arXiv:2610.37021v1 [cs.CV] 27 Oct 2026

Ground-to-Satellite Knowledge Distillation for Crop Classification under Temporal and Geographic Shift

Miro Miranda[1,*], **Rocco Sedona**[1,*], **Valerio Marsocci**[2], **Gabriele Cavallaro**[1,3], **Claudia Paris**[4]

[1]Jülich Supercomputing Centre, Forschungszentrum Jülich, Germany
[2]European Space Agency (ESA/ESRIN), Italy
[3]University of Iceland, Iceland
[4]University of Twente (ITC), the Netherlands

Abstract

Continuous and up-to-date crop type mapping is essential for agricultural monitoring but is often hampered by limited reference data. The growing availability of ground-level field photographs offers a promising way to address this challenge. When combined with satellite time series, these images provide complementary perspectives on the same parcel, capturing crop characteristics from distinctly different perspectives. However, field photographs are not always available for each satellite pixel at inference time, limiting their potential in scalable crop type mapping. We investigate whether ground photographs can provide privileged supervision for crop classification built on pretrained satellite embeddings. A photo teacher based on frozen CLIP features supervises classifiers trained on AlphaEarth and TESSERA representations; photographs are used only during training. On a paired field photo and satellite imagery dataset with 17 crop classes, knowledge distillation consistently improves over satellite-only classification; for frozen embeddings the margin grows under temporal shift, reaching up to $+5.5$ macro-F1 points, and widens in geographic holdouts ($+3.7$ vs $+3.0$ in-distribution). These results show that ground-view supervision can complement satellite embeddings and improve their robustness without changing inference requirements.

1 Introduction

Ground photographs offer a complementary perspective to satellite data and are becoming increasingly available for crop monitoring through both crowdsourcing platforms [Garcin et al., 2021] and field data collection protocols [Nakalembe et al., 2025, d'Andrimont et al., 2022]. For instance, the Land Use/Cover Area Frame survey (LUCAS) pairs georeferenced survey points across Europe with ground photographs, providing fine-grained crop characteristics such as leaf shape, structure, and texture that are difficult to capture from satellite data alone [d'Andrimont et al., 2022]. Such information can remain informative when satellite signature change across space or time and thus contributes to robust crop type mapping [Abbas et al., 2025]. However, the drawback is obvious: ground photographs are limited to surveyed locations and are rarely available on demand, which makes them impractical when rapid, scalable inference is required. This is a learning-using-privileged-information setting [Vapnik and Vashist, 2009], for which Knowledge Distillation (KD) is a practical implementation [Hinton et al., 2015, Lopez-Paz et al., 2016]. We hypothesize that paired ground photographs provide a complementary semantic view of crop type that can improve robustness when scaling across regions and years [Wang et al., 2025]. From an operational perspective, crop type maps require frequent

*Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026). Workshop: 2nd Workshop on Advances in Representation Learning for Earth Observation.

updating to account for the highly dynamic nature of agricultural systems.
Recent advances in geospatial foundation-models (GFMs) offer the potential to streamline data preparation and substantially reduce Earth Observation (EO) data requirements by leveraging frozen embeddings for continuous mapping. However, limited reference data remains a key constraint, as classifiers trained on labels from specific geographic and temporal domains can suffer from distribution shifts, leading to misclassification [Wang et al., 2025]. Such shifts arise because satellite observations are influenced not only by crop characteristics but also by climate, phenology, land management, and acquisition conditions [Wang et al., 2025], causing the same crop to exhibit different satellite signatures across domains. Ground photographs provide a complementary view of the same parcel that is less affected by these factors. Distilling the ground view into the classifier may therefore anchor predictions to crop-relevant cues that are more consistent across geographic and temporal domains. This is particularly relevant when classifiers are trained on frozen GFM embeddings, where the downstream model has limited ability to compensate for region- or time-specific biases encoded in the representation.
Building on the hypothesis that ground photographs provide complementary semantic information about crops, we investigate whether this information can be transferred to satellite-based classifiers to improve their robustness across geographic and temporal domains. Because ground photographs are not consistently available across space and time, we use a ground photo teacher only during training. Ground photographs are encoded with CLIP ViT-B/32 [Radford et al., 2021], pretrained on approximately 400 million web-collected image–text pairs; a small photo teacher classifier is trained per seed and then frozen. Satellite students learn from the labels together with the teacher's logits and embeddings. At inference, the student receives satellite data only. We keep the teacher and loss fixed while changing the satellite branch, considering: (i) frozen embeddings from AlphaEarth [Brown et al., 2025] and TESSERA [Feng et al., 2025], (ii) sequence transformers trained from scratch on monthly satellite time series, and (iii) a partially fine-tuned TESSERA encoder.

To assess the effectiveness of the proposed multimodal framework, we make three controlled comparisons. First, under the same teacher and training budget, we compare the KD–Cross Entropy (CE) difference across satellite models based on frozen and from-scratch representations. Second, we evaluate these models under temporal transfer (2018→2022) and, for AlphaEarth, held-out ecoregions, i.e., geographically distinct ecological regions [Dinerstein et al., 2017]. Third, we test partial fine-tuning of TESSERA while explicitly quantifying the mismatch introduced by our monthly-composite processing of satellite data. Across the main frozen and from-scratch regimes, KD consistently improves in-domain macro-F1 across the satellite branches, by up to 5.1 points on frozen TESSERA, and the frozen-embedding margins grow further under temporal shift and in the Mediterranean geographic holdout.

2 Method

Photo Teacher. A frozen CLIP ViT-B/32 encoder [Radford et al., 2021] maps ground photographs to a feature vector v. Each views is individually L2-normed and averaged to create a single 512-dimensional representation over the location. For each seed, a small photo teacher head is trained on the training split and then frozen: a two-layer projection to the shared 128-dimensional embedding $e_t(v)$ and a one-hidden-layer classifier producing the crop logits $l_t(e_t(v))$. The CLIP encoder itself remains frozen; the head is trained because frozen CLIP provides neither the shared embedding nor crop-class logits, the two distillation targets. Teacher targets are cached only for the student-training points.

Satellite Student. Each satellite student maps the input z to a shared dimensional embedding $e_s(z)$ and logits $l_s(e_s(z))$.

We use (a) a small MLP head on frozen GFM embeddings, mirroring the teacher: a two-layer projection (width 256, LayerNorm, GELU) to $e_s(z)$ and a one-hidden-layer classifier on concatenated z and $e_s(z)$; (b) a sequence transformer trained from scratch on monthly composites; and (c) TESSERA v1.1 [Feng et al., 2025], unfreezing the last two transformer blocks per modality and the fusion reducer. For (c), the encoder and head learning rates are 10^-4 and 10^-3, respectively; earlier layers remain frozen. Following standard knowledge distillation [Hinton et al., 2015], with class label y, temperature T, and cross-entropy CE, the loss is:

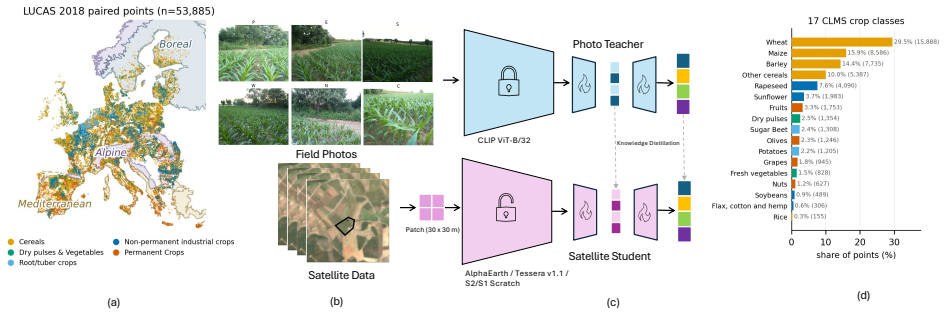


Figure 1: Dataset and model overview. (a) Spatial distribution of the selected agricultural LUCAS 2018 points, including the Alpine, Mediterranean, and Boreal held-out ecoregions. (b) Example of the Sentinel-2 annual composite (S2), the *in-situ* photograph with the four cardinal views (E, W, N, S) and a close-up photograph (C). (c) KD workflow including photo teacher and satellite distilled student model. (d) Distribution of the 17 CLMS crop classes.

$$\mathcal{L} = \mathrm{CE}(l_s, y) + \lambda_{\mathrm{KL}} T^2 \,\mathrm{KL}(\sigma(l_t/T) \,\|\, \sigma(l_s/T)) + \lambda_{\cos}\,(1 - \cos(e_s, e_t))\,. \tag{1}$$

The CE control uses an identical student, optimization, and class-weighted CE term, with both distillation weights set to zero. Thus each reported KD–CE contrast changes only the privileged supervision. More implementation details are provided in Appendix A.

3 Experimental setup

Data. To test the proposed approach, we use LUCAS 2018 survey points with ground photographs [d'Andrimont et al., 2022] for training, focusing on the 17 classes of the Copernicus High Resolution Layer Croplands (CLMS) products [Fahl et al., 2026]. The paired dataset contains 53,885 points with a label, photo, and time series of satellite data. The canonical split follows 70/15/15 for train/validation/test and is fixed across the main comparisons. The 2×2 km LUCAS grid ensures spatially separated, non-overlapping observations, reducing potential spatial dependence between splits. Satellite inputs are either released 64-D AlphaEarth [Brown et al., 2025] and TESSERA [Feng et al., 2025] embeddings, or 12 monthly composites of Sentinel-2 (10 bands) and Sentinel-1 (VV/VH) over a 3×3 pixel patch (30m$\times$30m) centered around the LUCAS point. Figure 1 provides an overview of the spatial distribution of crop types (a), Sentinel-2 composites, the *in-situ* photograph, and the crop type distribution (d).
Temporal transfer. Models trained on LUCAS 2018 are evaluated without adaptation on LUCAS 2022. The class mapping and normalization constants fitted on 2018 are reused, and geographical locations belonging to the 2018 train or validation sets are excluded from the 2022 transfer set. The resulting transfer set contains 36,428 points with a satellite feature and one of the 17 classes, at locations unseen during training.
Geographic transfer. Geographic transfer is evaluated entirely within 2018, isolating spatial from temporal domain shift. EEA ecoregions define fixed geographic test domains [European Environment Agency, 2016]: Mediterranean, Alpine, and Boreal are held out in turn, while the remaining 2018 points are split into training and validation sets. CE and KD are paired on the same split. Our input construction does not follow TESSERA's inference pattern: irregular acquisitions are reduced to 12 monthly composites, missing months become mean-valued observations with a valid mid-month day-of-year, and orbit-merged S1 uses averaged ascending/descending normalization statistics. A linear probe scores 36.2 ± 1.6 macro-F1 on embeddings from this frozen monthly-input path and 43.6 ± 1.6 on the released TESSERA embeddings, a -7.4 point difference. We therefore treat the released embedding and our monthly-input setting as separate regimes rather than using their difference in macro-F1 as an estimate of the effect of fine-tuning. The reported main experiments use five seeds. Macro-F1 is computed on the same evaluation set for each paired CE/KD run, using the fixed 17-label taxonomy.

Table 1: Macro-F1 (%), mean over five seeds, for matched CE and KD students. All students are trained on 2018 only; the transfer block evaluates them on the 2022 survey (unseen locations; Appendix A). Δ is the KD$-$CE mean of per-seed paired differences, computed before rounding. ΔT is the difference to teacher performance (69.7 macro-F1).

	2018 test				2018→2022 transfer		
Satellite branch	CE	+KD	Δ	ΔT	CE	+KD	Δ
Frozen AlphaEarth	63.5	66.5	+2.9	−3.2	52.6	56.0	+3.4
Frozen TESSERA	66.7	71.7	+5.1	+2.0	57.6	63.1	+5.5
S2 SITS (scratch)	62.0	65.2	+3.2	−4.5	54.0	56.4	+2.4
S1+S2 SITS (scratch)	66.1	68.7	+2.6	−1.0	57.5	58.7	+1.2
TESSERA fine-tuned	71.2	**72.1**	+0.8	+2.4	63.0	**63.6**	+0.7

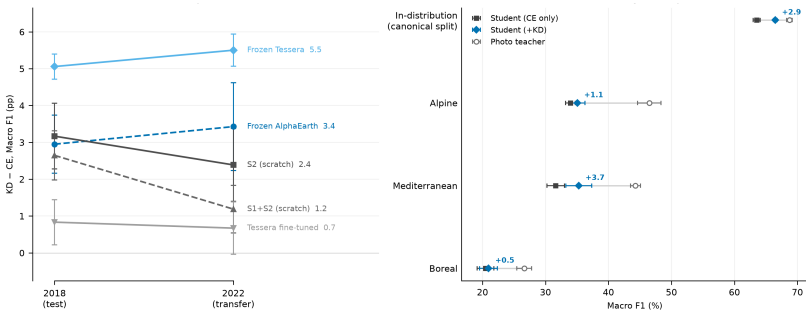


Figure 2: Left: Overall paired KD$-$CE macro-F1 margins per branch from 2018→2022. Margins increase for frozen models, decrease for scratch models, and remain below one point for fine-tuned models. Right: AlphaEarth macro-F1 gains in held-out ecoregions. Only the Mediterranean exceeds the in-distribution gain (+3.7 vs. +2.9); Alpine gains +1.1, while Boreal shows unstable improvements (+0.5, SD 1.4, mixed signs). Error bars indicate ± 1 SD, paired for margins.

4 Results and Discussion

Table 1 shows a consistent in-domain improvement of 2.6–5.1 macro-F1 points for the frozen and from-scratch students, largest on frozen TESSERA. Partial TESSERA fine-tuning (Sec. 2) gives the best absolute result, 72.1 macro-F1, with up to +2.4 macro-F1 over the teacher, but the paired KD margin is smaller at +0.8 macro-F1. Because this branch uses the monthly input path above, this margin is not a controlled estimate of the effect of fine-tuning alone.

Transfer Experiments. Figure 2 (left) shows the paired KD$-$CE margin for every branch under temporal transfer: the frozen margins grow (+2.9 to +3.4 macro-F1 points for AlphaEarth, +5.1 to +5.5 for TESSERA), the scratch margins shrink (+3.2 to +2.4 and +2.6 to +1.2), and the partially fine-tuned TESSERA branch stays below one point. Geographic holdout (right) widens the margin in Mediterranean (+3.7 vs +3.0 in-distribution); Alpine is positive but smaller (+1.1), and Boreal is unstable (+0.5, sd 1.4, mixed signs); Mediterranean is thus the only clear geographic gain.

Discussion. Ground-view KD helps most clearly when the satellite representation is fixed, and its margin can grow under domain shift. Operationally, the photo teacher is needed only during training; deployment requires a satellite-only head on the same embedding product. KD also outperforms alternative cross-modal objectives such as contrastive learning (Appendix C). The comparison has limits. First, TESSERA fine-tuning uses the dataset's monthly composites, for input consistency with the scratch transformers, whereas the frozen TESSERA model uses released embeddings produced through a different input path, a mismatch the −7.4 probe difference quantifies. Second, the geographic results are clear in the Mediterranean holdout, Boreal shows how quickly a small holdout can become noisy. Lastly, the transfer performance depend on the class itself, especially for the frozen TESSERA (Appendix D).

5 Conclusion

Ground photographs can provide privileged supervision for satellite crop classification without being required at inference. Their benefit is strongest for fixed pretrained satellite representations and persists, or increases, under temporal and geographic shift. Our ablations show that this gain is driven primarily by distillation of the teacher's predictive structure rather than by cross-modal feature alignment. The gain is consistent across all satellite branches with frozen-embedding margins grow further under temporal shift and in the Mediterranean holdout. This knowledge-distillation framework is particularly powerful from an operational perspective: field observations can be used to improve satellite models without imposing field-data requirements on large-scale deployment. This is especially relevant for scaling agricultural mapping and monitoring, where satellite imagery enables broad and repeated coverage but ground observations are costly to collect.

Acknowledgements

This work was supported by the European Space Agency (ESA) Φ-Lab as part of the FAST-EO project, under Contract No. 4000143501/23/I-DT and the the AgroVision project, under contract No. 4000147606. The authors gratefully acknowledge the computing time granted on the supercomputer JUWELS Booster at Jülich Supercomputing Centre (JSC).

References

Adel Abbas, Claudia Paris, and Andrew Nelson. Multimodal satellite-field data integration for scalable crop-type mapping with limited labels. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–12, 2025.

Christopher F. Brown, Michal R. Kazmierski, Valerie J. Pasquarella, William J. Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, Noel Gorelick, Lihui Lydia Zhang, Sophia Alj, Emily Schechter, Sean Askay, Oliver Guinan, Rebecca Moore, Alexis Boukouvalas, and Pushmeet Kohli. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025.

Raphaël d'Andrimont, Momchil Yordanov, Laura Martinez-Sanchez, Peter Haub, Oliver Buck, Carsten Haub, Beatrice Eiselt, and Marijn van der Velde. LUCAS cover photos 2006–2018 over the EU: 874,646 spatially distributed geo-tagged close-up photos with land cover and plant species label. *Earth System Science Data*, 14:4463–4477, 2022. doi: 10.5194/essd-14-4463-2022.

Eric Dinerstein, David Olson, Anup Joshi, Carly Vynne, Neil D Burgess, Eric Wikramanayake, Nathan Hahn, Suzanne Palminteri, Prashant Hedao, Reed Noss, et al. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67(6):534–545, 2017.

European Environment Agency. Biogeographical regions, Europe 2016. Dataset, version 1, 2016. `https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3`.

Fernando Fahl, Marijn van der Velde, Martin Claverie, Momchil Yordanov, Laura Martinez Sanchez, and Emanuele Lugato. Europe wide agricultural land use systems by harmonizing the copernicus high resolution layers on grassland and croplands. *Scientific Data*, 2026.

Zhengpeng Feng, Clement Atzberger, Sadiq Jaffer, Jovana Knezevic, Silja Sormunen, Robin Young, Madeline C. Lisaius, Markus Immitzer, David A. Coomes, Anil Madhavapeddy, Andrew Blake, and Srinivasan Keshav. TESSERA: Temporal embeddings of surface spectra for earth representation and analysis. *arXiv preprint arXiv:2506.20380*, 2025.

Camille Garcin, alexis joly, Pierre Bonnet, Antoine Affouard, Jean-Christophe Lombardo, Mathias Chouet, Maximilien Servajean, Titouan Lorieul, and Joseph Salmon. Pl@ntnet-300k: a plant image dataset with high label ambiguity and a long-tailed distribution. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021. URL `https://openreview.net/forum?id=eLYinD0TtIt`.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 18661–18673, 2020.

David Lopez-Paz, Léon Bottou, Bernhard Schölkopf, and Vladimir Vapnik. Unifying distillation and privileged information. In *International Conference on Learning Representations*, 2016.

Catherine Nakalembe, Ivan Zvonkov, Hannah Kerner, Diana Botchway Frimpong, Kenneth Mwangi, Jane Kioko, Bhanu Tokas, Kartik Jawanjal, Iman Akhtar Smith, Anjali Paliyam, et al. Helmets labeling crops: Kenya crop type dataset created via helmet-mounted cameras and deep learning. *Scientific Data*, 12(1):1496, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Vladimir Vapnik and Akshay Vashist. A new learning paradigm: Learning using privileged information. *Neural Networks*, 22(5–6):544–557, 2009.

Hengbin Wang, Yu Yao, Junyi Liu, Xindan Zhang, Yuanyuan Zhao, Shaoming Li, Zhe Liu, Xiaodong Zhang, and Yelu Zeng. Unsupervised cross-regional and cross-year adaptation by climate indicator discrepancy for crop classification. *Journal of Remote Sensing*, 5:0439, 2025.

A Additional experimental details

Sequence Transformer. The time series of satellite images is encoded using a lightweight sequence transformer model. Monthly spectral features are linearly projected into a shared latent space. The sequence is processed by pre-normalized Transformer blocks with multi-head self-attention, feed-forward layers, residual connections, and dropout. Mean pooling across all months, followed by LayerNorm, produces a single sequence representation.

Photo Teacher Precomputed photo features are mapped into a shared latent space using a two-layer projection head with LayerNorm and GELU activation. The resulting embedding is passed to an MLP classifier for crop type prediction. The trained photo teacher is frozen during student training and provides both embedding- and logit-level supervision.

Satellite student. Precomputed satellite features are transformed into the teacher's latent space using an equivalent two-layer projection head. The projected embedding is concatenated with the original satellite feature vector and passed to an MLP classifier. Training combines supervised cross-entropy with logit- and embedding-level KD from the frozen photo teacher.

Main KD configuration. The main KD runs use temperature $T = 2$, logit-KD weight 1.0, embedding-cosine weight 0.25, and the same class-weighted CE term as the matched CE control. The student architecture, optimizer, split, and seed are paired within each KD–CE comparison.

Transfer controls. The 2022 evaluation reuses the 2018 class map and scaler and excludes points used for 2018 training or validation. In the geographic experiments the held-out ecoregion is fixed as test data; only the train/validation partition of the remaining points changes with seed.

TESSERA comparison. The frozen released-embedding experiment and the partially fine-tuned monthly-input experiment are deliberately reported as different regimes. A direct test of the effect of fine-tuning would require both CE and KD with the encoder frozen on the same monthly inputs.

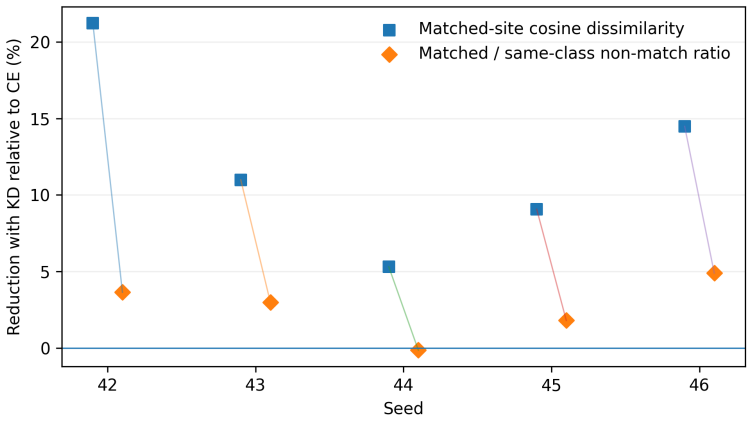


Figure 3: Temporal representation stability across five training seeds, computed on the 1,046 label-stable locations observed in both surveys. Squares: relative reduction from CE to KD, $\Delta_{\text{site}} = 100(D_{\text{CE}} - D_{\text{KD}})/D_{\text{CE}}$, where D is the matched-location dissimilarity of Eq. (2), i.e. the cosine dissimilarity between the 2018 and 2022 representations of the same location, averaged within each crop class and then across classes. Diamonds: relative reduction $\Delta_{\text{ratio}} = 100(R_{\text{CE}} - R_{\text{KD}})/R_{\text{CE}}$ of the matched-to-nonmatched ratio R of Eq. (3), i.e. the same matched-location dissimilarity divided by the mean cross-year dissimilarity between a 2018 location and the other 2022 locations of its crop, so that a reduction in R indicates location-specific stability beyond increased class compactness.

B Temporal representation stability

We further examined whether ground-view distillation changes the temporal stability of the learned satellite representation. This analysis was performed on the partially fine-tuned TESSERA student, using the encoder features before the projection and classification heads. We matched the 2018 and 2022 LUCAS surveys by point identifier and retained only locations observed in both years whose crop label was unchanged. This resulted in 1,046 repeated locations from 10 crop classes; locations whose crop label changed between surveys were excluded because a change in representation is expected in that case.

For a retained location i, we quantify the change between years by the cosine dissimilarity

$$d_{\cos}\left(\mathbf{h}_i^{2018}, \mathbf{h}_i^{2022}\right) = 1 - \frac{\mathbf{h}_i^{2018} \cdot \mathbf{h}_i^{2022}}{\|\mathbf{h}_i^{2018}\|_2 \, \|\mathbf{h}_i^{2022}\|_2}, \tag{2}$$

where $\mathbf{h}_i^y$ denotes the student representation of location i in year y. We first average this quantity within each crop class and then macro-average over classes, such that frequent crops do not dominate the measurement. As shown in Fig. 3, KD reduced the matched-location cross-year dissimilarity by 12.2% on average across the five training seeds, with a reduction for every seed.

Because lower matched-location dissimilarity can also arise from increased within-class compactness, this quantity alone does not isolate a location-specific temporal effect. We therefore used a same-class control. For each crop class c, the matched dissimilarity was compared with the mean cross-year dissimilarity between a 2018 location and all *other* retained 2022 locations of the same crop:

$$R_c = \frac{\frac{1}{|\mathcal{I}_c|} \sum_{i \in \mathcal{I}_c} d_{\cos}\left(\mathbf{h}_i^{2018}, \mathbf{h}_i^{2022}\right)}{\frac{1}{|\mathcal{I}_c|} \sum_{i \in \mathcal{I}_c} \frac{1}{|\mathcal{I}_c| - 1} \sum_{\substack{j \in \mathcal{I}_c \\ j \neq i}} d_{\cos}\left(\mathbf{h}_i^{2018}, \mathbf{h}_j^{2022}\right)}, \tag{3}$$

where $\mathcal{I}_c$ is the set of repeated, label-stable locations belonging to class c. The ratios R_c are again macro-averaged across classes. KD reduced this matched-to-nonmatched ratio by 2.6% on average, with lower values in four of five seeds. Thus, most of the reduction in cross-year dissimilarity co-occurs with increased intra-class compactness, while a smaller additional same-location stability effect remains. We therefore regard the result as consistent with the semantic-anchoring interpretation of ground-view distillation with strong regularization by the photo teacher's predictive structure.

Table 2: Macro F1 and Weighted F1 for the frozen AlphaEarth CE baseline, contrastive objectives (InfoNCE, SupCon), photo probe, and fusion, mean ± std over the same five seeds, split, and evaluation protocol as Table 1. Photographs are used only during training for the satellite-only KD and contrastive models; the photo probe and fusion model require photographs at inference. The first two rows repeat the matched CE and KD students of Table 1 as anchors. The CE control row is the in-harness baseline of the contrastive comparison (identical backbone, data, and optimizer as the CL rows); its head differs architecturally from the student, which accounts for the small offset between the two CE rows.

Model	MacroF1%	WeightedF1%
Frozen AlphaEarth CE student [z]	63.5 ± 0.6	67.0 ± 0.4
+ KD [z]	66.5 ± 0.3	68.6 ± 0.4
CE control [z]	63.8 ± 0.7	67.4 ± 0.5
+ CLIP (InfoNCE)	63.7 ± 0.7	67.2 ± 0.6
+ SupCon	63.4 ± 0.4	67.1 ± 0.4
Photo probe [CLIP, v]	69.7 ± 0.4	69.2 ± 0.3
Fusion [z+v]	77.7 ± 0.3	77.0 ± 0.3

C Contrastive Learning & Modality Fusion

We further tested alternative cross-modal objectives, such as contrastive learning (CL), including a supervised contrastive (SupCon) [Khosla et al., 2020] and CLIP-style instance contrastive (InfoNCE) [van den Oord et al., 2018, Radford et al., 2021]. SupCon adds a supervised contrastive loss across modalities with positives defined by shared crop class. In contrast, InfoNCE adds an instance-contrastive loss with positives defined by the same site, aligning each satellite point with its own photo against the rest of the batch. We tested the CL approaches on the frozen AlphaEarth embeddings, coupled with the CLIP photo embeddings and compared against the AlphaEarth and photo probe. We further include the performance for combining both modalities. For the fusion model both modalities are concatenated at the input level and fed into the classifier. Results for 2018 are shown in Table 2. Importantly, the CL objectives do not improve over the matched CE control (per-seed paired differences are small and mixed in sign), in contrast to the KD objective repeated in the table ($+2.9$ macro-F1 paired over the CE student). The same dissociation holds within the distillation harness: adding only the logit-KL term to CE reproduces the full gain ($+2.9$ macro-F1, positive on every seed), whereas adding only the embedding-cosine term is null ($+0.2$ macro-F1, mixed signs). The CL objectives shape embeddings rather than predictive distributions, so their null result is consistent with the cosine-only control. In contrast, the macro-F1 of the photo probe is consistently about 6 pp higher than the CL objectives and both CE rows, which use the AlphaEarth embeddings only. The combination of both modalities provides consistently the best results with approximately 8 pp higher macro-F1 than the photo probe model. This is expected, since this model receives complementary information that boosts crop classification performance but it cannot be used for inference in locations where no field photos are available.

D Per-class analysis

We further show the per-class results of the marcro-F1 score on the 2018 test set in Table 3 and for the temporal transfer in Table 4. Notably, KD improves crop classification for most classes, except individual cases such as rice (89.0 to 88.9) for the frozen AlphaEarth embeddings. Moreover, paired KD$-$CE macro-F1 margins per crop class for the temporal transfer is shown in Figure 4, evaluated at 2022. Notably, performance gains depends on crop type and model branch, with $+38.9$ macro-F1 points for Flax/Hemp on the frozen TESSERA model and -1.6 on Rice for the TESSERA fine-tuned branch. To demonstrate that the reported KD gain is not only driven by single class results, we show the macro-F1 score for a held-one-class-out validation in Figure 5 with stable gains, except for frozen TESSERA. For frozen TESSERA under temporal transfer, excluding the combined flax, cotton, and hemp class reduces the gain from 5.5 to 3.4 percentage points.

Table 3: Per-class F1 (%) on the 2018 test set, mean ± standard deviation over five seeds, for matched CE and KD students. Classes are sorted by 2018 test support.

	Frozen AlphaEarth		Frozen TESSERA		S2 SITS (scratch)		S1+S2 SITS (scratch)		TESSERA fine-tuned	
Class	CE	+KD	CE	+KD	CE	+KD	CE	+KD	CE	+KD
Wheat	70.0 ± 0.9	71.3 ± 0.9	74.0 ± 0.2	76.0 ± 0.7	71.4 ± 1.3	73.7 ± 0.8	73.4 ± 2.2	76.3 ± 1.0	77.2 ± 1.1	78.3 ± 0.6
Maize	87.8 ± 0.4	89.0 ± 0.2	89.0 ± 0.3	89.5 ± 0.2	88.6 ± 0.8	90.0 ± 0.3	89.9 ± 0.4	91.1 ± 0.3	91.4 ± 0.4	91.3 ± 0.4
Barley	55.6 ± 1.1	55.6 ± 0.9	64.3 ± 1.0	63.6 ± 0.6	57.4 ± 1.3	57.9 ± 1.2	59.6 ± 0.6	60.8 ± 0.8	65.5 ± 0.8	64.8 ± 0.9
Other cereals	46.6 ± 0.5	46.6 ± 0.4	51.4 ± 0.9	49.8 ± 1.2	44.1 ± 0.8	43.3 ± 1.1	48.6 ± 0.6	49.3 ± 0.9	53.5 ± 1.3	53.6 ± 1.0
Rapeseed	77.4 ± 1.6	79.4 ± 0.6	80.6 ± 1.0	83.3 ± 0.3	79.0 ± 0.1	79.8 ± 0.5	82.1 ± 0.9	84.0 ± 0.5	82.1 ± 0.5	83.4 ± 0.4
Sunflower	74.5 ± 1.3	77.4 ± 1.0	83.2 ± 1.5	86.2 ± 0.5	79.0 ± 1.3	80.0 ± 2.1	85.2 ± 1.0	86.1 ± 0.7	86.5 ± 1.1	87.2 ± 1.3
Fruits	60.8 ± 1.0	64.9 ± 1.3	61.4 ± 1.2	64.0 ± 1.3	53.1 ± 1.7	58.4 ± 2.7	55.7 ± 3.7	60.7 ± 1.3	60.6 ± 0.9	63.6 ± 1.8
Dry pulses	38.3 ± 1.5	42.6 ± 1.3	59.1 ± 3.3	64.0 ± 0.8	44.2 ± 1.0	48.7 ± 2.5	54.3 ± 2.6	58.3 ± 1.5	62.2 ± 1.0	62.6 ± 2.8
Sugar Beet	81.8 ± 1.1	84.5 ± 1.6	84.2 ± 0.9	86.3 ± 0.7	83.6 ± 1.2	86.4 ± 1.6	86.9 ± 2.0	87.6 ± 1.4	87.7 ± 0.7	88.0 ± 0.8
Olives	71.3 ± 1.9	74.3 ± 0.4	74.4 ± 1.7	76.0 ± 1.3	67.5 ± 1.4	69.7 ± 1.4	69.0 ± 1.2	69.1 ± 2.1	73.4 ± 1.9	74.0 ± 1.0
Potatoes	60.3 ± 1.7	67.2 ± 1.6	68.0 ± 1.9	71.8 ± 1.3	62.2 ± 1.3	66.6 ± 1.8	69.8 ± 2.5	72.6 ± 1.1	75.6 ± 2.3	78.0 ± 1.3
Grapes	60.3 ± 1.0	61.1 ± 1.1	65.7 ± 1.4	67.1 ± 2.0	52.0 ± 3.4	55.7 ± 0.9	56.5 ± 0.8	57.1 ± 0.8	62.0 ± 2.2	62.7 ± 2.6
Fresh vegetables	30.6 ± 2.0	34.4 ± 1.6	39.3 ± 1.9	44.5 ± 0.5	30.6 ± 3.6	30.9 ± 3.1	33.7 ± 2.7	34.7 ± 2.9	39.1 ± 5.2	39.8 ± 4.0
Nuts	41.3 ± 3.8	42.9 ± 2.1	39.3 ± 2.2	44.2 ± 4.0	33.9 ± 3.3	40.3 ± 3.0	38.3 ± 1.0	41.7 ± 1.5	41.5 ± 3.9	41.5 ± 2.7
Soybeans	56.5 ± 3.7	69.1 ± 3.3	76.4 ± 1.7	82.5 ± 0.6	59.0 ± 3.8	70.2 ± 4.2	64.3 ± 2.7	71.9 ± 2.4	76.7 ± 1.2	78.4 ± 2.1
Flax, cotton and hemp	77.8 ± 1.2	80.9 ± 2.0	35.8 ± 0.5	80.7 ± 2.6	70.4 ± 4.6	73.4 ± 1.2	73.2 ± 4.9	77.2 ± 2.6	85.8 ± 1.5	84.5 ± 2.4
Rice	89.0 ± 3.8	88.9 ± 2.2	87.4 ± 2.9	90.1 ± 2.5	77.9 ± 5.6	82.7 ± 4.3	83.2 ± 4.6	90.1 ± 4.2	90.4 ± 1.8	93.1 ± 1.9
Macro F1	63.5 ± 0.6	66.5 ± 0.3	66.7 ± 0.2	71.7 ± 0.2	62.0 ± 1.1	65.2 ± 1.0	66.1 ± 1.0	68.7 ± 0.6	71.2 ± 0.7	72.1 ± 0.5

Table 4: Per-class F1 (%) under 2018→2022 temporal transfer, mean ± standard deviation over five seeds: the same trained students evaluated on the 2022 survey (unseen locations, train-year scaler and class map). Classes are sorted by 2018 test support.

	Frozen AlphaEarth		Frozen TESSERA		S2 SITS (scratch)		S1+S2 SITS (scratch)		TESSERA fine-tuned	
Class	CE	+KD	CE	+KD	CE	+KD	CE	+KD	CE	+KD
Wheat	64.2 ± 0.4	65.8 ± 0.9	65.9 ± 1.2	70.8 ± 1.1	64.0 ± 1.4	66.4 ± 0.6	66.6 ± 2.5	69.7 ± 0.9	71.2 ± 1.5	72.3 ± 0.5
Maize	78.5 ± 0.9	83.3 ± 0.2	86.3 ± 0.4	88.3 ± 0.2	83.8 ± 0.7	85.7 ± 0.9	85.8 ± 0.6	87.2 ± 0.2	88.1 ± 0.3	88.6 ± 0.3
Barley	48.0 ± 1.9	48.3 ± 0.2	55.5 ± 0.5	57.3 ± 1.1	49.5 ± 1.0	50.1 ± 0.3	52.3 ± 0.8	53.1 ± 0.8	57.7 ± 1.4	58.5 ± 0.4
Other cereals	42.2 ± 1.5	42.7 ± 1.9	47.6 ± 1.6	48.3 ± 1.5	41.9 ± 1.7	41.9 ± 1.8	46.0 ± 1.1	46.2 ± 0.5	50.3 ± 1.4	49.0 ± 1.9
Rapeseed	68.4 ± 1.1	69.6 ± 0.4	71.8 ± 0.9	73.5 ± 0.5	67.8 ± 1.2	70.1 ± 0.8	73.3 ± 0.2	74.2 ± 0.1	72.7 ± 0.5	74.3 ± 0.5
Sunflower	45.7 ± 2.2	51.0 ± 3.5	73.2 ± 1.4	74.2 ± 1.1	65.7 ± 1.3	68.4 ± 1.6	73.5 ± 0.9	75.1 ± 1.5	75.8 ± 1.2	76.2 ± 2.5
Fruits	59.9 ± 1.1	62.4 ± 1.1	53.9 ± 1.6	57.4 ± 1.0	54.3 ± 0.6	55.8 ± 1.6	54.2 ± 1.9	56.8 ± 1.3	58.4 ± 0.8	59.8 ± 0.8
Dry pulses	20.8 ± 2.3	24.9 ± 2.2	38.3 ± 3.3	45.0 ± 1.1	29.2 ± 1.5	33.7 ± 1.1	37.3 ± 2.7	40.1 ± 2.0	41.8 ± 0.8	42.3 ± 2.4
Sugar Beet	64.6 ± 5.2	72.1 ± 1.5	64.9 ± 5.5	73.6 ± 4.5	76.2 ± 2.1	80.1 ± 2.3	79.1 ± 3.2	81.6 ± 1.4	78.7 ± 1.3	80.8 ± 2.2
Olives	70.0 ± 1.7	70.3 ± 1.4	67.4 ± 1.5	69.8 ± 1.8	63.6 ± 1.8	66.4 ± 1.0	64.0 ± 1.6	65.8 ± 0.7	68.4 ± 0.7	70.2 ± 1.1
Potatoes	41.0 ± 2.0	48.1 ± 2.5	56.9 ± 2.8	63.5 ± 2.1	51.9 ± 1.5	55.0 ± 1.9	57.5 ± 1.7	57.8 ± 1.5	63.7 ± 1.6	64.4 ± 1.1
Grapes	54.7 ± 3.0	54.8 ± 5.0	49.6 ± 4.7	52.3 ± 3.0	46.8 ± 3.1	46.6 ± 3.9	53.3 ± 1.9	51.8 ± 2.4	55.7 ± 3.5	55.2 ± 2.6
Fresh vegetables	22.7 ± 1.6	28.5 ± 1.5	30.2 ± 1.6	35.3 ± 1.7	28.3 ± 1.0	29.4 ± 3.2	29.1 ± 2.0	30.7 ± 1.1	34.3 ± 1.7	36.1 ± 1.2
Nuts	32.1 ± 2.8	31.6 ± 1.6	29.5 ± 3.4	32.9 ± 1.7	25.2 ± 1.6	28.0 ± 0.9	28.9 ± 1.4	29.2 ± 1.1	30.1 ± 0.9	30.2 ± 0.8
Soybeans	34.3 ± 2.6	42.0 ± 1.8	62.3 ± 3.4	64.9 ± 1.8	41.2 ± 1.5	45.6 ± 3.1	45.2 ± 1.7	45.5 ± 4.2	60.2 ± 1.9	58.8 ± 1.7
Flax, cotton and hemp	65.6 ± 3.8	70.7 ± 3.1	39.8 ± 0.6	78.7 ± 2.3	58.6 ± 3.6	61.9 ± 2.9	56.7 ± 1.8	58.4 ± 2.8	74.9 ± 2.6	78.4 ± 1.3
Rice	81.7 ± 3.1	86.5 ± 2.1	85.9 ± 1.2	86.8 ± 1.6	70.3 ± 4.9	73.7 ± 3.1	75.3 ± 3.9	75.0 ± 3.0	88.1 ± 1.5	86.5 ± 1.6
Macro F1	52.6 ± 1.2	56.0 ± 0.6	57.6 ± 0.7	63.1 ± 0.6	54.0 ± 1.0	56.4 ± 0.7	57.5 ± 0.7	58.7 ± 0.1	63.0 ± 0.6	63.6 ± 0.4

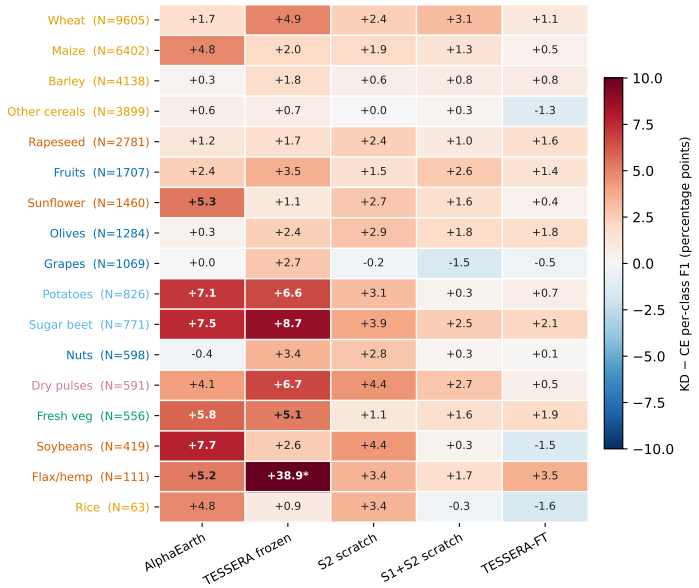


Figure 4: Heatmap for paired KD−CE macro-F1 margins per crop class for the temporal transfer (2022). Colorbar is cropped to 10.0 macro-F1 points.

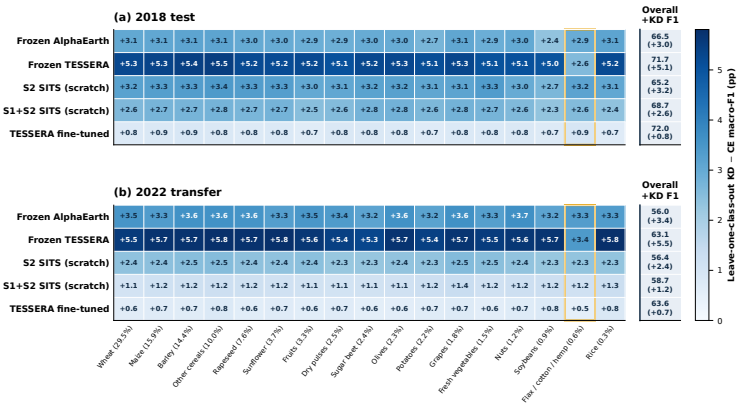


Figure 5: Heatmap for paired macro-F1 KD−CE margins fore a held-one-class-out validation over the model branches and yearsRre. Constantly stable performance gains, except for the frozen TESSERA.