0.056±0.002
	SSL-BT-FT	0.075±0.001	0.044±0.002	0.095±0.001	0.021±0.002	0.126±0.002	0.037±0.001	**0.100±0.001**	0.122±0.001	**0.072±0.001**	0.085±0.002	0.044±0.001
100	*FS-IN-FT*	**0.067±0.001**	**0.038±0.002**	**0.081±0.002**	**0.015±0.001**	**0.101±0.000**	**0.035±0.001**	**0.094±0.001**	**0.113±0.001**	**0.069±0.001**	**0.080±0.000**	0.039±0.001
	FS-R-FT	0.081±0.001	0.046±0.002	0.104±0.002	0.020±0.003	0.143±0.004	0.038±0.001	0.108±0.002	0.138±0.002	0.079±0.001	0.086±0.001	0.045±0.001
	SSL-BT-FT	0.070±0.001	0.041±0.002	0.090±0.001	0.016±0.001	0.117±0.000	0.037±0.001	0.098±0.000	0.117±0.002	0.071±0.001	0.082±0.001	**0.038±0.001**

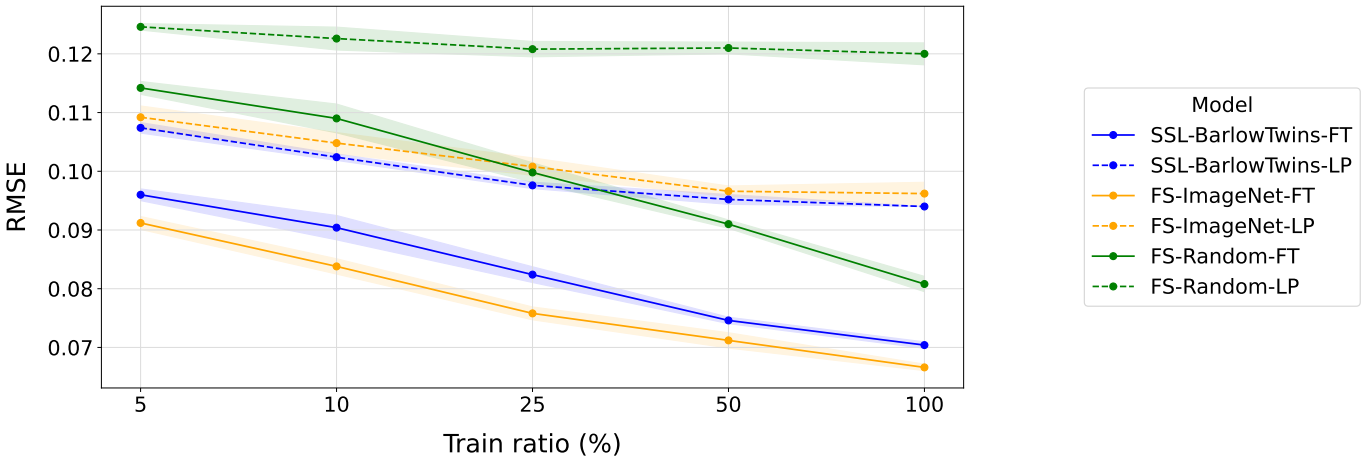
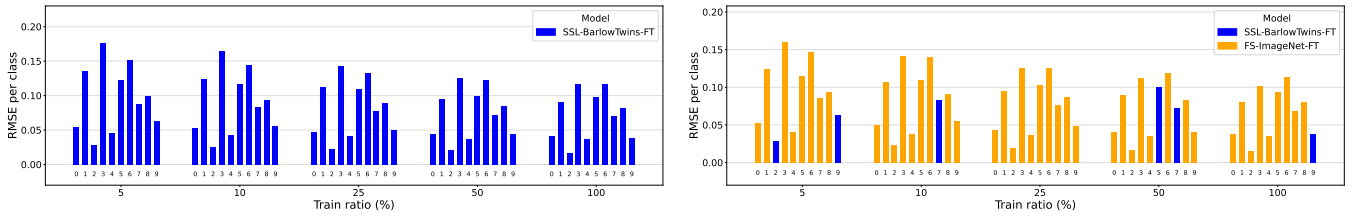


Fig. 4. RMSE results obtained for LULC fraction estimation on Land-1.0 using linear probing (LP) and fine-tuning (FT). The ResNet-18 feature extractor was pretrained under two paradigms: SSL using the Barlow Twins architecture (SSL-BarlowTwins-LP/FT) and fully-supervised learning initialized with ImageNet-1k (FS-ImageNet-LP/FT) and random weights (FS-Random-LP/FT). Lower RMSE values indicate better performance. The shaded area around each line represents the standard deviation calculated from five independent runs with different seeds.



(a) Comparison with baseline (b) Comparison with FS-ImageNet-FT

Fig. 5. RMSE per class and train ratio obtained for LULC fraction estimation using the Land-1.0 dataset after fine-tuning (FT) the models for 100 epochs. To enhance clarity, the left side showcases the comparison between the self-supervised learning (SSL) Barlow Twins model and the fully-supervised (FS) model initialized with random weights (baseline). The right side displays the SSL Barlow Twins model against the FS model initialized with ImageNet-1k weights. The per-class metrics are provided for each train ratio, with the downstream dataset comprising nine classes. Each bar represents the best performance achieved. Lower RMSE values indicate superior performance.

using Ray Tune. This process employs the same grid search space as previously used (excluding layer sizes that are irrelevant in this context) and spans 100 epochs with a grace period of 75 epochs.

To maintain simplicity, the naming convention of these models adopts the structure *X*-*Y*-*Z*. Here, *X* denotes the training method: fully-supervised (FS) or self-supervised learning (SSL[5]). Subsequently, *Y* represents the initialization weights: Barlow Twins, ImageNet, or Random. Lastly, *Z* corresponds to the transfer learning algorithm: linear probing (LP) or fine-tuning (FT). For instance, FS-ImageNet-FT represents a fully-

[5]The SSL model undergoes fully-supervised training on the downstream dataset in the same way as the other models. However, for simplicity, the abbreviation SSL is used instead of SSL-FS.

TABLE VI

PER-CLASS AND OVERALL F1 SCORES OBTAINED FOR LULC SCENE CLASSIFICATION ON THE LAND-1.0 DATASET, AVERAGED OVER FIVE RUNS. THE EVALUATION WAS PERFORMED AFTER FINE-TUNING (FT) THE SELF-SUPERVISED LEARNING (SSL) BARLOW TWINS MODEL AND THE FULLY-SUPERVISED (FS) MODELS INITIALIZED WITH IMAGENET-1K (IN) AND RANDOM (R) WEIGHTS. ALL MODELS WERE TRAINED FOR 100 EPOCHS. TR STANDS FOR TRAIN RATIO AND THE CLASSES RANGE FROM 0 TO 9. THE BEST F1 SCORES PER TRAIN RATIO ARE HIGHLIGHTED IN BOLD.

TR (%)	Model	Micro F1	Macro F1	Weighted F1	F1 per class									
					0	1	2	3	4	5	6	7	8	9
5	*FS-IN-FT*	**0.689±0.011**	0.588±0.016	**0.683±0.012**	**0.629±0.026**	0.713±0.024	0.650±0.040	**0.731±0.015**	**0.352±0.150**	**0.598±0.024**	**0.736±0.013**	0.490±0.057	0.456±0.029	0.527±0.066
	FS-R-FT	0.551±0.016	0.437±0.012	0.542±0.014	0.208±0.062	0.583±0.011	0.684±0.050	0.568±0.021	0.166±0.061	0.473±0.036	0.605±0.028	0.465±0.077	0.261±0.048	0.362±0.064
	SSL-BT-FT	0.677±0.007	**0.591±0.015**	0.673±0.007	0.602±0.047	**0.720±0.008**	**0.738±0.022**	0.722±0.015	0.274±0.121	0.566±0.028	0.705±0.026	**0.541±0.032**	**0.464±0.020**	**0.580±0.038**
10	*FS-IN-FT*	**0.714±0.009**	**0.630±0.019**	**0.709±0.010**	**0.652±0.027**	**0.751±0.009**	**0.771±0.035**	**0.753±0.012**	**0.391±0.105**	**0.619±0.018**	**0.750±0.011**	**0.563±0.031**	0.472±0.037	0.576±0.096
	FS-R-FT	0.576±0.007	0.462±0.026	0.565±0.007	0.293±0.147	0.609±0.011	0.680±0.088	0.597±0.021	0.118±0.101	0.461±0.028	0.630±0.020	0.457±0.103	0.320±0.045	0.458±0.077
	SSL-BT-FT	0.696±0.004	0.613±0.018	0.691±0.005	0.625±0.049	0.731±0.011	0.737±0.065	0.732±0.015	0.294±0.096	0.602±0.014	0.730±0.008	0.555±0.025	**0.482±0.030**	**0.637±0.032**
25	*FS-IN-FT*	**0.748±0.004**	**0.664±0.005**	**0.746±0.004**	0.678±0.040	**0.803±0.002**	**0.764±0.038**	**0.800±0.004**	0.365±0.087	**0.654±0.009**	**0.763±0.011**	**0.596±0.019**	**0.555±0.008**	**0.669±0.038**
	FS-R-FT	0.618±0.009	0.531±0.013	0.613±0.007	0.481±0.095	0.652±0.020	0.704±0.044	0.637±0.009	0.206±0.107	0.542±0.017	0.660±0.027	0.508±0.046	0.395±0.045	0.528±0.046
	SSL-BT-FT	0.732±0.008	0.655±0.013	0.730±0.009	**0.683±0.023**	0.770±0.014	0.732±0.075	0.779±0.017	**0.407±0.038**	0.629±0.011	0.761±0.002	0.589±0.042	0.547±0.010	0.649±0.049
50	*FS-IN-FT*	**0.762±0.003**	**0.697±0.008**	**0.760±0.003**	**0.687±0.036**	**0.814±0.008**	**0.782±0.030**	**0.814±0.006**	**0.518±0.065**	**0.662±0.002**	**0.780±0.008**	0.609±0.018	0.543±0.028	**0.758±0.024**
	FS-R-FT	0.652±0.042	0.580±0.021	0.648±0.044	0.570±0.077	0.709±0.037	0.758±0.043	0.675±0.073	0.205±0.156	0.579±0.064	0.658±0.048	0.610±0.047	0.473±0.011	0.565±0.049
	SSL-BT-FT	0.743±0.004	0.666±0.017	0.740±0.005	0.666±0.032	0.783±0.006	0.764±0.063	0.793±0.007	0.391±0.116	0.626±0.050	0.768±0.011	**0.630±0.027**	**0.544±0.028**	0.701±0.009
100	*FS-IN-FT*	0.750±0.014	0.684±0.017	0.749±0.012	0.651±0.031	0.806±0.007	0.759±0.093	0.807±0.020	**0.509±0.052**	0.637±0.029	0.763±0.023	**0.638±0.027**	0.548±0.045	0.717±0.046
	FS-R-FT	0.714±0.006	0.640±0.012	0.711±0.004	0.644±0.030	0.749±0.011	0.760±0.041	0.740±0.012	0.334±0.052	0.649±0.008	0.740±0.013	0.628±0.017	0.510±0.027	0.649±0.071
	SSL-BT-FT	**0.770±0.003**	**0.689±0.014**	**0.767±0.004**	**0.706±0.011**	**0.811±0.019**	**0.759±0.037**	**0.822±0.004**	0.411±0.083	**0.672±0.021**	**0.791±0.011**	0.614±0.050	**0.589±0.017**	**0.719±0.015**

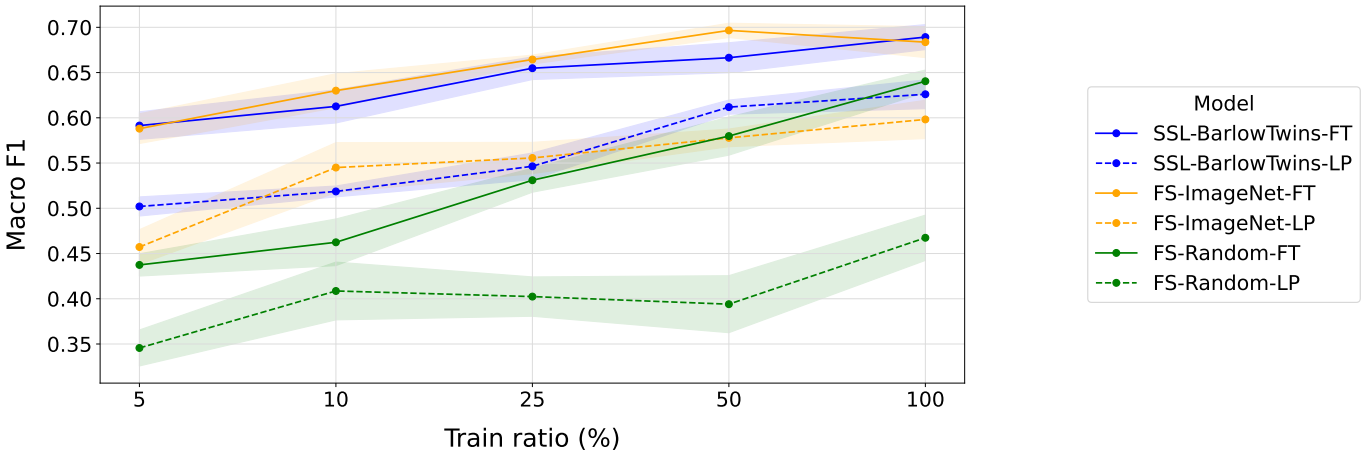
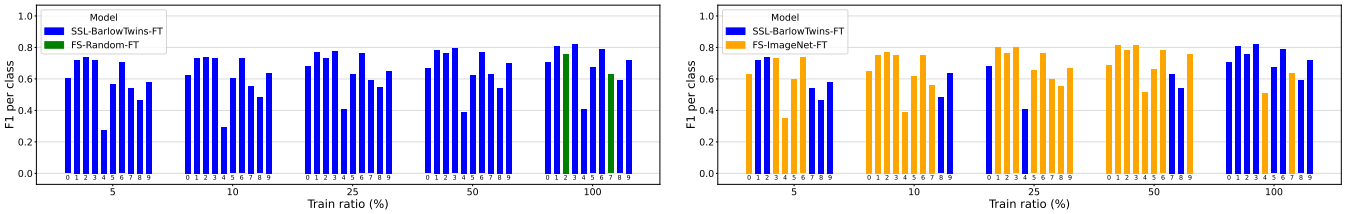


Fig. 6. Macro F1 scores obtained for LULC scene classification on Land-1.0 using linear probing (LP) and fine-tuning (FT). The ResNet-18 feature extractor was pretrained under two paradigms: self-supervised learning (SSL) using the Barlow Twins architecture (SSL-BarlowTwins-LP/FT) and fully-supervised learning initialized with ImageNet-1k (FS-ImageNet-LP/FT) and random weights (FS-Random-LP/FT). Higher F1 scores indicate superior performance. The shaded area around each line represents the standard deviation calculated from five independent runs with different seeds



(a) Comparison with baseline (b) Comparison with FS-ImageNet-FT

Fig. 7. F1 score per class and train ratio obtained for the LULC scene classification challenge on the Land-1.0 dataset after fine-tuning (FT) the models for 100 epochs. To enhance clarity, the left side showcases the comparison between the self-supervised learning (SSL) Barlow Twins model and the fully-supervised (FS) model initialized with random weights (baseline). The right side displays the SSL Barlow Twins model against the FS model initialized with ImageNet-1k weights. The per-class metrics are provided for each train ratio, with the downstream dataset comprising nine classes. Each bar represents the best performance achieved. Higher F1 scores indicate superior performance.

supervised model initialized with ImageNet-1k weights and fine-tuned for the target downstream task. When only the first two components are mentioned, such as FS-ImageNet, it refers to the general performance of the fully-supervised model, regardless of the transfer learning protocol.

1) LULC Fraction Estimation: We evaluated the influence of varying training ratios on the performance of the deep learning models for the LULC fraction estimation classification task (Table V). To simulate real-world scenarios with limited data availability, we sampled images from the original Land-1.0 dataset to create different training subsets. The complete training set consists of 15,038 samples, from which we employed stratified sampling to preserve the original distribution across subsets. We considered subset sizes of 5%, 10%, 25%, 50%, and 100% of the initial training dataset to understand the impact of data availability on model performance. The validation and test sets, comprising 2,153 and 4,298 samples, respectively, remained consistent throughout the experiments

to ensure a fair comparison of results.

Fig. 4 displays the RMSE values achieved by each target model across different training ratios and transfer learning algorithms for the LULC fraction estimation task.

In general, under the less demanding LP scenario, SSL-BarlowTwins-LP achieves the best results compared to FS-ImageNet-LP and FS-Random-LP. In contrast, in the most demanding TF scenario, both SSL-BarlowTwins-TF and FS-ImageNet-TF achieve highly competitive results, consistently outperforming the baseline FS-Random-FT across all train ratios and transfer learning protocols. The best results were observed under the FT scenario. This indicates that weights pretrained on an in-distribution dataset, such as Sentinel2GlobalLULC, still require FT on the downstream Land-1.0 dataset. Notably, SSL-BarlowTwins-LP outperforms FS-Random-FT for smaller train ratios, even though the latter has undergone FT on the same target downstream dataset.

Based on average values, SSL-BarlowTwins-LP outperforms the FS-Random-LP model by 18.85% (0.099 vs. 0.122) under the LP setting. Similarly, in FT, it shows an improvement of 16.16% (0.083 vs. 0.099). When comparing SSL-BarlowTwins with FS-ImageNet, the performance is generally close. Specifically, *SSL-BarlowTwins* performs better in LP by 2.94% (0.099 vs. 0.102) but is slightly worse in FT by 6.41% (0.083 vs. 0.078).

To gain a deeper understanding of these differences, pairwise comparisons were conducted between the three fully-supervised models, considering the per-class results and employing the FT transfer learning method. The FT procedure, which is more powerful than LP, facilitates unlocking the full potential of each model. By adjusting all the weights in the network, FT enables the extraction of more relevant and task-specific features, capturing intricate patterns and nuances that LP might overlook.

Fig. 5a compares the performance of SSL-BarlowTwins-FT and FS-Random-FT models in the context of LULC fraction estimation. The former consistently achieves superior results across all classes and training ratios, spanning 60 different scenarios. This underscores the effectiveness of the SSL pretrained weights. In contrast, Fig. 5b provides a second pairwise comparison, showing that FS-ImageNet-FT model outperforms SSL-BarlowTwins-FT in 54 out of 60 scenarios.

2) Supplementary Test on LULC Scene Classification: Scene classification serves as an additional experiment, complementing the results presented above on LULC fraction estimation. These results are detailed in Table VI.

Fig. 6 presents the macro F1 scores. In the LP setting, the SSL-BarlowTwins-LP model significantly outperforms the FS-Random-LP across all training ratios. Moreover, it surpasses the FS-ImageNet-LP model in more than half of the training ratios studied. In FT, the SSL-BarlowTwins-FT and FS-ImageNet-FT models perform similarly, and both outperform the FS-Random-FT model. Notably, the SSL-BarlowTwins-LP model achieves a higher F1 score than the FS-Random-FT model for almost all the training ratios studied, which is significant given the increased computational resources required for FT.

Considering average results, SSL-BarlowTwins-LP significantly outperforms the baseline FS-Random-LP model by 38.86% (0.561 vs. 0.404) and 21.32% (0.643 vs. 0.530) in LP and FT, respectively. When compared to FS-ImageNet-LP, SSL-BarlowTwins-LP shows an average performance improvement of 2.56% (0.561 vs. 0.547). However, the opposite is observed with the FT protocol, where FS-ImageNet-FT narrowly outperforms SSL-BarlowTwins-FT by 1.53% (0.653 vs. 0.643). Overall, both SSL-BarlowTwins and FS-ImageNet demonstrate strong performance in this downstream task, regardless of the transfer learning protocol used.

The difference between SSL-BarlowTwins-FT and FS-Random-FT becomes more pronounced in the pairwise comparison (Fig. 7a), with SSL-BarlowTwins-FT outperforming the baseline model in 58 out of 60 cases. For nearly every class and training ratio, SSL-BarlowTwins-FT consistently provides more accurate predictions than random initialization after FT. Furthermore, the FS-ImageNet-FT model outperforms the SSL counterpart in 41 out of 60 cases (Fig. 7b). Specifically, in the scenario using 100% of training data, SSL-BarlowTwins-FT excels in 8 out of 10 classes. This performance underscores the robustness of the model and its potential as a more efficient alternative for deep learning applications in these domains.

VI. Discussion

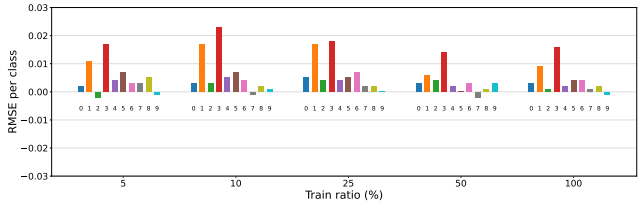
In this study, we compared the performance of SSL models with popular fully-supervised approaches for the analysis of RGB satellite patches, focusing on the critical task of LULC fraction estimation. Additionally, we explored the supplementary task of LULC scene classification to provide a more comprehensive evaluation. Specifically, we examined the effectiveness of transferring SSL models pretrained using in-domain knowledge compared to traditional supervised approaches. The results of these comparisons are discussed in detail below, highlighting the strengths and potential limitations of each approach in the RS context.

A. Barlow Twins: Leading in Clustering Performance

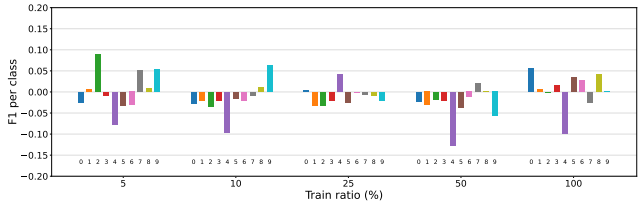
Among the four SSL models based on convolutional architectures, Barlow Twins distinctly demonstrated superior clustering performance. This method's ability to produce well-defined and compact clusters with clear boundaries between different classes indicates its exceptional capacity to capture the intricate patterns and nuances in RS data. The effectiveness of Barlow Twins in delineating these classes suggests that it generates highly discriminative features, which are essential for enhancing the performance of downstream tasks such as classification, segmentation, and change detection in RS applications.

B. Outperforming Random Initialization

Accurately capturing spatial characteristics is crucial for many RS applications. This study demonstrates the effectiveness of the Barlow Twins model in achieving this goal after



(a) LULC fraction estimation



(b) LULC scene classification

Fig. 8. Performance differences per class between the fine-tuned fully-supervised models initialized with the self-supervised learning pretrained and ImageNet-1k weights, the latter being the reference for the y-axis. Each color represents one of the 10 classes included in the downstream dataset, and the variations are shown for different train ratios.

pretraining on the Sentinel2GlobalLULC dataset. This advantage becomes evident when comparing the SSL-BarlowTwins and FS-Random models in the target downstream tasks.

Across both LP and FT scenarios, SSL-BarlowTwins consistently outperforms the baseline FS-Random model, showcasing the effectiveness of SSL to capture the spatial characteristics of the RS-specific classes within the pretraining dataset. Notably, the SSL-BarlowTwins model not only outperforms the FS-Random model but also shows remarkable consistency across various training ratios and downstream tasks, underscoring its potential as a powerful tool for RS applications.

C. Transfer Learning Effects Compared to ImageNet-1k

In the case of LP, SSL-BarlowTwins-LP consistently outperforms FS-ImageNet-LP across various training ratios and downstream tasks. This consistent performance highlights the nuanced impact of different transfer learning protocols on model performance, showcasing the superior adaptability of the SSL-BarlowTwins model in capturing relevant features for RS applications.

Fig. 8 shows the class-wise differences between the SSL-BarlowTwins-FT and FS-ImageNet-FT models, with the latter used as the reference. The performance differences between the two models for both downstream tasks are minimal. For FT, the analysis indicates an average difference of $\overline{\Delta\text{RMSE}} = 0.0053$ between the SSL pretrained and ImageNet-1k weights for LULC fraction estimation. In the scene classification task, the average difference in macro F1 score is $\overline{\Delta\text{F1}} = 0.0308$. These results highlight the competitive performance of the SSL-BarlowTwins-FT model, suggesting it is a viable alternative to traditional supervised approaches.

D. Maximizing Small Pretraining Data

The performance parity between SSL-BarlowTwins and FS-ImageNet is particularly noteworthy, especially considering

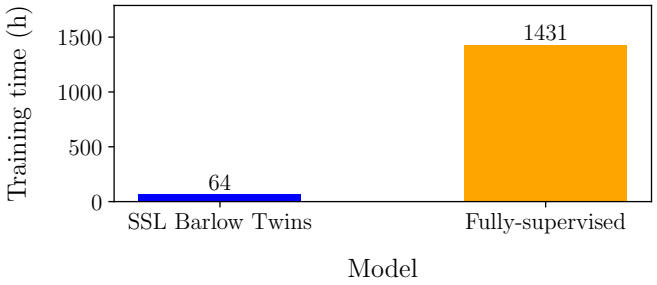


Fig. 9. Comparison of pretraining times between the self-supervised learning (SSL) Barlow Twins model, utilizing the Sentinel2GlobalLULC dataset, and its fully-supervised counterpart trained on ImageNet-1k. Both training processes were conducted on a single NVIDIA Tesla V100 for 500 epochs and employed a ResNet-18 backbone architecture with weights randomly initialized.

the significant differences in the number of classes and training images between the two pretraining datasets. The ImageNet-1k dataset, designed for visual object recognition, contains 1,000 classes and 1,281,167 training images. In contrast, the Sentinel2GlobalLULC dataset used to pretrain Barlow Twins comprises only 29 classes and 175,381 training samples of RGB satellite images.

This comparison highlights two key points. First, the robustness of SSL-BarlowTwins in effectively leveraging in-distribution data from smaller and more specialized datasets. Second, it underscores the potential of SSL-BarlowTwins in scenarios where massive datasets such as ImageNet-1k cannot be used for training. This adaptability is crucial, especially in fields with limited data availability.

E. Superior Training Time Efficiency

Fig. 9 illustrates the pretraining times required by the SSL Barlow Twins model, which employs a ResNet-18 architecture on the Sentinel2GlobalLULC dataset, compared to the same backbone when fully-supervised trained on the ImageNet1k dataset. Both scenarios utilize a single NVIDIA Tesla V100 and pretraining spans 500 epochs. The fully-supervised model, initialized randomly, demands approximately 59 days and 15 hours to complete the pretraining process. This extensive period is necessary to develop and adjust weights that are transferable for fine-tuning in downstream tasks.

In contrast, the SSL model completes pretraining on the Sentinel2GlobalLULC dataset in just 2 days and 16 hours. Despite the substantial reduction in the number of training samples when comparing both datasets, this discrepancy does not lead to a drastic decrease in accuracy. Remarkably, both approaches achieve similar accuracy results, yet the SSL approach offers a 22-fold acceleration compared to fully-supervised training. This significant reduction in training time highlights the efficiency and potential of SSL methods in scenarios where computational resources and time are limited.

VII. Conclusions

The role of SSL models in RS is gaining momentum due to their ability to learn valuable representations from unlabeled

data. In this study, we explore the potential of SSL in the challenging task of LULC fraction estimation using RGB satellite patches. Additionally, LULC scene classification was also explored as a supplementary experiment. The aim is to determine whether SSL models pretrained with in-domain RS datasets can outperform traditional supervised approaches, such as fully-supervised models with ImageNet-1k and random initialization.

The findings revealed that the features learned from SSL models, obtained through a pretext task on a small in-domain RS dataset of pure-pixel RGB satellite patches, were successfully transferred to perform LULC fraction estimation and scene classification on a smaller dataset of mixed pixels with varying train ratios. The experimental results demonstrated that the SSL model achieved competitive or slightly better results when trained on a smaller high-quality in-domain dataset compared to the supervised model trained on the extensive ImageNet-1k dataset. This underscores the impressive performance of SSL approaches pretrained using in-distribution datasets without labels, showcasing the efficient learning of SSL in terms of dataset size and training time.

This research opens new avenues for exploring SSL in RS data analysis, specifically for LULC problems. By leveraging the strengths of SSL models, we can enhance the accuracy and efficiency of LULC estimation and scene classification, potentially transforming practices in environmental monitoring and land management.

REFERENCES

[1] F. F. Sabins Jr and J. M. Ellis, *Remote sensing: Principles, interpretation, and applications*. Waveland Press, 2020.

[2] O. Dubovik, G. L. Schuster, F. Xu, Y. Hu, H. Bösch, J. Landgraf, and Z. Li, "Grand challenges in satellite remote sensing," p. 619818, 2021.

[3] R. P. Sishodia, R. L. Ray, and S. K. Singh, "Applications of remote sensing in precision agriculture: A review," *Remote Sensing*, vol. 12, no. 19, p. 3136, 2020.

[4] S. Khanal, K. Kc, J. P. Fulton, S. Shearer, and E. Ozkan, "Remote sensing in agriculture—accomplishments, limitations, and opportunities," *Remote Sensing*, vol. 12, no. 22, p. 3783, 2020.

[5] Y. Gao, M. Skutsch, J. Paneque-Gálvez, and A. Ghilardi, "Remote sensing of forest degradation: a review," *Environmental Research Letters*, vol. 15, no. 10, p. 103001, 2020.

[6] T. Wellmann, A. Lausch, E. Andersson, S. Knapp, C. Cortinovis, J. Jache, S. Scheuer, P. Kremer, A. Mascarenhas, R. Kraemer *et al.*, "Remote sensing in urban planning: Contributions towards ecologically sound policies?" *Landscape and urban planning*, vol. 204, p. 103921, 2020.

[7] P. C. Pandey and L. K. Sharma, *Advances in remote sensing for natural resource monitoring*. John Wiley & Sons, 2021.

[8] S. Kumar, R. S. Meena, S. Sheoran, C. K. Jangir, M. K. Jhariya, A. Banerjee, and A. Raj, "Remote sensing for agriculture and resource management," in *Natural Resources Conservation and Advances for Sustainability*. Elsevier, 2022, pp. 91–135.

[9] Q. Yuan, H. Shen, T. Li, Z. Li, S. Li, Y. Jiang, H. Xu, W. Tan, Q. Yang, J. Wang *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sensing of Environment*, vol. 241, p. 111716, 2020.

[10] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sensing*, vol. 12, no. 15, p. 2495, 2020.

[11] S. Gaur and R. Singh, "A comprehensive review on land use/land cover (lulc) change modeling for urban development: current status and future prospects," *Sustainability*, vol. 15, no. 2, p. 903, 2023.

[12] M. N. Ahmad, Z. Shao, and A. Javed, "Modelling land use/land cover (lulc) change dynamics, future prospects, and its environmental impacts based on geospatial data models and remote sensing data," *Environmental Science and Pollution Research*, vol. 30, no. 12, pp. 32 985–33 001, 2023.

[13] S. D. Sasmito, P. Taillardat, J. N. Clendenning, C. Cameron, D. A. Friess, D. Murdiyarso, and L. B. Hutley, "Effect of land-use and land-cover change on mangrove blue carbon: A systematic review," *Global Change Biology*, vol. 25, no. 12, pp. 4291–4302, 2019.

[14] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis *et al.*, "Deep learning for computer vision: A brief review," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PMLR, 2020, pp. 1597–1607.

[16] P. Berg, M.-T. Pham, and N. Courty, "Self-supervised learning for scene classification in remote sensing: Current state of the art and perspectives," *Remote Sensing*, vol. 14, no. 16, p. 3995, 2022.

[17] Y. Benhammou, D. Alcaraz-Segura, E. Guirado, R. Khaldi, B. Achchab, F. Herrera, and S. Tabik, "Sentinel2globallulc: A sentinel-2 rgb image tile dataset for global land use/cover mapping with deep learning," *Scientific Data*, vol. 9, no. 1, p. 681, 2022.

[18] Y. Wang, N. A. A. Braham, Z. Xiong, C. Liu, C. M. Albrecht, and X. X. Zhu, "Ssl4eo-s12: A large-scale multi-modal, multi-temporal dataset for self-supervised learning in earth observation," *arXiv preprint arXiv:2211.07044*, 2022.

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

[20] Y. Wang, C. M. Albrecht, N. A. A. Braham, L. Mou, and X. X. Zhu, "Self-supervised learning in remote sensing: A review," *arXiv preprint arXiv:2206.13188*, 2022.

[21] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.

[22] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct 2017. [Online]. Available: http://dx.doi.org/10.1109/JPROC.2017.2675998

[23] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[24] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *GIS '10*, 2010.

[25] A. Ghanbarzade and H. Soleimani, "Self-supervised in-domain representation learning for remote sensing image scene classification," 2023.

[26] C. J. Reed, X. Yue, A. Nrusimha, S. Ebrahimi, V. Vijaykumar, R. Mao, B. Li, S. Zhang, D. Guillory, S. Metzger *et al.*, "Self-supervised pretraining improves self-supervised pretraining," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2584–2594.

[27] C. Tao, J. Qi, M. Guo, Q. Zhu, and H. Li, "Self-supervised remote sensing feature learning: Learning paradigms, challenges, and future works," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[28] H. Jung, Y. Oh, S. Jeong, C. Lee, and T. Jeon, "Contrastive self-supervised learning with smoothed representation for remote sensing," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.

[29] D. Muhtar, X. Zhang, and P. Xiao, "Index your position: A novel self-supervised learning method for remote sensing images semantic segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2022.

[30] K. Ayush, B. Uzkent, C. Meng, K. Tanmay, M. Burke, D. Lobell, and S. Ermon, "Geography-aware self-supervised learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10 181–10 190.

[31] R. Balestriero, M. Ibrahim, V. Sobal, A. Morcos, S. Shekhar, T. Goldstein, F. Bordes, A. Bardes, G. Mialon, Y. Tian *et al.*, "A cookbook of self-supervised learning," *arXiv preprint arXiv:2304.12210*, 2023.

[32] R. Balestriero and Y. LeCun, "Contrastive and non-contrastive self-supervised learning recover global and local spectral embedding methods," *arXiv preprint arXiv:2205.11508*, 2022.

[33] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *International Conference on Machine Learning*. PMLR, 2021, pp. 12 310–12 320.

[34] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," *arXiv preprint arXiv:2003.04297*, 2020.
[35] X. Chen and K. He, "Exploring simple siamese representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15 750–15 758.
[36] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9729–9738.
[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[38] Y.-H. H. Tsai, S. Bai, L.-P. Morency, and R. Salakhutdinov, "A note on connecting barlow twins with negative-sample-free contrastive learning," *arXiv preprint arXiv:2104.13712*, 2021.
[39] Y. Benhammou, J. Rodríguez-Ortega, D. Alcaraz-Segura, and S. Tabik, "Andalunmixingrgb: A dataset of sentinel-2 rgb imagery acquired in andalusia region of spain, enriched with environmental ancillary data and annotated for blind spectral unmixing using deep learning (license cc by 4.0)," 2023. [Online]. Available: https://dx.doi.org/10.21227/y355-9h30
[40] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, "Tune: A research platform for distributed model selection and training," *arXiv preprint arXiv:1807.05118*, 2018.
[41] S. Tong, Y. Chen, Y. Ma, and Y. Lecun, "Emp-ssl: Towards self-supervised learning in one training epoch," *arXiv preprint arXiv:2304.03977*, 2023.
[42] G. Wang, K. Wang, G. Wang, P. H. Torr, and L. Lin, "Solving inefficiency of self-supervised representation learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9505–9515.
[43] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
[44] I. Susmelj, M. Heller, P. Wirth, J. Prescott, and M. E. et al., "Lightly," *GitHub. Note: https://github.com/lightly-ai/lightly*, 2020.

Andres J. Sanchez-Fernandez received a B.Sc. degree in Industrial Technology Engineering with a specialization in Automatic Systems and an M.Sc. degree in Mechatronic Engineering from the University of Malaga (Malaga, Spain) in 2016 and 2017, respectively. In 2018, he joined the Department of Computer Architecture at the same university to pursue a Ph.D. in Mechatronic Engineering, focusing on parallel programming on heterogeneous CPU-GPU systems, which he completed in February 2022, receiving an Excellent, Cum Laude, and International Doctorate distinction. Following his Ph.D., he worked at CénitS-COMPUTAEX (2022–2023) on data science and machine learning projects applied to renewable energies. Currently, he is a postdoctoral researcher at the Department of Computer Systems Engineering and Telematics at the University of Extremadura (Caceres, Spain). His research interests include data science, deep learning for satellite image processing, parallel programming, and large-scale data processing on heterogeneous systems.

Sergio Moreno-Álvarez received his B.Sc. in Computer Engineering (2012-2017), the M.Sc. in Computer Engineering (2017-2019) and the Ph.D. degree (2022) at the University of Extremadura. His Ph.D. has been awarded by the Geoscience and Remote Sensing Society's Spanish Chapter, recognizing his achievements during this period. His research interests cover High Performance Computing, Deep Learning and its optimization through parallelism approaches and Remote Sensing. He has served as reviewer in multiple IEEE journals, including IEEE Transactions on Geoscience and Remote Sensing, IEEE Geoscience and Remote Sensing Letters, IEEE Transactions on Emerging Topics in Computational Intelligence or IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, among others. He has participated as author/co-author in more than 15 JCR journal articles and multiple presentations at international and national conferences. Currently he is a member and professor at the Department of Computer Languages and Systems in the National University of Distance Education (UNED), Madrid, Spain.

Juan A. Rico-Gallego received the Computer Science Engineering degree and the Ph.D. degree in computer science from the University of Extremadura, Cáceres, Spain, in 2002 and 2016, respectively. Formerly a software consultant, he is an associate professor at the Dept. of Computer Systems Engineering of the University of Extremadura (Spain), and he is the general director of the Foundation for Advanced Computing and Technologies of Extremadura (COMPUTAEX) and the Extremadura Supercomputing Center (CénitS). His research interests are in analytical performance models on heterogeneous platforms and the usage of deep and reinforcement learning techniques to high-performance computing (HPC) platform optimization problems. He codevelops AzequiaMPI, an efficient thread-based full MPI 1.3 standard implementation on shared memory platforms. His other research interests include current message passing interface (MPI) implementations and applications and quantum computing technologies and development.

Siham Tabik received the B.Sc. degree in physics from the University Mohammed V, Rabat, Morocco, in 1998, and the Ph.D. degree in computer science from the University of Almería, Almería, Spain. She is currently Full Professor with the Department of Computer Science and Artificial intelligence, University of Granada, Granada, Spain. Her research interests include artificial intelligence, computer vision and remote sensing for biodiversity conservation and global climate change.