remote sensing

MDPI

Article

Cross-Domain Mixup for Parcel-Level Crop Mapping on a Multi-Year Sentinel-2 Dataset from Slovakia

Antonela-Adelina Dinescu and Corneliu Florea *

Image Processing and Analysis Laboratory, National University of Science and Technology Politehnica Bucharest, Splaiul Independentei 313, 060042 Bucharest, Romania; antonela.dinescu@upb.ro
* Correspondence: corneliu.florea@upb.ro

Highlights

What are the main findings?

- Geographically distinct European regions (e.g., France and Slovakia) exhibit noticeably different data distributions for the same crop types.
- CDMix, based on convex cross-domain mixing of source and target samples, achieves competitive performance in the tested settings.

What are the implications of the main findings?

- Geographic domain shifts can reduce model performance across fully labeled source and target datasets, motivating supervised domain adaptation.
- CDMix, when applied to pretrained models, can help mitigate the domain gap.

Abstract

Reliable crop-type mapping from satellite image time series is affected by distribution shifts across geographic regions, agricultural years, and heterogeneous label systems. To address this challenge, we propose Cross-Domain Mixup (CDMix), a supervised domain-adaptation method designed to leverage a larger labeled source dataset to improve performance on a smaller labeled target dataset under distribution shifts. We also introduce PixelSet-Slovakia, a new multi-year, parcel-level Sentinel-2 dataset covering three Slovak study regions and several growing seasons. Using a common backbone, we compare CDMix against three families of adaptation strategies: (i) no adaptation, (ii) weight transfer through fine-tuning and encoder freezing, and (iii) feature-space alignment using Maximum Mean Discrepancy (MMD) and Correlation Alignment (CORAL). All methods are evaluated in two scenarios: geographic supervised adaptation across datasets from two countries and temporal supervised adaptation across different growing seasons. Across both tested source–target settings, CDMix generally achieves competitive performance when initialized from pretrained representations. Under the region-held-out validation protocol, several pretrained adaptation strategies outperform training from scratch.

Keywords: crop-type mapping; supervised domain adaptation; transfer learning; satellite image time series; Sentinel-2; CDMix; temporal generalization; PixelSet-Slovakia

Check for updates

Academic Editor: Muhammad Moshiur Rahman

Received: 18 July 2026
Revised: 17 August 2026
Accepted: 20 August 2026
Published: 23 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Agricultural land use lies at the intersection of food security, climate mitigation, and an increasingly complex regulatory landscape. Understanding which crops are cultivated on specific parcels, and how these patterns evolve over time, is essential for both scientific analysis and policy implementation [1]. Satellite image time series (SITS) provide a

powerful means to observe crop phenology at large scales, offering consistent, objective, and temporally rich data [2]. A key challenge in this field is the difference in the data distributions and the label definitions across geographic regions and agricultural years. The performance of a model can decrease when it is applied beyond the domain in which it was trained. These challenges motivate the investigation of supervised adaptation across labeled domains.

In the European Union, Sentinel-based area monitoring is integrated into the Common Agricultural Policy (CAP) [3–5]. Beyond CAP, related requirements for spatially explicit monitoring are emerging in other policy domains: the Renewable Energy Directive (EU) 2023/2413 (RED III) links the eligibility of agricultural biomass to sustainability and land-use criteria [6], and Regulation (EU) 2023/1115 (EUDR) requires geolocation data for deforestation-free supply chains [7]. We emphasize that these policy settings provide context and motivation only; none of them is addressed or evaluated experimentally in this study. Within the CAP administrative framework, the Land Parcel Identification System (LPIS) defines parcel boundaries, while the Geospatial Aid Application records the crop types declared by farmers [8]. These systems provide large-scale parcel-level annotations that can support machine learning without dedicated field surveys [9]. Differences among national LPIS nomenclatures may result in non-equivalent crop definitions, complicating cross-country model transfer. The EuroCrops initiative seeks to mitigate this issue through a hierarchical taxonomy for harmonizing crop labels across countries [10]. In the present study, the labels are explicitly mapped to a common prediction space as detailed in Section 2.3.4.

In this paper, we study *supervised domain adaptation* for parcel-level crop classification under two forms of distribution shift: geographic and target-year. Both domains are fully labeled, and the question is whether the larger labeled source domain can improve training on a smaller labeled target domain. The method classifies declared agricultural parcels whose LPIS boundaries are given, within three Slovak study regions. It therefore requires parcel boundaries as prior data and does not constitute an end-to-end, wall-to-wall crop-mapping system or an operational national crop product. Within this scope, the paper makes the following contributions.

First, we release PixelSet-Slovakia, a multi-year (2020–2025), parcel-level Sentinel-2 dataset covering three Slovak administrative regions (Nitriansky, Banskobystrický, and Košický kraj). The dataset was constructed from LPIS declarations using a Google Earth Engine processing pipeline. To our knowledge, this is the first multi-year Sentinel-2 time-series PixelSet dataset covering multiple regions of Slovakia.

Second, we propose Cross-Domain Mixup (CDMix), an input-space regularization strategy that builds on our preliminary work [11] and on domain-mixup adaptation [12]. CDMix forms convex combinations of randomly paired source and target samples, with a mixing coefficient that follows a deterministic warm-up schedule, and trains with a weighted dual-domain loss. For comparison, we evaluate target-only training from scratch, weight transfer through full fine-tuning and encoder freezing, feature-space alignment using Maximum Mean Discrepancy (MMD) [13] and Correlation Alignment (CORAL) [14], domain-specific batch normalization (DSBN) [15], and asymmetric projected alignment. All methods use the Pixel-Set Encoder with Temporal Attention Encoder (PSE-TAE) as a common backbone.

Third, we evaluate the methods in two supervised settings: geographic adaptation from PixelSet-France to PixelSet-Slovakia in 2020, and target-year adaptation with previous-year support, using PixelSet-Slovakia data from 2020–2024 together with labeled 2025 data. Both settings follow a three-fold leave-one-region-out (LORO) protocol. Since the held-out region guides checkpoint selection, the reported scores constitute validation-based

evidence rather than independent test or final benchmark performance. We additionally report a zero-shot evaluation for each scenario.

1.1. Related Work

1.1.1. Sentinel-2 Based Datasets for Crop Mapping

In recent years, the remote sensing community has released numerous large-scale benchmark datasets for crop recognition. These datasets differ in the region they cover, their temporal span, the type of annotation they use, and the source of their labels, as summarized in Table 1.

Table 1. Public datasets based on Sentinel-2 data for crop recognition and mapping. *Format* denotes how each dataset represents its input unit: **Patch** (fixed-size windows for semantic/panoptic segmentation), **Parcel** (pixel time series sampled within a parcel), **Parcel (agg.)** (a single spatially aggregated time series per parcel), or **Pixel** (individual per-pixel time series). The source domain used in this study is the PixelSet-France dataset of [16] (France, tile T31TFM, 2017); it is distinct from the semantic/panoptic benchmark [17] (2018–2019).

Dataset	Format	Location	Years	Samples	Classes	GT Source
PASTIS PatchSet [17]	Patch	France	2018–2019	∼2433 patches	18	LPIS
PixelSet-France [16]	Parcel	France	2017	∼192,000 parcels	20	LPIS
BreizhCrops [18]	Parcel (agg.)	France	2017	∼610,000 parcels	13	LPIS
Dacia5 [19]	Patch	Romania	2020–2024	47 × 5 parcels	15	in situ
Sen4AgriNet [20]	Patch	Spain, France	2017–2020	>800,000 patches	9	LPIS
TimeSen2Crop [21]	Pixel	Austria	2017–2021	∼1.3 M pixels	16	LPIS
PixelSet-Slovakia	Parcel	Slovakia	2020–2025	75,493 observations	18	LPIS

In the last column, GT denotes the ground-truth source; agg. denotes spatial aggregation.

A more fundamental distinction is the manner in which each dataset represents its spatial units, since this determines how a model ingests the data. In the patch-based representation approach, the scene is partitioned into fixed-size image windows, and crop mapping is formulated as a dense prediction task corresponding either to semantic segmentation or to panoptic segmentation. A representative instance of this approach is the PASTIS PatchSet [17], which established a widely used benchmark for both tasks; Sen4AgriNet [20] adopts the same patch-based formulation while broadening it to additional years and a second country. At the parcel level, vector parcel boundaries derived from LPIS are used to represent each agricultural plot as a single, self-contained input. Two variants of this parcel-level representation can be distinguished. In the first, the parcel is described by a set of pixel time series sampled within its boundary, as in the PixelSet-France dataset introduced with PSE-TAE [16] and in the proposed PixelSet-Slovakia dataset. In the second, the parcel is summarized by a single aggregated time series, as in BreizhCrops [18]. We constructed PixelSet-Slovakia following the representation introduced by Sainte Fare Garnot et al. [16]. By removing the need for tiling and stitching, this object-centric representation is considerably more efficient in terms of both storage and training cost [16].

Nevertheless, the majority of existing studies concentrate on a limited number of well-studied regions, predominantly in Western Europe. The selected Slovak regions constitute a relevant and challenging case study of model adaptation to underrepresented regions and across agricultural years.

1.1.2. Domain Adaptation in Remote Sensing

Domain adaptation provides a framework for transferring knowledge between source and target domains whose data distributions differ [22]. In remote sensing, such differ-

ences may arise across geographic regions, acquisition periods, sensors, or observation conditions [23]. Supervised transfer learning has also been investigated for crop classification by reusing pretrained representations [24]. A substantial part of the remote-sensing literature has focused on unsupervised domain adaptation, in which labeled source data are combined with an unlabeled target domain. Examples include temporal adaptation for SITS-based land-cover mapping [25], architecture search with self-training [26], and spatial and temporal adaptation for Sentinel-2 crop classification [27].

The adaptation strategies operate at different stages of the learning process. Input-space approaches modify or combine samples before feature extraction, with domain mixup providing one mechanism for constructing intermediate source–target observations [12]. Feature-space approaches instead reduce discrepancies between the learned source and target representations. MMD has been used in deep adaptation networks [13], hyperspectral-image classification [28], and spatially and temporally transferable crop mapping [27]. CORAL aligns second-order feature statistics [14] and has also been incorporated into multi-source hyperspectral-image adaptation [29]. Domain-specific batch normalization provides a complementary strategy by maintaining separate normalization statistics for different domains [15].

These studies demonstrate the diversity of adaptation strategies available in remote sensing, but they predominantly examine unlabeled target domains or evaluate individual adaptation mechanisms in isolation. The setting considered here is different: the target-domain training partition is fully labeled but smaller than the source domain. Comparative evidence under this form of supervised domain adaptation remains limited, particularly when multiple adaptation families are evaluated using a common backbone and spatially disjoint validation.

1.1.3. Relation to Previous Work and Paper Organization

A preliminary version of the proposed strategy was presented in [11], where Mixup-based transfer was investigated using a source dataset and a limited Romanian target dataset. The present work extends that study in several aspects: it introduces PixelSet-Slovakia, adds a target-year supervised adaptation setting and separate zero-shot evaluations, considers temporal, spectral, and spatial mixing variants, and broadens the experimental comparison to multiple adaptation families evaluated under region-held-out validation.

The rest of the paper is organized as follows: Section 2 describes the datasets, label harmonization, formal problem, model, and adaptation strategies. Section 3 presents the experimental protocol and validation results. Section 4 discusses the findings and limitations, while Section 5 concludes the paper.

2. Materials and Methods

2.1. Formal Problem Definition

Intuitively, the setting is the following: we are given a large labeled source dataset and a smaller labeled target dataset whose data distribution differs from that of the source. A model trained only on the smaller target dataset is limited by the amount of target data, so we ask whether the larger source dataset can be exploited during training to improve performance on the target, despite the distribution shift between the two. The rest of this section formalizes this setting.

A parcel is represented as $\mathbf{x} \in \mathbb{R}^{T \times C \times S}$, an unordered set of S pixels, each described by $C = 10$ spectral bands over $T = 24$ temporal observations, with a crop label y drawn from a common prediction space $\mathcal{Y}$ of $|\mathcal{Y}| = 20$ classes, established through the label harmonization of Section 2.3.4. We are given a labeled source domain $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{n_s}$

and a smaller, fully labeled target-domain training partition $\mathcal{D}_t^{\text{tr}} = \{(\mathbf{x}_j^t, y_j^t)\}_{j=1}^{n_t}$, with $n_s > n_t$ and $P_s(\mathbf{x}, y) \neq P_t(\mathbf{x}, y)$. Because both training domains are fully labeled, this is supervised domain adaptation under data-volume asymmetry, and target labels enter every training loss. The goal is a classifier $f : \mathbb{R}^{T \times C \times S} \to \mathcal{Y}$ that generalizes in a held-out target validation region despite the shift between P_s and P_t. In the separate zero-shot evaluations, the classifier is trained only on $\mathcal{D}_s$ and applied to the target domain without any target-label parameter updates.

2.2. Datasets and Experimental Domains

The study uses two parcel-level Sentinel-2 datasets whose roles depend on the experimental setting. Both datasets use a common set of 10 spectral bands: B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12. The 60 m atmospheric bands B1, B9, and B10 are not used. For PixelSet-Slovakia, pixel values were extracted at a nominal spatial scale of 10 m. The native 20 m bands were implicitly resampled to the 10 m sampling grid using the default nearest-neighbor procedure.

PixelSet-France is a parcel-level dataset introduced together with the PSE-TAE architecture [16]. It covers Sentinel-2 tile T31TFM in France during the 2017 agricultural year. The second dataset is the proposed PixelSet-Slovakia, a new multi-year parcel-level Sentinel-2 dataset constructed from Slovak LPIS declarations, as described in Section 2.3. It covers the 2020–2025 agricultural years in three selected Slovak administrative regions. Although both datasets use the same spectral bands, their original crop-label nomenclatures differ. The mapping of the introduced dataset to the prediction space is described in Section 2.3.4, with the complete label correspondence provided in Appendix A.

The assignment of source and target domains is scenario-specific. In the geographic supervised-adaptation setting, the PixelSet-France dataset from 2017 is used as the source domain and PixelSet-Slovakia from 2020 as the target domain. In the target-year supervised-adaptation setting, the Slovak data from 2020–2024 provide the historical source domain, while the labeled data from 2025 form the target domain. All samples in the target-domain training partitions are labeled. Separate zero-shot evaluations are conducted without using labels from the relevant target domain or target year for model updating.

PixelSet-Slovakia includes Nitriansky kraj, Banskobystrický kraj, and Košický kraj. These administrative regions define the spatial partitions used in the LORO evaluation. In each fold, two regions are used for target-domain training, and the remaining region is reserved for validation. Their geographic locations are shown in Figures 1 and 2, while the corresponding region-by-year dataset statistics are provided in Appendix B.



Figure 1. Geographic distribution of the data used in this study. The western hatched area corresponds to Sentinel-2 tile T31TFM in France, used for the 2017 PixelSet-France dataset. The three blue areas indicate the Slovak regions included in PixelSet-Slovakia for 2020–2025.

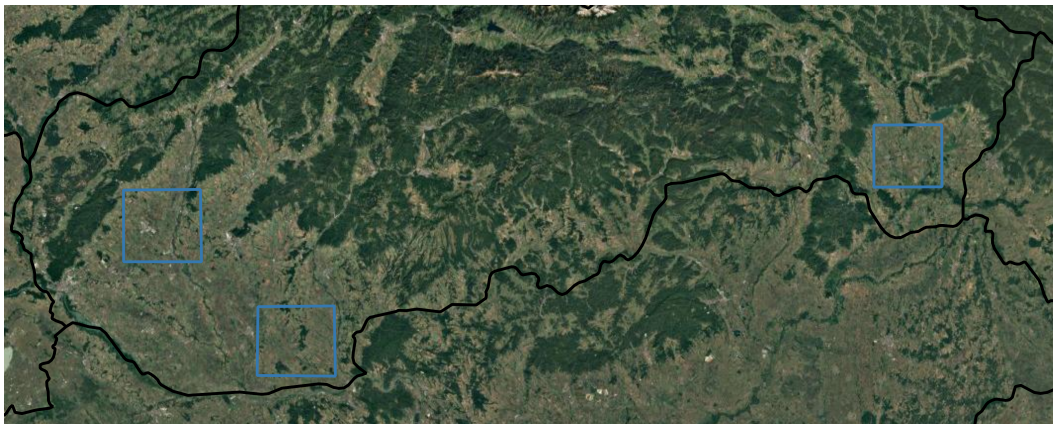


Figure 2. Location of the Slovak administrative regions included in PixelSet-Slovakia. These regions define the spatial partitions used in the LORO evaluation and do not represent national coverage of Slovakia.

2.3. Target Dataset Construction

2.3.1. LPIS Labels and Vector Preprocessing

Parcel geometries and crop labels were obtained from the official Slovak LPIS [30], which provides, for each agricultural year, the parcel polygons and the farmer-declared crop, encoded by a nomenclature of approximately 180 classes. In a geographic information system workflow, the geometries were reprojected to a reference frame compatible with the Sentinel-2 grid, restricted to the three study regions, and cleaned of invalid or degenerate polygons. The original declared parcel boundaries were retained for extraction; no geometric erosion or inward buffering was applied.

The observations are treated as parcel–year instances; recurring parcel entities were identified geometrically for dataset characterization (Appendix B). The resulting per-year and per-class composition, after the label harmonization of Section 2.3.4, is summarized in Table 2. The complete region-by-year characterization and the extraction-pipeline details are provided in Appendix B.

Table 2. Class distribution of the PixelSet-Slovakia dataset by agricultural year, expressed as the number of parcel–year instances assigned to each harmonized crop class. A parcel declared in multiple years contributes one separate instance in each corresponding year; consequently, the annual observations should not be interpreted as independent physical parcels. The Total column aggregates the six annual datasets. Class names correspond to the harmonized labels defined in Section 2.3.4.

Crop Class	2020	2021	2022	2023	2024	2025	Total
Meadow	1025	1061	1072	1021	937	892	6008
Soft winter wheat	3462	2792	3443	3678	2381	2942	18,698
Corn	2290	2421	2071	1597	1805	2097	12,281
Winter barley	528	473	573	586	474	466	3100
Winter rapeseed	1150	998	1131	1347	1315	1124	7065
Spring barley	822	926	566	655	681	541	4191
Sunflower	787	1069	1132	918	1109	1078	6093
Grapevine	527	637	634	622	608	556	3584
Beet	107	132	129	143	159	119	789
Winter triticale	0	0	0	0	38	0	38
Winter durum wheat	0	0	0	553	749	824	2126
Fruits veg flowers	258	261	254	237	190	254	1454
Potatoes	128	124	118	106	118	118	712
Leguminous fodder	546	540	529	409	397	447	2868
Soybeans	659	809	858	562	942	871	4701
Orchard	237	245	240	224	225	233	1404
Mixed cereal	0	0	0	0	0	32	32
Sorghum	28	35	48	51	99	88	349
Total	12,554	12,523	12,798	12,709	12,227	12,682	75,493

The class distribution is strongly imbalanced, reflecting the retained LPIS declarations without any over- or undersampling. The methodology, inherited from the original PSE-TAE setup, accounts for this at both stages: training uses Focal Loss, which down-weights well-classified examples, and evaluation emphasizes balanced accuracy and mean Intersection-over-Union (IoU), which weight each represented class equally. The rarest classes nevertheless retain limited statistical support, so their per-class results are interpreted cautiously.

2.3.2. Sentinel-2 Time Series and Preprocessing

Time series were derived from the Google Earth Engine Sentinel-2 surface-reflectance collection (`COPERNICUS/S2_SR_HARMONIZED`), which is already atmospherically corrected, retaining the same ten bands as the source domain (Section 2.2). All Google Earth Engine operations were executed through the `earthengine-api` library (v1.7.36). The preprocessing chain (Figure 3) comprises spatio-temporal filtering, cloud masking, temporal compositing, gap filling, and parcel-level extraction.

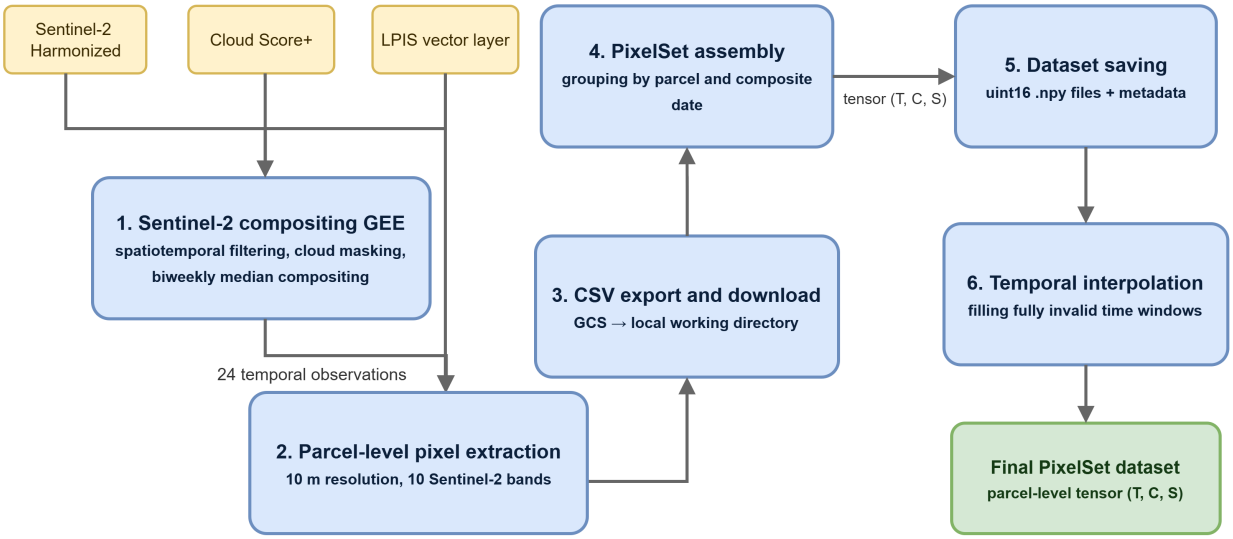


Figure 3. Preprocessing chain used to generate the target Sentinel-2 time series: spatio-temporal filtering, cloud masking with Cloud Score+, approximately bi-weekly median compositing, temporal gap filling, and parcel-level extraction in (T, C, S) format.

Observations were restricted to the LPIS extent and the annual analysis period. Within this window, all available Sentinel-2 observations passing the cloud criteria were extracted: images with a scene-level `CLOUDY_PIXEL_PERCENTAGE` of at least 80% were discarded, and pixels were retained only where the Cloud Score+ clarity score satisfied `cs_cdf` ≥ 0.60. Because acquisition dates and cloud coverage vary across years and locations, the resulting time series are irregular, with unequal numbers of valid observations per period. To bring these irregular series onto a common grid, the retained imagery was subsampled through median compositing into 24 approximately bi-weekly composites, with target dates on the 1st and 15th of each month, and the median being robust to any remaining invalid pixels. This temporal grid is identical across all six years, so the Temporal Attention Encoder receives fixed-length sequences on a common temporal representation. Composites left without valid data were filled along the temporal axis, independently per pixel and band: interior gaps by linear interpolation between the nearest valid observations, and leading or trailing gaps with the nearest available observation. Directly observed values were left unchanged.

2.3.3. PixelSet Extraction and Assembly

For each LPIS polygon, the Sentinel-2 pixels inside the original declared boundary were stored as a tensor of shape $(T = 24,\ C = 10,\ S_p)$, where S_p is the number of available pixel locations within parcel p (Figure 4), preserving both the temporal signal and the internal spatial variability of the parcel [16]. For model input, the data loader provides a fixed set of $S = 128$ pixels per parcel: larger parcels are randomly subsampled without replacement, while smaller ones are padded by repeating an available pixel, with a binary mask excluding the padded positions from spatial pooling. Pixel sampling is controlled by the random seed of the corresponding run. The automated extraction pipeline itself is described in Appendix B.

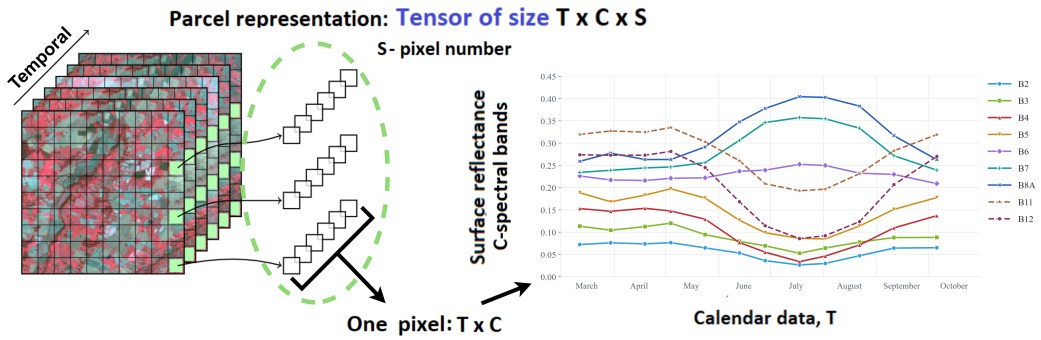


Figure 4. PixelSet representation of an agricultural parcel. The stored parcel contains a variable number S_p of pixel-level multispectral time series on a fixed temporal grid, with each pixel represented by a time series of shape (T, C). During model loading, a fixed input of $S = 128$ pixels is produced through random subsampling or masked padding.

2.3.4. Label-Space Harmonization

Domain adaptation requires the source and target domains to share a common prediction space. The original Slovak crop names were translated into English and matched semantically to the classes of the PixelSet-France dataset, following the general harmonization principle formalized by the EuroCrops Hierarchical Crop and Agriculture Taxonomy (HCAT) [10], although HCAT codes were not used directly. Only sufficiently unambiguous correspondences were retained: broader categories, different crop uses, mixed or undersown variants, and seasonally unspecified descriptions were assigned a null mapping and excluded. The mapping was established from the 2020 nomenclature and extended to crop names appearing in later years; the complete correspondence, including the retained Slovak descriptions, the mapping rules, and illustrative examples of the conservative decisions, is provided in Appendix A. After harmonization, PixelSet-Slovakia contains 18 retained crop classes. Source-domain background and void categories have no corresponding declared crop class in the target dataset and were excluded from evaluation. Some retained crop classes are nevertheless absent from particular years or validation regions; such classes are excluded from the corresponding per-class metric averages rather than being assigned a score of zero.

2.3.5. Residual Label-System Uncertainty

The harmonization provides a consistent operational prediction space but cannot guarantee complete semantic equivalence between national LPIS nomenclatures; residual label-system mismatch may therefore contribute to the observed performance differences, alongside spectral–temporal and class-distribution shifts.

2.4. Contextual Characterization of Domain Differences

Coarse-resolution climatic variables are reported separately to provide contextual information about the study areas and years. They are not inputs to the model; we use them to describe environmental differences between domains, not to establish a causal contribution to model performance. Figure 5 presents semi-monthly daytime and nighttime land-surface temperature, accumulated precipitation, and surface soil moisture. The same descriptors are reported for both experimental settings. They were derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) MOD11A2 product, the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) dataset, and the Soil Moisture Active Passive (SMAP) L4 product, respectively. The class composition of the target dataset is documented in Section 2.3.4 and in the complete region-by-year characterization provided in Appendix B; the source-domain composition is reported in [16]. Residual differences in label semantics are treated as label-system uncertainty rather than as a directly measured form of concept shift.

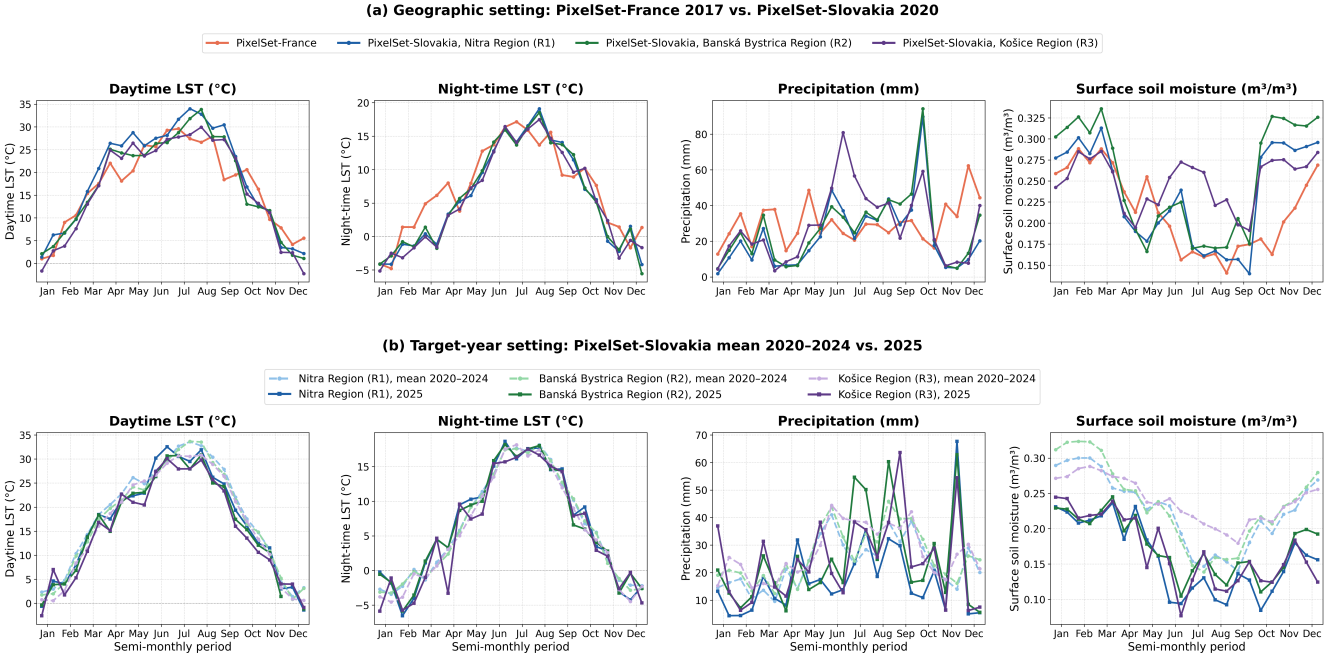


Figure 5. Climatic descriptors reported as contextual information for the two experimental settings. The upper row compares the PixelSet-France source area in 2017 with the three PixelSet-Slovakia regions in 2020. The lower row compares the regional means for 2020–2024 with the corresponding observations from 2025. From left to right, the panels show daytime land-surface temperature, nighttime land-surface temperature, precipitation, and surface soil moisture. In the lower row, lighter dashed lines represent the 2020–2024 means, whereas darker solid lines represent 2025.

Figure 6 provides a class-level description of the spectral–temporal differences observed between the PixelSet-France and PixelSet-Slovakia datasets. The comparison uses the same Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI) diagnostics for both domains. These profiles are descriptive and are not used to attribute performance differences to a particular environmental or labeling factor.

2.5. Classification Backbone: PSE-TAE

All experiments use a single fixed backbone, PSE-TAE [16], so that differences in performance are attributable to the adaptation strategy rather than to the model.

The Pixel-Set Encoder (PSE) processes each timestep independently, treating the parcel pixels as an unordered set. A shared multilayer perceptron embeds each pixel's spec-

tral vector into a 64-dimensional latent space; the per-timestep parcel representation is then obtained by concatenating the mean and standard deviation of the pixel embeddings (computed only over real pixels indicated by a mask), yielding a permutation- and cardinality-invariant 128-dimensional vector that is refined by a second perceptron.

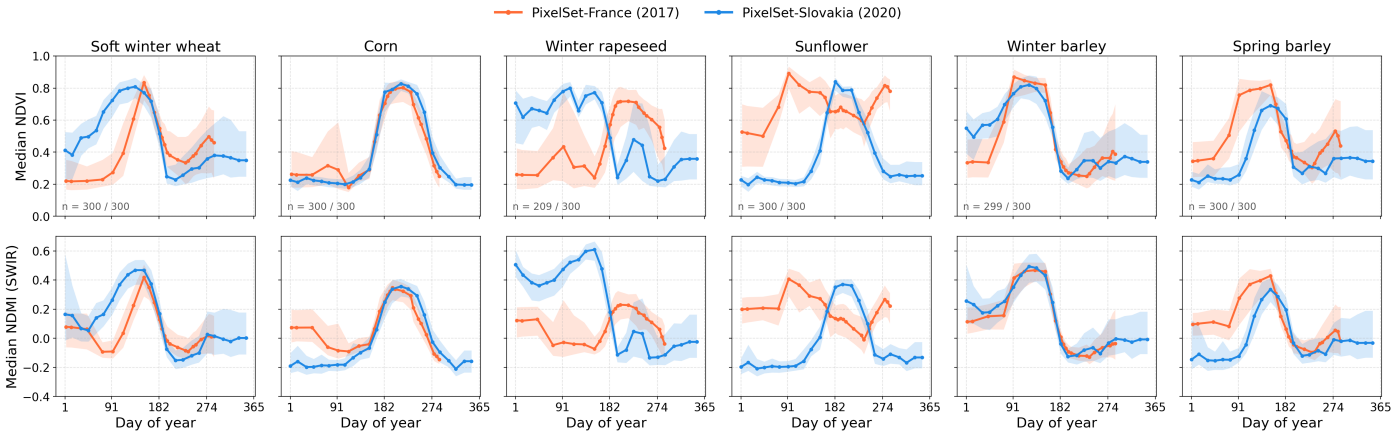


Figure 6. Median temporal profiles of NDVI (**top row**) and NDMI (**bottom row**) for six major harmonized classes (soft winter wheat, corn, winter rapeseed, sunflower, winter barley, and spring barley). The PixelSet-France source from 2017 is shown in orange and PixelSet-Slovakia from 2020 in blue. Shaded bands show the interquartile range (25th–75th percentile) across parcels. The horizontal axis represents the 24 approximately bi-weekly composite time steps, corresponding to target dates on the 1st and 15th of each month. The profiles provide a descriptive comparison of the observed spectral–temporal patterns and are not used to identify their causal origin.

The Temporal Attention Encoder (TAE) summarizes the resulting sequence of per-timestep representations into a single global vector through multi-head self-attention, using a single synthetic master query per sequence and a bespoke positional encoding based on the real calendar offsets between acquisitions. The 128-dimensional global parcel embedding z produced at this stage is the feature on which our feature-space alignment methods operate (Section 2.6). A multilayer perceptron classification head maps z to the 20 classes, so the model performs object-level (parcel) classification rather than per-pixel segmentation. Training uses Focal Loss [31]

$$\mathcal{L}_{\mathrm{FL}} = -(1 - p_t)^{\gamma} \log(p_t), \tag{1}$$

where p_t is the softmax probability of the true class and $\gamma = 1$, a value retained from the original PSE-TAE configuration [16] and not re-tuned in this study. The factor $(1 - p_t)$ down-weights easily classified examples and concentrates the gradient on rare and difficult classes, which is essential under the pronounced class imbalance of agricultural data.

2.6. Domain-Adaptation Strategies

Building upon PSE-TAE as the backbone, we propose the CDMix strategy, which is evaluated against several baselines. All strategies share the same classification objective (Equation (1)) and the same protocol (Section 3.1); they differ in how source knowledge is transferred to, or aligned with, the target.

2.6.1. CDMix: Cross-Domain Mixup

The proposed adaptation method builds on Mixup, introduced for image classification by Zhang et al. [32], which generates synthetic training examples as combinations of sample pairs and their labels. Given two training examples $(\mathbf{x}_i, y_i)$ and $(\mathbf{x}_j, y_j)$, the mixed sample $(\tilde{\mathbf{x}}, \tilde{y})$ is defined as

$$\tilde{\mathbf{x}} = \lambda \mathbf{x}_i + (1-\lambda)\mathbf{x}_j, \quad \tilde{y} = \lambda y_i + (1-\lambda) y_j, \tag{2}$$

where $\lambda \sim \text{Beta}(\alpha, \alpha)$ is sampled from a Beta distribution parameterized by the hyperparameter $\alpha > 0$. Mixup can be interpreted as a regularizer that encourages locally linear behavior between training samples by filling the low-density regions of the input space with synthetic interpolations [32]. This linearity assumption may also constitute a limitation, particularly when the underlying data manifold is highly non-linear: interpolated samples may lie outside the true data distribution, leading to manifold intrusion, where synthetic examples conflict with genuine observations and introduce label ambiguity [33]. In the cross-domain setting considered here, this limitation has an additional facet: when crop calendars differ between the source and target domains (Figure 6), a convex combination of two time series need not correspond to a physically realistic crop trajectory. Input-space mixing should therefore be read as a regularization mechanism rather than as the synthesis of observations. In this work, we adapt Mixup to parcel-level spatio-temporal classification, where each input is a tensor of size $T \times C \times S$ (Section 2.3), and apply it between samples originating from two different domains: one drawn from the smaller labeled target region and the other from a larger, high-resource labeled reference region. This cross-domain mixing is inspired by MixMatch [34] and by Adversarial Domain Mixup [12]. Source and target samples are paired at random, without any class-matching constraint. Let $(\mathbf{x}_t, y_t)$ and $(\mathbf{x}_s, y_s)$ denote batches sampled from the target and reference regions, respectively. The mixed input is computed as

$$\widetilde{\mathbf{x}_{mix}} = \lambda \cdot \mathbf{x}_s + (1-\lambda) \cdot \mathbf{x}_t, \tag{3}$$

where the interpolation follows one of the variants described below. In a baseline version, rather than explicitly interpolating the labels, the loss is computed as a weighted combination of the individual classification (Focal) losses,

$$\mathcal{L}(\widetilde{\mathbf{x}_{mix}}) = \lambda \cdot \mathcal{L}(y_s, \hat{y}) + (1-\lambda) \cdot \mathcal{L}(y_t, \hat{y}), \tag{4}$$

where $\hat{y}$ denotes the predicted class probabilities for $\widetilde{\mathbf{x}_{mix}}$, and $\mathcal{L}(\cdot)$ is the classification loss (the Focal Loss of Equation (1)). Unlike the original Mixup, λ is not drawn from a Beta distribution: the proposed method mixes source and target batches under a deterministic coefficient $\lambda \in [0, 1]$ that decreases linearly from 1 (source only) to 0 (target only) over W warm-up epochs e,

$$\lambda(e) = \max\left(0,\ 1 - \frac{e-1}{W-1}\right), \quad e \in \{1, \ldots, W\}. \tag{5}$$

Four variants are evaluated:

1. linear—the baseline presented earlier in Equation (4);
2. temporal;
3. spectral (bands);
4. spatial (pixels).

The last three variants instead splice a λ-determined fraction of timesteps, bands, or pixels from the source into the target sample (Figure 7). The schedule lets the model first stabilize on the source and then migrate smoothly toward the target distribution.

2.6.2. Baseline Adaptation Methods

Several alternatives for domain adaptation have been considered. These range from trivial baselines to recent, well-established methods.

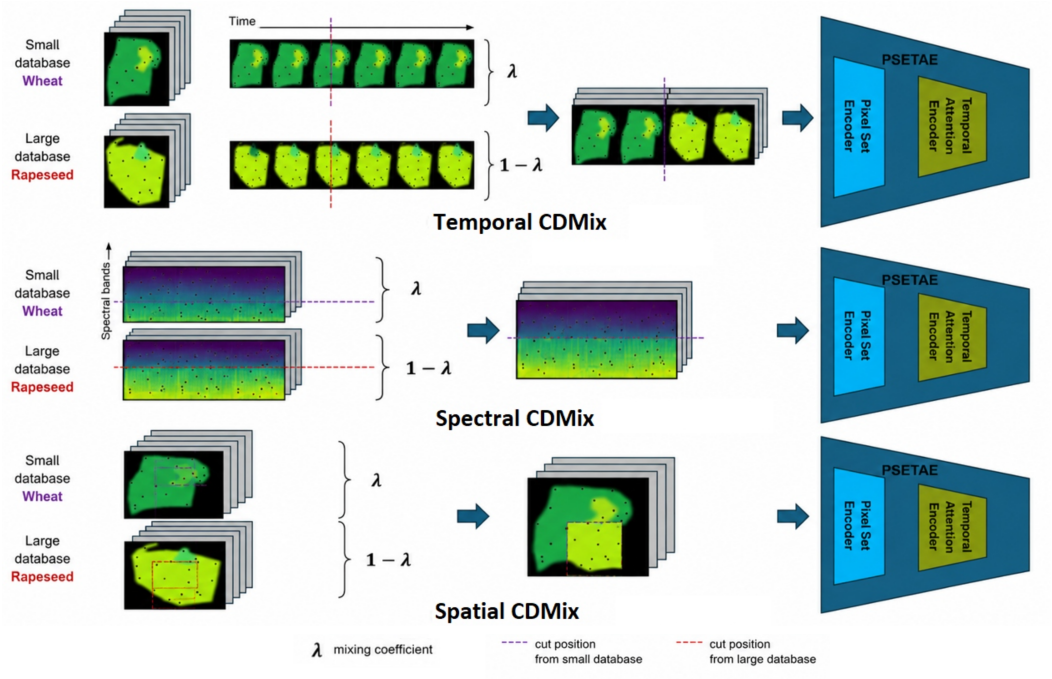


Figure 7. Conceptual diagram of the input-space CDMix strategy. At each iteration, a batch from the large source dataset and a batch from the smaller target dataset are combined using a mixing coefficient λ, which decreases from 1 to 0 during the warm-up period, gradually shifting the training from source-dominated to target-dominated.

No Adaptation

This family groups the three transfer-free references reported in Tables 3 and 4. The target-only reference is a model trained from scratch on the target domain only, with random initialization; it quantifies the performance attainable without any transfer and serves as the lower-bound reference against which all transfer and alignment strategies are compared. The zero-shot reference applies the source-trained model directly to the target data, without any target labels or parameter updates, and measures the severity of the raw domain shift. Finally, the source reference reports the performance of the source model within its own domain.

Table 3. Geographic scenario (France 2017 → Slovakia 2020). All figures are the mean ± standard deviation over the three LORO folds; the per-region validation IoU of the principal configurations is reported in Appendix D. Methods grouped by adaptation family and ordered by validation IoU within each family (the reference rows follow a fixed logical order: target-only, zero-shot, source).

Method	IoU	Bal. Acc (%)	Acc (%)
No adaptation (reference)			
Scratch (target only)	0.417 ± 0.126	57.4 ± 10.4	73.9 ± 11.1
Zero-shot (France → Slovakia) [b]	0.064	16.6	25.5
Source ref. (PixelSet-France) [a]	0.515	57.45	94.39
Weight transfer			
Fine-tune all	0.465 ± 0.106	58.7 ± 10.5	77.5 ± 9.3
Freeze TAE (temporal)	0.449 ± 0.088	60.1 ± 7.4	78.4 ± 5.2
Freeze PSE (spatial)	0.443 ± 0.112	57.6 ± 11.2	76.5 ± 9.8
Freeze encoder	0.313 ± 0.060	46.1 ± 5.9	68.8 ± 7.5

Table 3. *Cont.*

Method	IoU	Bal. Acc (%)	Acc (%)
Input-space alignment (CDMix)			
CDMix temporal, $W = 20$, pretr. ⋆	**0.475 ± 0.086**	**61.6 ± 7.0**	78.9 ± 5.5
CDMix temporal, $W = 20$, scratch	0.469 ± 0.072	61.4 ± 5.6	81.0 ± 3.8
CDMix pixels, $W = 20$, scratch	0.466 ± 0.084	60.0 ± 7.8	77.7 ± 7.1
CDMix temporal, $W = 50$, scratch	0.457 ± 0.057	60.0 ± 5.3	76.2 ± 5.7
CDMix temporal, $W = 50$, pretr.	0.454 ± 0.103	61.4 ± 8.3	77.2 ± 8.1
CDMix pixels, $W = 20$, pretr.	0.451 ± 0.106	58.5 ± 10.5	76.4 ± 7.9
CDMix bands, $W = 50$, pretr.	0.451 ± 0.108	59.4 ± 8.7	78.1 ± 6.5
CDMix, $W = 20$, pretr.	0.448 ± 0.089	59.8 ± 8.9	**81.1 ± 3.8**
CDMix bands, $W = 50$, scratch	0.447 ± 0.089	59.4 ± 5.8	78.8 ± 4.3
CDMix bands, $W = 20$, scratch	0.446 ± 0.083	59.8 ± 5.3	79.6 ± 4.6
CDMix, $W = 20$, scratch	0.445 ± 0.090	60.1 ± 5.6	76.6 ± 6.5
CDMix, $W = 50$, scratch	0.441 ± 0.102	59.5 ± 6.1	73.9 ± 8.9
CDMix bands, $W = 20$, pretr.	0.440 ± 0.098	58.0 ± 10.4	77.3 ± 9.2
CDMix, $W = 50$, pretr.	0.440 ± 0.140	57.2 ± 14.5	72.1 ± 16.2
CDMix pixels, $W = 50$, pretr.	0.414 ± 0.161	51.2 ± 18.4	62.5 ± 25.7
CDMix pixels, $W = 50$, scratch	0.409 ± 0.162	51.3 ± 17.8	61.9 ± 26.3
Feature-space & normalization alignment			
DSBN	0.468 ± 0.098	60.4 ± 8.5	78.4 ± 8.1
Projected align. (MMD), pretr.	0.450 ± 0.107	58.3 ± 11.4	74.7 ± 13.7
Projected align. (MMD), scratch	0.418 ± 0.104	56.4 ± 7.0	74.4 ± 8.4
CORAL, scratch	0.404 ± 0.096	57.3 ± 6.0	75.4 ± 6.0
MMD, scratch	0.389 ± 0.113	55.0 ± 10.7	72.0 ± 12.7
CORAL, pretr.	0.386 ± 0.095	53.9 ± 10.5	68.5 ± 11.6
MMD, pretr.	0.369 ± 0.086	52.3 ± 8.6	67.6 ± 11.9

[a] Source-domain reference (trained and evaluated on the source PixelSet-France); not a transfer result and not directly comparable to the target rows. [b] France-trained source model applied directly to the complete Slovak 2020 target with no target labels or parameter updates; evaluated once on the full target, hence no fold deviation. Best values per column (among target rows) in bold; the best overall configuration of each scenario is marked with a star (⋆).

Weight Transfer

The second family reuses the source-pretrained weights. Early work showed that pretraining on a large dataset followed by fine-tuning outperforms training from scratch when the target dataset is comparatively small [35]. More recent studies have nuanced this picture, with the benefit depending on the size of the target dataset relative to the source, but the general approach has been repeatedly confirmed for remote-sensing and agricultural data [24].

Here, in full fine-tuning, all weights are restored from the source checkpoint except the final classification layer, which is reinitialized, and the whole network is then trained on the target. In encoder freezing, part of the pretrained encoder is frozen, and only the remaining weights are trained: we consider freezing the spatial encoder (PSE), the temporal encoder (TAE), or the entire encoder (training only the decoder). These configurations probe how reusable the pretrained representations are under each type of shift.

Feature-Space Alignment

This family of methods aligns the source and target domains directly in the feature space, following the deep feature-alignment approach of Long et al. [13]. Feature-space alignment has since been widely adopted for domain adaptation, including applications to

remote-sensing imagery [28] and, more specifically, to Sentinel-based crop classification [27]. However, these methods were not designed for the tensor-structured spatio-temporal representations used in datasets such as PixelSet-France and PixelSet-Slovakia. Consequently, their direct application is not straightforward and requires appropriate modifications.

Table 4. Temporal scenario—target-year supervised adaptation with previous-year support (Slovakia 2020–2024 → 2025). All figures are the mean ± standard deviation over the three LORO folds; the per-region validation IoU of the principal configurations is reported in Appendix D. Methods grouped by adaptation family and ordered by validation IoU within each family (the reference rows follow a fixed logical order: target-only, zero-shot, source). Best values per column in bold; ⋆ marks the best overall configuration.

Method	IoU	Bal. Acc (%)	Acc (%)
No adaptation (reference)			
Scratch (target only)	0.418 ± 0.082	56.6 ± 8.5	77.3 ± 4.8
Zero-shot (2020–2024 → 2025) [b]	0.478 ± 0.066	62.5 ± 8.1	80.3 ± 0.2
Source ref. (2020–2024) [a]	0.476 ± 0.109	60.1 ± 13.8	80.2 ± 6.4
Weight transfer			
Freeze encoder	0.498 ± 0.099	64.6 ± 10.4	82.5 ± 1.3
Freeze PSE (spatial)	0.486 ± 0.089	63.0 ± 10.1	80.9 ± 2.8
Fine-tune all	0.479 ± 0.085	63.2 ± 9.4	81.5 ± 3.2
Input-space alignment (CDMix)			
CDMix temporal, $W = 50$, pretr. ⋆	**0.503 ± 0.095**	65.0 ± 10.1	83.5 ± 2.5
CDMix temporal, $W = 20$, pretr.	0.495 ± 0.089	64.1 ± 10.0	**83.7 ± 2.8**
CDMix, $W = 20$, pretr.	0.494 ± 0.082	**65.1 ± 9.7**	83.0 ± 2.8
CDMix, $W = 50$, pretr.	0.493 ± 0.097	64.1 ± 12.0	82.6 ± 1.6
CDMix bands, $W = 20$, pretr.	0.486 ± 0.089	63.3 ± 10.3	81.2 ± 3.8
CDMix, $W = 50$, scratch	0.485 ± 0.092	62.3 ± 11.3	82.1 ± 2.2
CDMix pixels, $W = 20$, pretr.	0.480 ± 0.100	61.3 ± 11.4	80.4 ± 4.7
CDMix temporal, $W = 50$, scratch	0.479 ± 0.071	62.4 ± 8.5	82.4 ± 3.3
CDMix, $W = 20$, scratch	0.475 ± 0.066	61.9 ± 7.7	81.5 ± 2.2
CDMix temporal, $W = 20$, scratch	0.474 ± 0.066	62.1 ± 7.7	81.4 ± 3.4
CDMix bands, $W = 50$, pretr.	0.473 ± 0.077	62.0 ± 10.3	82.9 ± 2.1
CDMix pixels, $W = 50$, pretr.	0.469 ± 0.136	59.6 ± 13.8	79.7 ± 4.7
CDMix pixels, $W = 20$, scratch	0.466 ± 0.074	61.5 ± 8.4	80.7 ± 2.4
CDMix pixels, $W = 50$, scratch	0.462 ± 0.107	59.3 ± 9.4	81.2 ± 3.6
CDMix bands, $W = 50$, scratch	0.458 ± 0.073	61.4 ± 8.3	80.4 ± 5.3
CDMix bands, $W = 20$, scratch	0.456 ± 0.076	59.9 ± 8.6	80.0 ± 2.9
Feature-space & normalization alignment			
CORAL, pretr.	0.480 ± 0.077	63.3 ± 8.9	81.4 ± 3.2
Projected align. (MMD), pretr.	0.478 ± 0.079	63.0 ± 9.2	81.4 ± 3.4
DSBN	0.472 ± 0.083	62.0 ± 9.9	81.2 ± 3.4
MMD, pretr.	0.469 ± 0.080	61.8 ± 9.7	81.4 ± 2.7
MMD, scratch	0.430 ± 0.088	58.9 ± 8.6	78.2 ± 4.0
Projected align. (MMD), scratch	0.428 ± 0.085	59.4 ± 9.4	77.3 ± 6.1
CORAL, scratch	0.419 ± 0.081	57.0 ± 9.5	76.9 ± 5.2

[a] Source-domain reference (trained on the 2020–2024 data of the two training regions and evaluated on the 2020–2024 data of the held-out region); not a transfer result. [b] Fold-specific source model (2020–2024, two training regions) applied directly to the 2025 data of the held-out region, with no 2025 labels or parameter updates; zero-shot with respect to both the evaluation region and the target year.

In our case, the feature-space alignment operates on the 128-dimensional global parcel embedding z produced by the TAE (Section 2.5). At each iteration, a source batch and a target batch of equal size are passed through the shared backbone, so the alignment

term is always estimated on balanced batches, and an alignment term is added to the classification loss:

$$\mathcal{L} = \mathcal{L}_{\mathrm{FL}} + \lambda_a \mathcal{L}_{\mathrm{align}}, \tag{6}$$

where λ_a weights the alignment and $\mathcal{L}_{\mathrm{align}} \in \{\mathcal{L}_{\mathrm{MMD}}, \mathcal{L}_{\mathrm{CORAL}}\}$. In all runs, the alignment weight is fixed to $\lambda_a = 1$ and applied throughout training; no systematic sweep over λ_a was performed.

Maximum Mean Discrepancy (MMD) measures the distance between the source and target embedding distributions as the distance between their mean embeddings in a reproducing-kernel Hilbert space [13,27]. For source embeddings $\{z_i^s\}_{i=1}^{n_s}$ and target embeddings $\{z_j^t\}_{j=1}^{n_t}$ in a batch, the squared empirical MMD is

$$\mathcal{L}_{\mathrm{MMD}} = \frac{1}{n_s^2}\sum_{i,i'} k(z_i^s, z_{i'}^s) + \frac{1}{n_t^2}\sum_{j,j'} k(z_j^t, z_{j'}^t) - \frac{2}{n_s n_t}\sum_{i,j} k(z_i^s, z_j^t), \tag{7}$$

where k is a Gaussian radial basis function (RBF) kernel; following common practice, we use a mixture of RBF kernels with several bandwidths to reduce sensitivity to the kernel scale. Minimizing Equation (7) pulls the first-order statistics of the two domains together while the classification loss preserves discriminative power.

Correlation Alignment (CORAL) aligns the second-order statistics of the two domains by matching their feature covariance matrices [14,29]. With C_s and C_t the $d \times d$ covariance matrices ($d = 128$) of the source and target embeddings in a batch,

$$\mathcal{L}_{\mathrm{CORAL}} = \frac{1}{4d^2}\,\|C_s - C_t\|_F^2, \tag{8}$$

where $\|\cdot\|_F$ is the Frobenius norm. CORAL is parameter-free and computationally lighter than MMD, providing a complementary, covariance-based view of feature alignment. Because the target domain is labeled from LPIS, the classification term in Equation (6) can be computed on both domains; the alignment term itself is computed on the marginal embedding distributions.

Since a marginal criterion can be counterproductive under strong label shift, we additionally evaluate two mechanisms that target the diagnosed first-order gap more directly.

Domain-Specific Batch Normalization (DSBN), as an architectural alternative to an explicit alignment loss, implies that every batch-normalization layer of the backbone is duplicated into a domain-specific pair that keeps separate running statistics and affine parameters for the source and target domains, while all other weights remain shared [15]. The active branch is selected before each forward pass, so a per-domain first-order (mean and scale) shift is absorbed by each domain's own normalization with no additional loss term or hyperparameter. Both branches are initialized from the pretrained statistics, and the target branch is used at inference.

Projected Feature Alignment is unlike the plain MMD and CORAL baselines above, which are symmetric and act inside a single shared, fully trainable backbone. The projected variant is asymmetric and uses a distinct configuration, so that although it employs the MMD discrepancy internally, it is not the plain-MMD baseline.

A frozen copy of the pretrained source backbone provides a fixed reference embedding z^s. Next, a learned linear projector $T : \mathbb{R}^{128} \to \mathbb{R}^{128}$ maps it into the target embedding space, and the discrepancy (MMD in our runs) is minimized between $T(z^s)$ and z^t. Only the projector T and the target branch receive gradients, and the classification loss is computed on the labeled target alone. The source anchor therefore never moves, and the target features are never pulled toward the source; instead, the projector learns the mapping that

reconciles the two, closing the first-order gap while preserving the discriminative target structure. A class-conditional variant of the marginal losses is left as a natural extension.

3. Results

3.1. Experimental Protocol

All experiments followed a three-fold *leave-one-region-out* (LORO) protocol, applied identically in the geographic supervised-adaptation scenario and in the temporal scenario of target-year supervised adaptation with previous-year support (Sections 3.2 and 3.3). In each fold, the models are trained using two of the three PixelSet-Slovakia regions and evaluated on the entire remaining region. The held-out region was rotated across the three folds, and performance is reported as the mean ± standard deviation over the three regional folds. This spatial partitioning prevents neighboring parcels from being distributed between the training and validation sets and, because all observations from the held-out region are excluded from training, also prevents the same physical parcels from contributing to both sets across different years. The three study regions are geographically disjoint and separated by at least approximately 29 km, thereby substantially reducing the local spatial dependence between training and validation samples compared with the previously used random parcel-level partition. Models were kept consistent across folds: the same architecture, training procedure, and hyperparameter settings were used in every fold; hyperparameters were shared across folds rather than tuned per fold, and only the stopping epoch was selected by monitoring the held-out region. Therefore, the results should be interpreted as region-held-out validation performance rather than as estimates from an independent test set. Results under the formerly used random 85/15 stratified split are provided in the Supplementary Materials (Tables S1 and S2).

The batch size is 128, with 128 pixels sampled from each parcel at each iteration. Models were trained for up to 100 epochs using Adam with a learning rate of 10^{-3} and Focal Loss with $\gamma = 1$. Performance is assessed using overall accuracy, balanced accuracy, defined as the mean per-class recall, and mean IoU. Given the strong class imbalance, balanced accuracy and mean IoU are used as the primary evaluation metrics, whereas overall accuracy is retained as a complementary measure. Metrics were averaged only over classes represented in the corresponding held-out region and evaluation year. Classes with no validation samples were excluded from the respective macro-average rather than assigned a score of zero. In particular, winter triticale, winter durum wheat, and mixed cereal are absent from the 2020 geographic target (Table 2). To assess sensitivity to stochastic training effects separately from regional variability, the principal configurations in both scenarios are additionally repeated using three random seeds. The results are reported separately in Appendix E.

3.2. Geographic Scenario: France to Slovakia

In the geographic supervised-adaptation scenario, the PixelSet-France dataset from 2017 constitutes the source domain and PixelSet-Slovakia from 2020 the target domain. For each LORO fold, target-domain training uses the 2020 parcels from two Slovak regions, while evaluation is performed on the complete 2020 data of the remaining region. This setting therefore combines geographic and inter-annual differences with changes in spectral–temporal patterns, class prevalence, and the original national label nomenclatures. A separate zero-shot geographic reference was obtained by applying the France-trained source model directly to the Slovak target without target-domain training or parameter updates. Table 3 summarizes the results.

Figure 8 illustrates these results spatially by rejoining the predictions to the LPIS parcel geometries of two held-out study areas and comparing the best CDMix configuration with the target-only reference at the level of individual parcels.



Figure 8. Parcel-level comparison between the best CDMix configuration (temporal, $W = 20$, pretrained) and the target-only reference for the held-out Nitriansky kraj (**left**, R1) and Banskobystrický kraj (**right**, R2). Colors indicate whether both models are correct (green), only CDMix is correct (blue), only the target-only reference is correct (red), or both are incorrect (purple). The maps show LORO validation outputs and do not constitute independently validated crop maps.

The zero-shot row quantifies the severity of the geographic shift. Applied directly to the Slovak target, without any target labels or parameter updates, the France-trained source model reaches an IoU of only 0.064. Direct geographic transfer is therefore not viable in this setting, and target supervision is indispensable. Inspection of the zero-shot predictions shows that only the most frequent and spectrally distinctive classes transfer to a useful degree (meadow, soft winter wheat, and partially corn and leguminous fodder), while the finer classes collapse.

Once Slovak labels are used, source pretraining becomes beneficial rather than harmful: full fine-tuning exceeds training from scratch on every regional fold. Encoder freezing shows, however, that this benefit requires sufficient model plasticity: freezing the full encoder substantially reduces IoU, while selective freezing of the temporal or spatial encoder yields lower IoU than full fine-tuning (Table 3). Under the pronounced geographic shift quantified in Section 3.4, the pretrained representations constitute a useful initialization but cannot be reused rigidly. The source-domain reference row (PixelSet-France) attains a high overall accuracy but a low balanced accuracy, illustrating the strong class imbalance of the source and the inadequacy of overall accuracy alone; it is shown only as context and is not a transfer result.

The CDMix family provides the best configuration of the scenario: CDMix temporal, $W = 20$, pretrained reaches an IoU of 0.475 with the highest balanced accuracy ($61.6 \pm 7.0\%$), improving over training from scratch on every regional fold and over full fine-tuning mainly on the most difficult fold. The gradual λ schedule (Equation (5)) lets the model stabilize on the source early and then migrate smoothly toward the target distribution, while the balanced-accuracy gain over the scratch reference (61.6% against 57.4%) indicates that the improvement extends to the rare and difficult classes rather than merely raising the mean over the dominant ones.

Within the family, the short warm-up ($W = 20$) is generally preferable to the long one ($W = 50$): the three top-ranked variants all use $W = 20$, whereas the long schedule

keeps the model under source-dominated gradients for half of the training budget and destabilizes the spatial-splice variant in particular (CDMix pixels, $W = 50$, 0.409–0.414, with the largest fold-to-fold variance of the family). Among the mixing variants, splicing along the temporal axis ranks highest, while the pixel, band, and linear variants remain close to one another (0.440–0.466); no band variant reaches the top of the family, a physically interpretable outcome, since splicing spectral channels across domains disrupts the inter-band structure on which vegetation discrimination relies. The margins between neighboring variants are nevertheless small relative to the fold-level standard deviations, so the variant ranking should be read as indicative. Across essentially all methods, the held-out Košický kraj (R3) is systematically the most difficult region, and CDMix brings its largest gains precisely on this fold (Appendix D).

3.3. Temporal Scenario: Target-Year Supervised Adaptation with Previous-Year Support

In the target-year supervised-adaptation scenario, the source domain consists of the 2020–2024 observations from the two training regions of each fold, while the target training data consist of their labeled 2025 observations. The evaluation uses the complete 2025 data of the held-out region, compounding inter-annual variability with the unseen-region shift of the LORO protocol. The held-out region enters neither training nor pretraining: for pretrained configurations, the initialization checkpoint is generated per fold from the same 2020–2024 training-region data. Since 2025 labels are available during adaptation, this evaluates supervised target-year adaptation rather than generalization to an unseen year; a complementary zero-shot target-year row applies the 2020–2024 model directly to the held-out region's 2025 data, without any 2025 labels or parameter updates; these data are unseen with respect to both region and year. Table 4 summarizes the results. The two settings answer different questions: the supervised rows measure how much the target season's own labels add on top of historical support, whereas the zero-shot row measures generalization to a season never seen during training.

Here, the weight-transfer family behaves oppositely to the geographic scenario. All weight-transfer configurations exceed training from scratch, and freezing now outperforms full fine-tuning (0.479 ± 0.085): freezing the entire encoder, the configuration that collapsed under geographic shift, reaches 0.498 ± 0.099, and freezing the spatial encoder follows closely (0.486 ± 0.089). Under the temporal shift quantified in Section 3.4, the pretrained representations transfer well, and freezing them acts as beneficial regularization by reducing the number of trainable parameters updated using a single target season.

The best overall configuration of the scenario comes from the CDMix family: CDMix temporal, $W = 50$, pretrained (IoU 0.503 ± 0.095), leading a narrow top band that includes full encoder freezing and pretrained CORAL, all comparable within fold-level variability. As in the geographic scenario, the temporal-splice variant with pretrained initialization ranks first in the family; the ordering of the two warm-up lengths, however, is not stable across scenarios ($W = 20$ ranks first geographically, and $W = 50$ here) and the corresponding differences lie within the fold-level standard deviations, so the warm-up length is best treated as a hyperparameter to be validated per scenario rather than as a quantity that scales with the shift magnitude.

The top of the table is dense: the leading configurations fall within a narrow 0.478–0.503 band that also includes the zero-shot reference, reflecting the moderate nature of the temporal shift, where several strategies are comparable within noise. The cross-year changes are more moderate than in the geographic case, with residual errors concentrated among phenologically close cereals (e.g., winter vs. spring barley, winter durum wheat vs. soft winter wheat), whose separation depends on fine temporal differences that inter-annual variability can shift or compress.

Figure 9 provides the corresponding spatial view for the temporal scenario on the held-out Nitriansky (R1) and Banskobystrický (R2) areas, where CDMix improves over the target-only reference on both folds (an IoU of 0.489 against 0.366 on R1, and 0.604 against 0.513 on R2); the corrected parcels are visible as the blue-dominated pattern of the agreement maps.

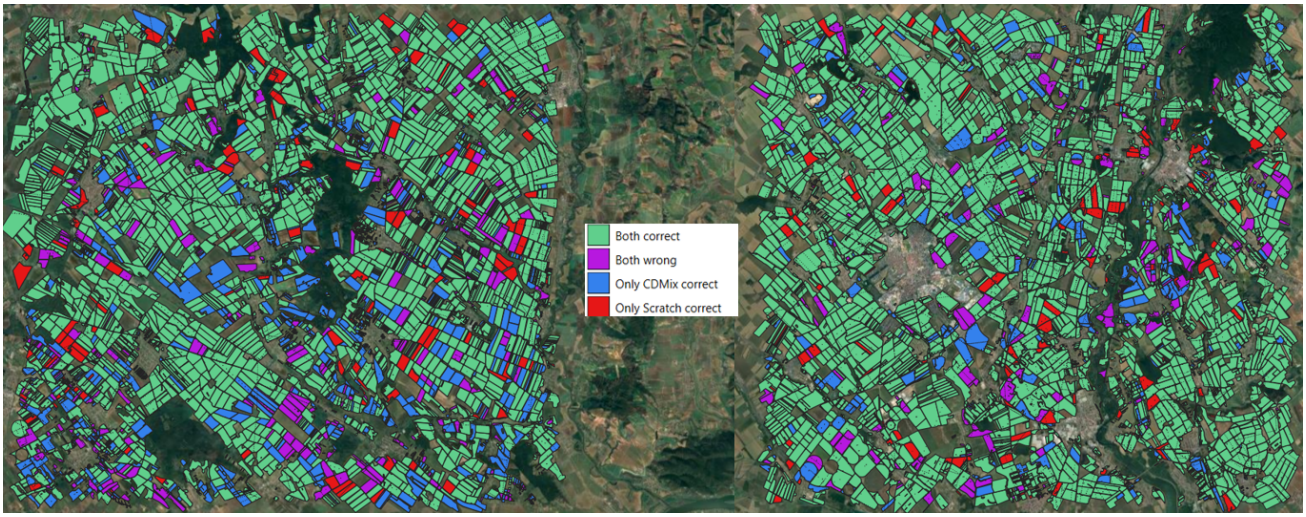


Figure 9. Parcel-level agreement maps for the temporal scenario (Slovakia 2020–2024 → 2025): the best CDMix configuration (temporal, $W = 50$, pretrained) versus the target-only reference for the held-out Nitriansky kraj (**left**, R1) and Banskobystrický kraj (**right**, R2). Colors are as in Figure 8: both correct (green), only CDMix correct (blue), only the target-only reference correct (red), both wrong (purple). The predictions are validation outputs under the LORO protocol and do not constitute an independently validated crop map.

3.4. Analysis of Feature-Space and Normalization-Based Alignment

The alignment family behaves very differently in the two scenarios, and this difference can be traced back to the magnitude and the nature of the two shifts. To quantify them symmetrically, both shifts are measured with the same diagnostics, computed with the same script, the same sample size (1024 embeddings per domain), and the same permutation protocol (500 permutations) on the 128-dimensional parcel embedding z.

The first diagnostic is the proxy $\mathcal{A}$-distance (PAD) [22]: a binary classifier is trained to distinguish source from target embeddings so that high accuracy indicates separable domains, and accuracy near the 50% chance level indicates aligned ones. In the geographic scenario, this classifier reaches about 98% accuracy (PAD $\approx$ 1.9), while in the temporal scenario, it reaches only 59.5% ($\approx$0.38), barely above chance. The normalized MMD tells the same story: 0.169 geographically against 0.0036 temporally, which is roughly 47 times larger, with both discrepancies statistically significant under the permutation test ($p = 0.002$). The temporal shift is therefore detectable but practically negligible, rather than absent. In both scenarios, the source and target covariances already coincide (CORAL distance $\approx 1.6 \times 10^{-5}$), so the gap is almost purely first-order: a shift of the embedding mean, combined with the label-prior shift of Section 2.3.4.

These diagnostics explain the behavior of the family in Table 3. CORAL has essentially nothing to align, so its loss mostly injects gradient noise. MMD draws the two marginal distributions together, while the label priors differ sharply, which pulls target parcels toward source-typical regions of the feature space and erodes class separability; applying either loss on top of pretrained weights is worse than training from scratch, confirming that the marginal penalty competes with the representation rather than refining it. The two mechanisms that address the first-order gap directly recover most of the lost per-

formance: DSBN absorbs the per-domain mean and scale shift architecturally, and the asymmetric projected alignment aligns a learned source-to-target map to a frozen pretrained reference instead of forcing the target onto the source. Neither, however, surpasses the CDMix optimum: once the target is labeled, mixing information in input space transfers more knowledge than a purely feature-level correction.

In the temporal scenario (Table 4), the same family behaves markedly better: all pretrained variants clearly exceed training from scratch and are comparable to full fine-tuning. This reversal is consistent with the diagnostics above. With nearly identical label priors across years, marginal alignment no longer competes with the classification objective and instead acts as useful regularization. Under the present fixed-weight marginal-alignment setup, MMD and CORAL are detrimental in the geographic scenario and more competitive in the temporal scenario. These results are specific to the tested implementation and should not be interpreted as evidence that feature-space alignment methods are generally ineffective.

4. Discussion

Beyond the methodological comparison, this work delivers PixelSet-Slovakia, a multi-year, parcel-level Sentinel-2 dataset with harmonized labels, built through a reproducible Google Earth Engine pipeline, characterized down to the region–year level, and released publicly. On this benchmark, within the fully supervised domain-adaptation setting, the proposed input-space mixing improved the geographic-transfer results over both target-only training and standard weight transfer. Three observations structure the interpretation of these results. The first observation concerns the consistency of the approach. In both scenarios, the strongest CDMix configuration combines pretrained initialization with temporal-splice mixing. Under the pronounced geographic shift, it exceeds both training from scratch and full fine-tuning, while under the mild temporal shift, it ranks first within a narrow band of comparable strategies. This supports, at the parcel level, a convergent observation of recent domain-generalization work: a lightweight adaptation over pretrained representations, paired with a mixing-based augmentation, is preferable to training from scratch. The warm-up length, by contrast, did not order consistently across the two scenarios: $W = 20$ ranked first geographically and $W = 50$ temporally, with differences within fold-level noise. It is therefore best treated as a hyperparameter to be validated per scenario.

The second observation is the opposite behavior of encoder freezing, which offers one of the clearest readings of the results. Under the geographic shift, freezing degrades performance, and full-encoder freezing collapses to 0.313 ± 0.060 because the source representations do not transfer directly. Under the temporal shift, the same mechanism acts as regularization, and full-encoder freezing sits within the top band (0.498 ± 0.099). This reversal is an empirical confirmation of domain-adaptation theory, in which the target error is bounded by the source error and a measure of divergence between domains [22]: as the divergence grows, rigid reuse of the representations becomes riskier and the need for active adaptation more pressing. The zero-shot evaluations measure the same asymmetry directly, with 0.064 IoU for France $\rightarrow$ Slovakia against 0.478 ± 0.066 for 2020–2024 $\rightarrow$ 2025.

The third observation concerns the reading of aggregate metrics. Overall accuracy remains a useful orientation indicator, but it is insufficient on an imbalanced dataset, where a model can score highly by favoring dominant classes. The primary interpretation therefore rests on validation IoU and balanced accuracy, with the per-class metrics used to verify that performance is coherently distributed across classes. These indicate that CDMix improves not only the global mean but also the balanced accuracy over rare classes, especially where the source-to-target transition is controlled gradually. The pooled confusion matrices in Appendix C detail this reading at the class level. In the geographic scenario, the largest recall

gains of the best CDMix configuration over the target-only reference occur for soybeans, the aggregated fruits–vegetables–flowers class, winter barley, meadow, and sunflower, while a few frequent classes (spring barley, corn, and sorghum) lose a small amount of recall. The aggregate improvement is therefore driven by the rarer and harder classes rather than by the dominant cereals. The benefit is concentrated on the most difficult regional fold, where transfer is hardest: on the eastern fold (Košický kraj), CDMix exceeds the target-only reference by 0.111 IoU and full fine-tuning by 0.036 IoU while remaining on par on the two easier folds. Given the fold-level variability, this observation is consistent with, though not proof of, a systematic advantage.

It is worth emphasizing that the target-year adaptation scenario still benefits from the historical data. Fine-tuning, freezing, and CDMix all clearly exceed training from scratch on the target season alone, and even the zero-shot application of the 2020–2024 model outperforms it. The mere proximity of the distributions does not remove the value of the source domain. This conclusion resonates with foundation-model studies [24], where even large-scale pretraining does not suppress the need for a light adaptation to new geographies and seasons.

Limitations

These findings are subject to several limitations. All reported scores are validation figures obtained under the LORO protocol, with the stopping epoch selected on the held-out region and without a fully independent test set. The fold-to-fold standard deviations are large relative to many of the differences between methods. Configurations whose intervals overlap, especially within the narrow top band of the temporal scenario, should therefore be read as comparable within noise rather than strictly ordered. The hyperparameter grid remains fixed and validation-selected, with the warm-up length chosen among $\{20, 50\}$. Ground truth derives solely from LPIS declarations, whose national nomenclature was harmonized to the source label space (Section 2.3.4), so residual differences in class definitions are a potential source of uncontrolled concept shift.

5. Conclusions

In the present work, we studied geographic and temporal generalization for parcel-level crop-type mapping under domain shift. We introduced a new multi-year, parcel-level Sentinel-2 dataset for Slovakia, named PixelSet-Slovakia, built from LPIS declarations and processed into a PixelSet representation through a reproducible Google Earth Engine pipeline.

We introduced a domain-adaptation method named Cross-Domain Mixup (CDMix) and used the PixelSet-Slovakia benchmark across a geographic scenario (France to Slovakia) and a temporal scenario of target-year supervised adaptation with previous-year support to evaluate its performance. The method was compared with three families of alternatives on a common PSE-TAE backbone. The families of adaptation strategies were: no adaptation, weight transfer, and feature-space alignment.

Under the tested PSE-TAE backbone, the selected source–target domains, and the LORO validation protocol, CDMix over pretrained representations (temporal-splice variant) was the best configuration under geographic shift, exceeding both training from scratch and full fine-tuning. In the temporal scenario, CDMix again ranked first, leading a narrow band of comparable strategies, all clearly preferable to training from scratch. The zero-shot evaluations quantify the asymmetry between the two shifts directly, with direct geographic transfer failing entirely, while the zero-shot cross-year evaluation remains competitive with supervised adaptation. In the current setup, explicit feature-space alignment proved the weakest family under geographic shift: marginal MMD and CORAL fell below the

no-adaptation reference, while normalization- and projection-based corrections of the first-order gap recovered fine-tuning-level performance without reaching the CDMix optimum. Together, these results indicate that input-space mixing was the strongest tested strategy under both shifts, with a clear margin under the geographic one, and a narrow one under the temporal one, where several pretrained strategies remain comparable and are all clearly preferable to training from scratch. The study is subject to several methodological constraints, including the use of a single backbone, one source–target pair per scenario, no parcel-boundary erosion, and a validation-selected warm-up length. Regional generalization was assessed only across the three Slovak study regions, rather than across all Slovak or European agricultural landscapes, which further limits the scope of these conclusions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18172857/s1.

Author Contributions: Conceptualization, A.-A.D. and C.F.; methodology, A.-A.D. and C.F.; software, A.-A.D.; validation, C.F.; formal analysis, A.-A.D. and C.F.; data curation, A.-A.D.; writing—original draft preparation, A.-A.D.; writing—review and editing, C.F.; supervision, C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The multi-year PixelSet-Slovakia dataset is archived on Zenodo at https://doi.org/10.5281/zenodo.21885996, and the data-processing and training code is available at https://github.com/AdelinaDns/CDMix (accessed on 21 July 2026). The source-domain PixelSet of [16] is publicly available at https://github.com/VSainteuf/pytorch-psetae (accessed on 18 May 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CAP	Common Agricultural Policy
CDMix	Cross-Domain Mixup
CORAL	Correlation Alignment
DSBN	Domain-Specific Batch Normalization
HCAT	Hierarchical Crop and Agriculture Taxonomy
IoU	Intersection-over-Union
LORO	Leave-One-Region-Out
LPIS	Land Parcel Identification System
MMD	Maximum Mean Discrepancy
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
PSE-TAE	Pixel-Set Encoder with Temporal Attention Encoder
RBF	Radial Basis Function
SITS	Satellite Image Time Series

Appendix A. Label-Space Harmonization Details

The harmonization was performed from the original Slovak crop descriptions available in the annual LPIS layers. These descriptions were translated into English and compared semantically with the classes of the PixelSet-France source dataset. The assignment was based exclusively on the label descriptions and was independent of the Sentinel-2 observations and model predictions. A conservative mapping rule was applied. A Slovak category was retained only when its crop type, seasonal specification, and intended use were sufficiently consistent with the corresponding source class. Labels describing broader categories, different crop uses, undersown variants, or unspecified seasonal forms were

assigned a NULL mapping and excluded. The mapping was initially established for the 2020 nomenclature and was subsequently extended to descriptions appearing in later years, using the same criteria. Table A1 summarizes the correspondence between the source classes and the Slovak LPIS descriptions retained in PixelSet-Slovakia.

Table A1. Class-level correspondence between the PixelSet-France source classes and the Slovak LPIS descriptions retained in PixelSet-Slovakia. The Slovak column reports the original descriptions or summarizes the inclusion rule when several semantically equivalent descriptions occur.

Label	PixelSet-France Class	Retained Slovak LPIS Descriptions	Mapping Rule
1	Meadow	*Trvalý trávny porast*; *Trávy alebo iné bylinné krmoviny*; and equivalent annual grassland descriptions	Permanent grassland and unambiguous grass or herbaceous-forage categories.
2	Soft winter wheat	*Pšenica letná ozimná*; *Pšenica ozimná*	Only descriptions explicitly identifying winter wheat.
3	Corn	*Kukurica*	Only the generic maize category was retained.
4	Winter barley	*Jačmeň ozimný*	Exact seasonal correspondence.
5	Winter rapeseed	*Kapusta repková pravá—ozimná*	Only descriptions explicitly identifying winter rapeseed.
6	Spring barley	*Jačmeň jarný*	Exact seasonal correspondence.
7	Sunflower	*Slnečnica ročná*	Generic sunflower without undersowing.
8	Grapevine	*Vinohrady*	Declared productive vineyards.
9	Beet	*Repa cukrová*	The source class was interpreted as sugar beet.
10	Winter triticale	Descriptions explicitly identifying winter triticale	The generic description *Tritikale* was not retained because the seasonal form was unspecified.
11	Winter durum wheat	*Pšenica tvrdá ozimná*	Only descriptions explicitly identifying winter durum wheat.
12	Fruits, vegetables, flowers	Unambiguous vegetable, small-fruit, horticultural, and flower descriptions	Permanent woody fruit crops were assigned to *Orchard*.
13	Potatoes	*Zemiaky konzumné (skoré)*; *Zemiaky konzumné (neskoré)*	Early and late ware potatoes; seed and undersown variants were excluded.
14	Leguminous fodder	Alfalfa, forage peas, clovers, vetches, and sainfoin explicitly used as forage	Only unambiguous forage-legume descriptions were retained.
15	Soybeans	*Sója fazuľová*	Exact crop correspondence.
16	Orchard	Orchards and unambiguous perennial fruit- or nut-tree descriptions	Grapevines and non-woody small-fruit categories were treated separately.
17	Mixed cereal	Descriptions explicitly identifying a cereal mixture	Generic mixed-crop descriptions were excluded.
18	Sorghum	*Cirok*	Only generic sorghum; forage and undersown variants were excluded.

Several examples illustrate the conservative nature of the mapping. The generic Slovak label *Kukurica* was assigned to the shared Corn class. By contrast, descriptions referring to silage, sweet, seed, green-fodder, or popcorn maize were assigned a NULL mapping. The same decision was applied to maize categories containing an undersowing specification because their intended use or management could differ from that represented by the generic source class.

Seasonal information was treated similarly. The generic description *Tritikale* was assigned a NULL mapping because it does not establish whether the crop corresponds

specifically to Winter triticale. Likewise, *Pšenica tvrdá* was excluded because its seasonal form is unspecified, whereas the explicit description *Pšenica tvrdá ozimná* was retained as Winter durum wheat. Spring rapeseed was not assigned to the Winter rapeseed class, and the generic mixed-crop description *Zmiešaná plodina* was excluded because it does not establish that the mixture consists specifically of cereals. Comparable distinctions were applied to crop purpose. *Repa cukrová* was retained as Beet, whereas fodder-beet descriptions were excluded. Similarly, the generic label *Cirok* was retained as Sorghum, while Sudan grass and undersown sorghum variants were assigned a `NULL` mapping.

The source-domain background and void categories did not correspond to declared Slovak crop classes. Records without a valid retained crop mapping, including records with missing crop descriptions, were excluded from evaluation.

Appendix A.1. Residual Label-System Uncertainty

The mapping provides a common operational prediction space, but it does not establish complete semantic equivalence between the French and Slovak LPIS systems. National nomenclatures may differ in their level of detail, reporting conventions, crop-purpose distinctions, and treatment of mixed or undersown crops. Consequently, part of the observed performance variation may reflect residual label-system mismatch in addition to spectral–temporal and class-distribution shifts. The experiments do not isolate the individual contribution of these factors.

Appendix B. Dataset Characterization and Parcel-Identity Audit

This appendix characterizes PixelSet-Slovakia at the level of the eighteen region–year subsets and summarizes the geometric audit of parcel identity across years.

Table A2 reports, for each region and year, the number of parcel–year instances, the number of represented harmonized classes, the total declared agricultural area, the mean and median parcel size, and the ratio between the most and least frequent represented class (imbalance). Across the eighteen subsets, the total declared area per subset ranges from approximately 46,997 to 67,806 ha, the mean parcel size lies between 11.5 and 16.3 ha, and the median parcel size between 3.6 and 9.0 ha; the annual totals over the three regions range from approximately 168,390 to 178,090 ha. The class composition is stable in its dominant classes—the most frequent class of every subset is either soft winter wheat or corn—while the rarest and absent classes vary: winter triticale appears only in 2024, mixed cereal only in 2025, winter durum wheat only from 2023 onwards, and grapevine is absent from Košický kraj in 2020–2022 and represented by a single parcel per year thereafter.

Appendix B.1. Parcel Identity Across Years and Leakage Considerations

Because LPIS parcels recur across years, the 75,493 parcel–year instances of Table 2 correspond to 30,386 unique physical parcel entities, identified geometrically by matching declared geometries across years (intersection-over-union ≥ 0.75). The mean persistence of an entity is 2.48 years, the year-to-year overlap of entities ranges from 64.1% to 76.5%, and 73.7% of the 2025 instances (9349 of 12,682) correspond to entities already observed in an earlier year. The annual observations of Table 2 are therefore not independent physical parcels. The LORO protocol of Section 3.1 accounts for this structure: because the three regions are pairwise disjoint and separated by at least approximately 29 km, no physical parcel entity can contribute to both the training and the validation partition of any fold, in either scenario.

Appendix B.2. Extraction Pipeline Notes

The processed annual vector layers were uploaded as Google Earth Engine assets and served as the source of geometries and labels, each parcel record carrying an identifier unique within its annual LPIS layer. Extraction was performed by an automated Google Earth Engine pipeline with intermediate storage in Google Cloud Storage, covering 2020–2025. Spectral values were sampled with `sampleRegions()` on the 10 m grid, using `tileScale=8` to avoid memory errors on large parcels; this parameter affects only the internal computation size and not the spatial sampling resolution. Parcels were sorted deterministically by their annual identifier and exported in fixed-size batches; this identifier-based batching introduces no spatial tiling. CSV files were assembled locally into the source-compatible PixelSet format in a streaming manner with periodic checkpointing, and spectral values were stored as `uint16`.

Table A2. Region-by-year characterization of PixelSet-Slovakia: number of parcel–year instances, number of represented harmonized classes, total declared agricultural area, mean and median parcel size, and class-imbalance ratio (most over least frequent represented class). R1 = Nitriansky kraj; R2 = Banskobystrický kraj; R3 = Košický kraj.

Region	Year	Instances	Classes	Area (ha)	Mean (ha)	Median (ha)	Imbalance
R1	2020	4792	15	67,379	14.1	3.6	290.0
R1	2021	4903	15	67,621	13.8	3.6	178.3
R1	2022	4905	15	67,806	13.8	3.6	99.5
R1	2023	4730	16	63,401	13.4	4.7	51.2
R1	2024	4660	17	62,867	13.5	5.8	109.5
R1	2025	4809	17	65,681	13.7	6.1	385.3
R2	2020	3569	15	57,306	16.1	5.0	71.5
R2	2021	3521	15	56,073	15.9	4.6	59.6
R2	2022	3562	15	55,277	15.5	4.5	49.2
R2	2023	3774	16	59,583	15.8	7.6	75.1
R2	2024	3587	17	58,341	16.3	8.2	60.3
R2	2025	3791	17	61,747	16.3	9.0	74.4
R3	2020	4193	14	50,214	12.0	3.9	457.3
R3	2021	4099	14	46,997	11.5	3.9	106.1
R3	2022	4331	14	50,788	11.7	4.0	187.9
R3	2023	4205	16	48,303	11.5	5.3	1421.0
R3	2024	3980	17	47,184	11.9	6.1	842.0
R3	2025	4082	17	50,663	12.4	6.5	1015.0

Appendix C. Per-Class Analysis and Pooled Confusion Matrices

This appendix reports the per-class reading of the main results, using confusion matrices pooled over the three LORO folds. Predictions falling into the source-side background and void categories, which are excluded from the metric averages (Section 2.3.4), are shown as rejections into a residual column; this rejection behavior originates in the target-trained models themselves (the target-only reference sends 1956 of 12,554 geographic-scenario predictions to this column) and is therefore not an artifact of source pretraining.

In the geographic scenario (Figure A1), the largest per-class recall gains of CDMix over the target-only reference concern soybeans (+0.73, from 0.07 to 0.80), the aggregated fruits–vegetables–flowers class (+0.20), winter barley (+0.10), meadow (+0.09), sunflower (+0.08), leguminous fodder (+0.08), and grapevine (+0.07). A few frequent classes lose a small amount of recall (spring barley −0.06, corn −0.04, sorghum −0.04, the latter through confusion with corn), and soft winter wheat is essentially unchanged. The improvement therefore comes from the rarer and harder classes rather than from the dominant cereals, which is consistent with the balanced-accuracy gain reported in Table 3.

In the temporal scenario (Figure A2), the recall gains of the best CDMix configuration over the target-only reference are again concentrated on the less frequent classes: grapevine (+0.39), potatoes (+0.23), leguminous fodder (+0.22), winter barley (+0.14), sorghum (+0.14), sunflower (+0.11), and orchard (+0.11), with soft winter wheat also improving (+0.09); soybeans is the one class that loses recall (−0.09). The 2025-only mixed-cereal class is never recovered by any configuration (recall 0.00), reflecting its minimal support and absence from the historical source years; per the protocol of Section 3.1, such minimally supported classes are interpreted cautiously and do not drive the conclusions.

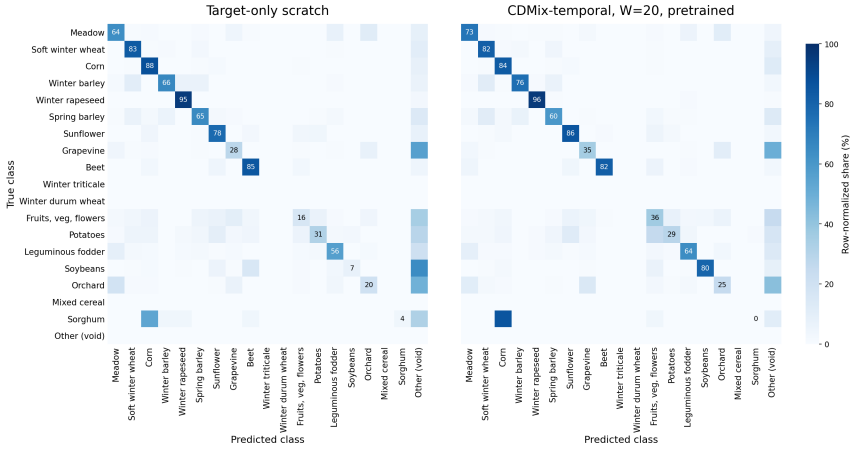


Figure A1. Row-normalized confusion matrices pooled over the three LORO folds for the geographic scenario (France 2017 → Slovakia 2020): target-only reference (**left**) and the best CDMix configuration (temporal, $W = 20$, pretrained; **right**). Rows and columns follow the harmonized classes of Table A1; the *Other (void)* column collects predictions rejected into the residual background and void categories, and empty rows correspond to classes absent from the 2020 target (winter triticale, winter durum wheat, mixed cereal).

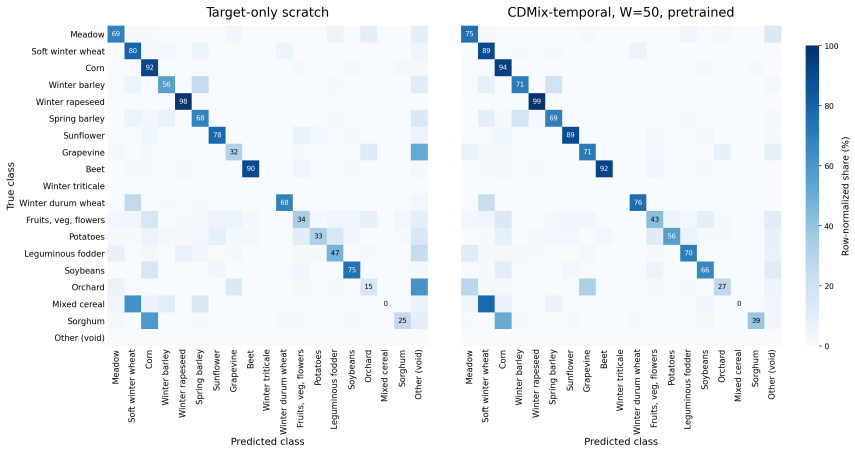


Figure A2. Row-normalized confusion matrices pooled over the three LORO folds for the temporal scenario (Slovakia 2020–2024 → 2025): target-only reference (**left**) and the best CDMix configuration (temporal, $W = 50$, pretrained; **right**). Rows and columns follow the harmonized classes of Table A1; the *Other (void)* column collects predictions rejected into the residual background and void categories, and the winter triticale row is empty because the class is absent from the 2025 target.

Appendix D. Per-Region Validation Results

This appendix reports the validation IoU obtained on each LORO fold, for the principal configurations of Tables 3 and 4: R1 = Nitriansky kraj, R2 = Banskobystrický kraj, R3 = Košický kraj. Each column corresponds to the fold in which that region is held out for

evaluation. All references and non-CDMix configurations are included; for CDMix, whose sixteen configurations differ only in the mixing variant, warm-up length, and initialization, the best warm-up length for each mixing variant and initialization is shown—the omitted configurations follow the same regional ordering. This ordering is consistent across both scenarios and essentially all methods: the eastern region (R3) is systematically the hardest and the central region (R2) the easiest, so the fold-level standard deviations of Tables 3 and 4 are interpreted throughout as regional heterogeneity. Per-region overall and balanced accuracy were not retained for every configuration and are therefore reported only as fold means in the main tables.

Table A3. Per-region validation IoU, geographic scenario (France 2017 → Slovakia 2020). Rows and grouping as in Table 3; for CDMix, the best warm-up length per mixing variant and initialization. Best value per column in bold. The zero-shot and source-reference rows are single evaluations without per-fold values.

Method	R1 IoU	R2 IoU	R3 IoU
No adaptation (reference)			
Scratch (target only)	0.470	0.508	0.273
Weight transfer			
Fine-tune all	0.492	**0.554**	0.348
Freeze TAE (temporal)	0.493	0.508	0.348
Freeze PSE (spatial)	0.479	0.534	0.318
Freeze encoder	0.333	0.361	0.246
Input-space alignment (CDMix)			
CDMix temporal, $W = 20$, pretr. ⋆	0.488	0.553	0.384
CDMix temporal, $W = 20$, scratch	0.492	0.527	**0.388**
CDMix pixels, $W = 20$, scratch	0.474	0.545	0.378
CDMix pixels, $W = 20$, pretr.	0.488	0.534	0.332
CDMix bands, $W = 50$, pretr.	**0.503**	0.522	0.326
CDMix, $W = 20$, pretr.	0.462	0.529	0.353
CDMix bands, $W = 50$, scratch	0.473	0.520	0.348
CDMix, $W = 20$, scratch	0.497	0.498	0.342
Feature-space & normalization alignment			
DSBN	0.501	0.545	0.358
Projected align. (MMD), pretr.	0.493	0.528	0.328
Projected align. (MMD), scratch	0.455	0.498	0.300
CORAL, scratch	0.449	0.469	0.293
MMD, scratch	0.460	0.449	0.258
CORAL, pretr.	0.431	0.450	0.277
MMD, pretr.	0.424	0.414	0.271

⋆ marks the best overall configuration of the scenario.

Table A4. Per-region validation IoU, temporal scenario (Slovakia 2020–2024 → 2025). Rows and grouping as in Table 4; for CDMix, the best warm-up length per mixing variant and initialization. Best value per column (among target rows) in bold.

Method	R1 IoU	R2 IoU	R3 IoU
No adaptation (reference)			
Scratch (target only)	0.366	0.513	0.376
Zero-shot (2020–2024 → 2025)	0.479	0.543	0.411
Source ref. (2020–2024)	0.513	0.561	0.353
Weight transfer			
Freeze encoder	**0.514**	0.588	0.393
Freeze PSE (spatial)	0.475	0.580	0.402
Fine-tune all	0.467	0.569	0.400

Table A4. *Cont.*

Method	R1 IoU	R2 IoU	R3 IoU
Input-space alignment (CDMix)			
CDMix temporal, $W = 50$, pretr. ⋆	0.489	**0.604**	0.416
CDMix, $W = 20$, pretr.	0.482	0.582	0.419
CDMix bands, $W = 20$, pretr.	0.469	0.582	0.407
CDMix, $W = 50$, scratch	0.490	0.575	0.391
CDMix pixels, $W = 20$, pretr.	0.453	0.590	0.397
CDMix temporal, $W = 50$, scratch	0.456	0.559	**0.423**
CDMix pixels, $W = 20$, scratch	0.475	0.535	0.388
CDMix bands, $W = 50$, scratch	0.419	0.542	0.413
Feature-space & normalization alignment			
CORAL, pretr.	0.465	0.563	0.411
Projected align. (MMD), pretr.	0.474	0.558	0.401
DSBN	0.469	0.557	0.391
MMD, pretr.	0.470	0.549	0.390
MMD, scratch	0.402	0.529	0.360
Projected align. (MMD), scratch	0.390	0.525	0.369
CORAL, scratch	0.383	0.512	0.363

⋆ marks the best overall configuration of the scenario.

Appendix E. Seed-Sensitivity Experiment

To separate the sensitivity to stochastic training effects from the regional variability captured by the LORO protocol, the principal configurations of both scenarios were repeated with three random seeds. Each seed controls the network initialization, the batch ordering, and the per-parcel pixel sampling. For every seed, the complete three-fold protocol was rerun and the three-fold mean validation IoU was computed; Table A5 reports this quantity per seed, together with its mean and standard deviation over the three seeds.

Table A5. Seed-sensitivity experiment. Each cell reports the three-fold mean validation IoU obtained with the corresponding random seed; the last column reports the mean ± standard deviation over the three seeds. Seed 1 corresponds to the runs reported in Tables 3 and 4.

Configuration	Seed 1	Seed 2	Seed 3	Mean ± Std
Geographic scenario (France 2017 → Slovakia 2020)				
CDMix temporal, $W = 20$, pretr.	0.475	0.475	0.492	0.481 ± 0.010
DSBN	0.468	0.462	0.435	0.455 ± 0.018
Fine-tune all	0.465	0.456	0.441	0.454 ± 0.012
Scratch (target only)	0.417	0.412	0.393	0.407 ± 0.013
Temporal scenario (Slovakia 2020–2024 → 2025)				
CDMix temporal, $W = 50$, pretr.	0.503	0.497	0.499	0.500 ± 0.003
CORAL, pretr.	0.480	0.480	0.489	0.483 ± 0.005
Fine-tune all	0.479	0.478	0.487	0.481 ± 0.005
Scratch (target only)	0.418	0.435	0.441	0.431 ± 0.012

The seed-induced standard deviation of the three-fold mean IoU lies between 0.003 and 0.018 across all repeated configurations—substantially (roughly 3–10×) below the fold-to-fold standard deviations reported in Tables 3 and 4 (approximately 0.06–0.16 IoU). Moreover, the best configuration of each scenario is identical under every seed, and the relative ordering of the repeated configurations is preserved across seeds with a single

exception: the two nearly tied baselines of the geographic scenario, DSBN and full fine-tuning, swap places under seed 3. The reported uncertainty is therefore dominated by regional heterogeneity rather than by stochastic training effects.

References

1. Weiss, M.; Jacob, F.; Duveiller, G. Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **2020**, *236*, 111402. [CrossRef]
2. Pluto-Kossakowska, J. Review on multitemporal classification methods of satellite images for crop and arable land recognition. *Agriculture* **2021**, *11*, 999. [CrossRef]
3. European Union. Commission Implementing Regulation (EU) 2018/746 Amending Regulation (EU) No 809/2014. 2018. Available online: https://eur-lex.europa.eu/eli/reg_impl/2018/746/oj/eng (accessed on 18 May 2025).
4. European Union Agency for the Space Programme (EUSPA). OBSERVER: How Copernicus Helps Implement the Common Agricultural Policy. 2022. Available online: https://eu-space.europa.eu/news/observer-how-copernicus-helps-implement-cap (accessed on 18 May 2025).
5. Regulation (EU) 2021/2116 of the European Parliament and of the Council of 2 December 2021 on the Financing, Management and Monitoring of the Common Agricultural Policy and Repealing Regulation (EU) No 1306/2013. 2021. Available online: https://eur-lex.europa.eu/eli/reg/2021/2116/oj/eng (accessed on 18 May 2025).
6. Directive (EU) 2023/2413 of the European Parliament and of the Council of 18 October 2023 Amending Directive (EU) 2018/2001 as Regards the Promotion of Energy from Renewable Sources (RED III). 2023. Available online: https://eur-lex.europa.eu/eli/dir/2023/2413/oj/eng (accessed on 18 May 2025).
7. Regulation (EU) 2023/1115 of the European Parliament and of the Council of 31 May 2023 on the Making Available on the Union Market of Certain Commodities and Products Associated with Deforestation and Forest Degradation (EUDR). 2023. Available online: https://eur-lex.europa.eu/eli/reg/2023/1115/oj/eng (accessed on 18 May 2025).
8. Jänicke, C.; Petersen, K.A.; Schmidts, P.; Müller, D.; Jepsen, M.R. Field and farm-level data on agricultural land use for the European Union. *Sci. Data* **2025**, *12*, 1050. [CrossRef] [PubMed]
9. Leonhardt, H.; Wesemeyer, M.; Eder, A.; Hüttel, S.; Lakes, T.; Schaak, H.; Seifert, S.; Wolff, S. Use cases and scientific potential of land use data from the EU's Integrated Administration and Control System: A systematic mapping review. *Ecol. Indic.* **2024**, *167*, 112709. [CrossRef]
10. Schneider, M.; Schelte, T.; Schmitz, F.; Körner, M. EuroCrops: The Largest Harmonized Open Crop Dataset Across the European Union. *Sci. Data* **2023**, *10*, 617. [CrossRef] [PubMed]
11. Andrei, B.; Dinescu, A.A.; Cândea, S.C.; Florea, C. Time Series Mixup for Crop Segmentation in Domain Adaptation. In Proceedings of the 2025 IEEE International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS), Bucharest, Romania, 2–4 September 2025; pp. 1–4. [CrossRef]
12. Xu, M.; Zhang, J.; Ni, B.; Li, T.; Wang, C.; Tian, Q.; Zhang, W. Adversarial domain adaptation with domain mixup. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 6502–6509.
13. Long, M.; Cao, Y.; Wang, J.; Jordan, M.I. Learning Transferable Features with Deep Adaptation Networks. In Proceedings of the International Conference on Machine Learning (ICML), Lille, France, 6–11 July 2015; pp. 97–105.
14. Sun, B.; Feng, J.; Saenko, K. Correlation alignment for unsupervised domain adaptation. In *Domain Adaptation in Computer Vision Applications*; Springer: Cham, Switzerland, 2017; pp. 153–171.
15. Chang, W.G.; You, T.; Seo, S.; Kwak, S.; Han, B. Domain-Specific Batch Normalization for Unsupervised Domain Adaptation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7346–7354. [CrossRef]
16. Garnot, V.S.F.; Landrieu, L.; Giordano, S.; Chehata, N. Satellite image time series classification with pixel-set encoders and temporal self-attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 12325–12334.
17. Sainte Fare Garnot, V.; Landrieu, L. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021.
18. Rußwurm, M.; Pelletier, C.; Zollner, M.; Lefèvre, S.; Körner, M. BreizhCrops: A Time Series Dataset for Crop Type Mapping. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Nice, France, 31 August–2 September 2020; Volume XLIII-B2-2020, pp. 1545–1551. [CrossRef]
19. Băicoianu, A.; Plajer, I.; Debu, M.; Stefan, M.; Ivanovici, M.; Florea, C.; Cataron, A.; Coliban, R.; Popa, S.; Oprisescu, S.; et al. DACIA5: A Sentinel-1 and Sentinel-2 dataset for agricultural crop identification applications. *Big Earth Data* **2025**, *9*, 1226–1257. [CrossRef]

20. Sykas, D.; Sdraka, M.; Zografakis, D.; Papoutsis, I. A sentinel-2 multiyear, multicountry benchmark dataset for crop classification and segmentation with deep learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 3323–3339. [CrossRef]
21. Weikmann, G.; Paris, C.; Bruzzone, L. Timesen2crop: A million labeled samples dataset of sentinel 2 image time series for crop-type classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 4699–4708. [CrossRef]
22. Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; Vaughan, J.W. A theory of learning from different domains. *Mach. Learn.* **2010**, *79*, 151–175. [CrossRef]
23. Ma, Y.; Chen, S.; Ermon, S.; Lobell, D.B. Transfer learning in environmental remote sensing. *Remote Sens. Environ.* **2024**, *301*, 113924. [CrossRef]
24. Gadiraju, K.K.; Vatsavai, R.R. Remote sensing based crop type classification via deep transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4699–4712. [CrossRef]
25. Capliez, E.; Ienco, D.; Gaetano, R.; Baghdadi, N.; Salah, A.H. Temporal-Domain Adaptation for Satellite Image Time-Series Land-Cover Mapping With Adversarial Learning and Spatially Aware Self-Training. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 3645–3675. [CrossRef]
26. Broni-Bediako, C.; Xia, J.; Yokoya, N. Unsupervised domain adaptation architecture search with self-training for land cover mapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 17–21 June 2024; pp. 543–553.
27. Wang, Y.; Feng, L.; Zhang, Z.; Tian, F. An unsupervised domain adaptation deep learning method for spatial and temporal transferable crop type mapping using Sentinel-2 imagery. *ISPRS J. Photogramm. Remote Sens.* **2023**, *199*, 102–117. [CrossRef]
28. Zhou, X.; Prasad, S. Deep feature alignment neural networks for domain adaptation of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5863–5872. [CrossRef]
29. Xu, T.; Han, B.; Li, J.; Du, Y. Domain-invariant feature and generative adversarial network boundary enhancement for multi-source unsupervised hyperspectral image classification. *Remote Sens.* **2023**, *15*, 5306. [CrossRef]
30. Agricultural Paying Agency (Pôdohospodárska Platobná Agentúra). Usage Limits: Usage Limits Declared by Applicants for Direct Support. National Open Data Catalog (Slovak Republic). 2025. Available online: https://data.gov.sk/set/cc261225-7153-44a3-8ebf-05af207515c9 (accessed on 5 August 2024).
31. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
32. Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond Empirical Risk Minimization. *arXiv* **2018**, arXiv:1710.09412.
33. Guo, H.; Mao, Y.; Zhang, R. Mixup as locally linear out-of-manifold regularization. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 3714–3722.
34. Berthelot, D.; Carlini, N.; Goodfellow, I.; Papernot, N.; Oliver, A.; Raffel, C.A. Mixmatch: A holistic approach to semi-supervised learning. In Proceedings of the Advances in Neural Information Processing Systems 32 (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.
35. Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 806–813.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.