Land Cover Mapping update via Unsupervised Domain Adaptation from satellite image time series data

Giuseppe Guarino (giuseppe.guarino@inrae.fr)
Post-Doc Fellowship
INRAE, INRIA, EVERGREEN, UMR TETIS

Dino Ienco (dino.ienco@inrae.fr)
Research Director, PhD, HDR
INRAE, INRIA, EVERGREEN, UMR TETIS
Evergreen team - https://team.inria.fr/evergreen/

Outline

1. Introduction to Remote Sensing
2. Domain Adaptation: Background
3. Paper Review
   - Spatial Shift: TimeMatch
   - Temporal Shift : SpADANN
   - Multi-Sensor Temporal Shift: M3SPADA
   - Multi-Source Temporal Shift: ConDA-Vine
   - Sensor Shift: SHeDD & SAHARA
4. Synthesis & Discussion

What is Remote Sensing?

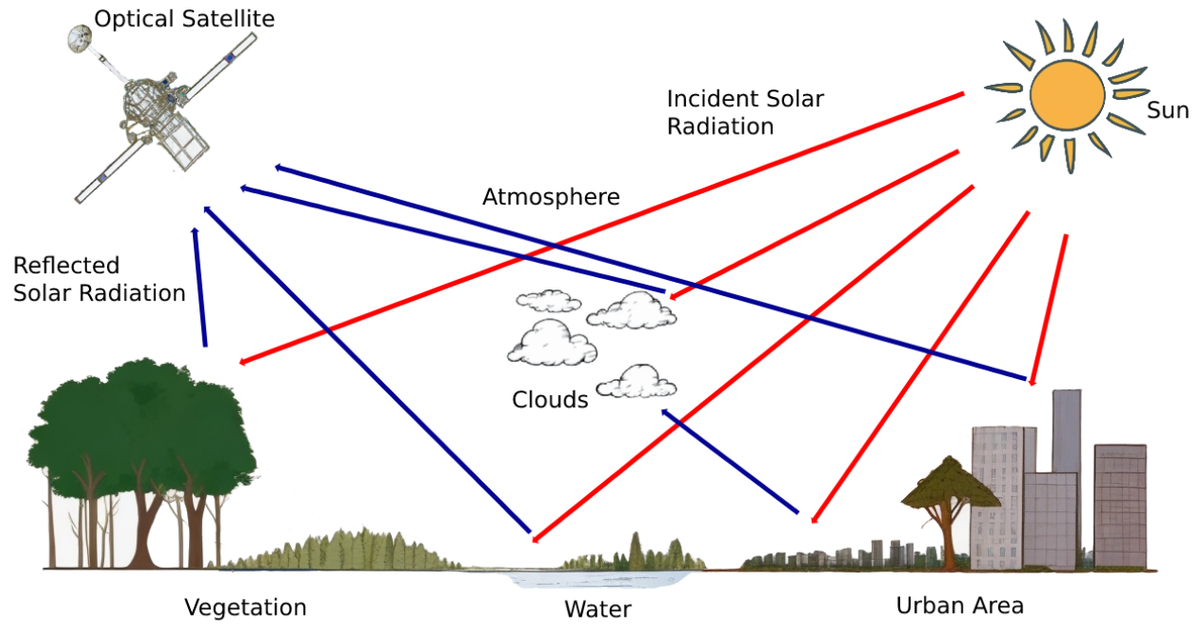
- **Remote Sensing (RS)**: the science and technology of acquiring information about an object, area, or phenomenon without direct physical contact with it

- **Platforms:** satellites, aircraft, drones (UAVs), ground-based stations

- **Core idea:** sensors measure electromagnetic radiation (EMR) reflected or emitted by the Earth's surface at various wavelengths

- **Earth Observation (EO)**: the systematic, continuous monitoring of the Earth's physical, chemical, and biological systems using RS technology

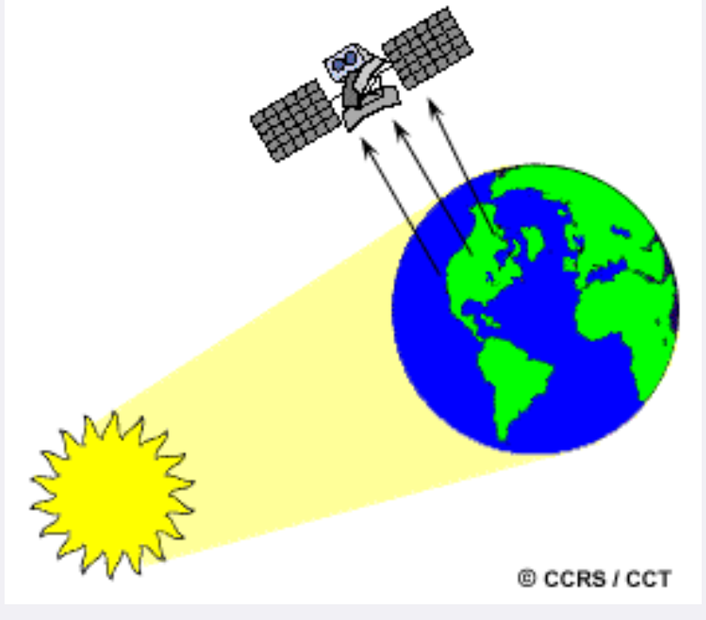
- **Applications:** agriculture, forestry, urban planning, disaster monitoring, climate science, military reconnaissance



Passive vs Active sensors

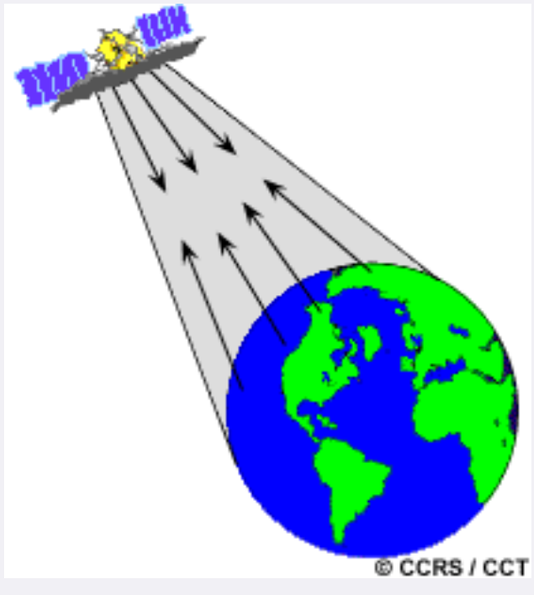
Passive Sensors

- Measure **naturally available radiation**: reflected sunlight (visible, NIR) or thermal emission
- The sensor does **not** emit its own energy
- Examples: optical cameras, multispectral/ hyperspectral sensors, thermal radiometers
- **Limitation**: affected by cloud cover, atmospheric conditions, limited to daytime (for reflected light)
- Most common in EO: Landsat, Sentinel-2, SPOT, MODIS



Active Sensors

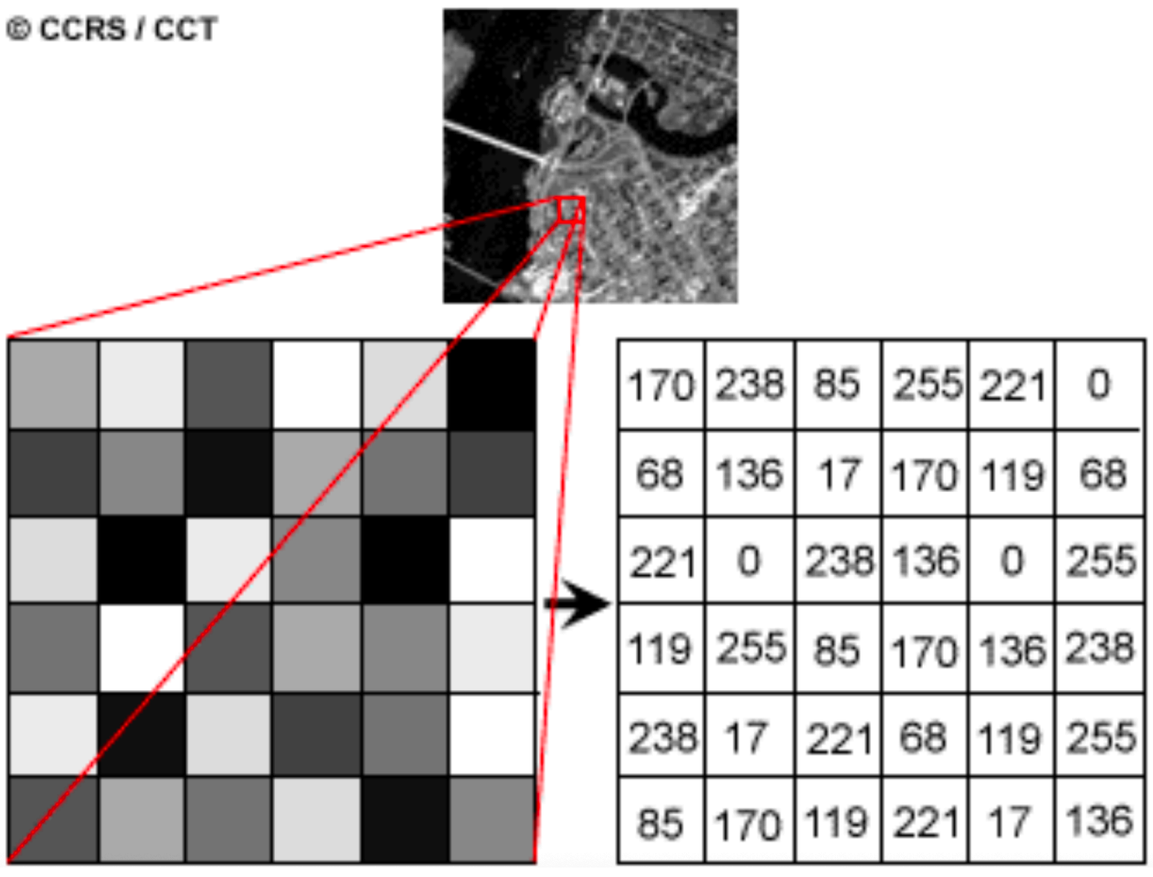
- **Emit their own energy** pulse and measure the return signal (backscatter)
- Independent of solar illumination
- Examples: SAR (Synthetic Aperture Radar), LiDAR, altimeters
- **Advantage**: day/night, all-weather capability; can penetrate clouds, rain, and even vegetation canopy
- Sentinel-1 (C-band SAR) is the primary EO SAR mission



Source: https://natural-resources.canada.ca/sites/nrcan/files/earthsciences/pdf/resource/tutor/fundam/pdf/fundamentals_e.pdf

The pixel

- The captured signal is organized on a grid of bins (pixels) which capture the brighness level of the covered zone.

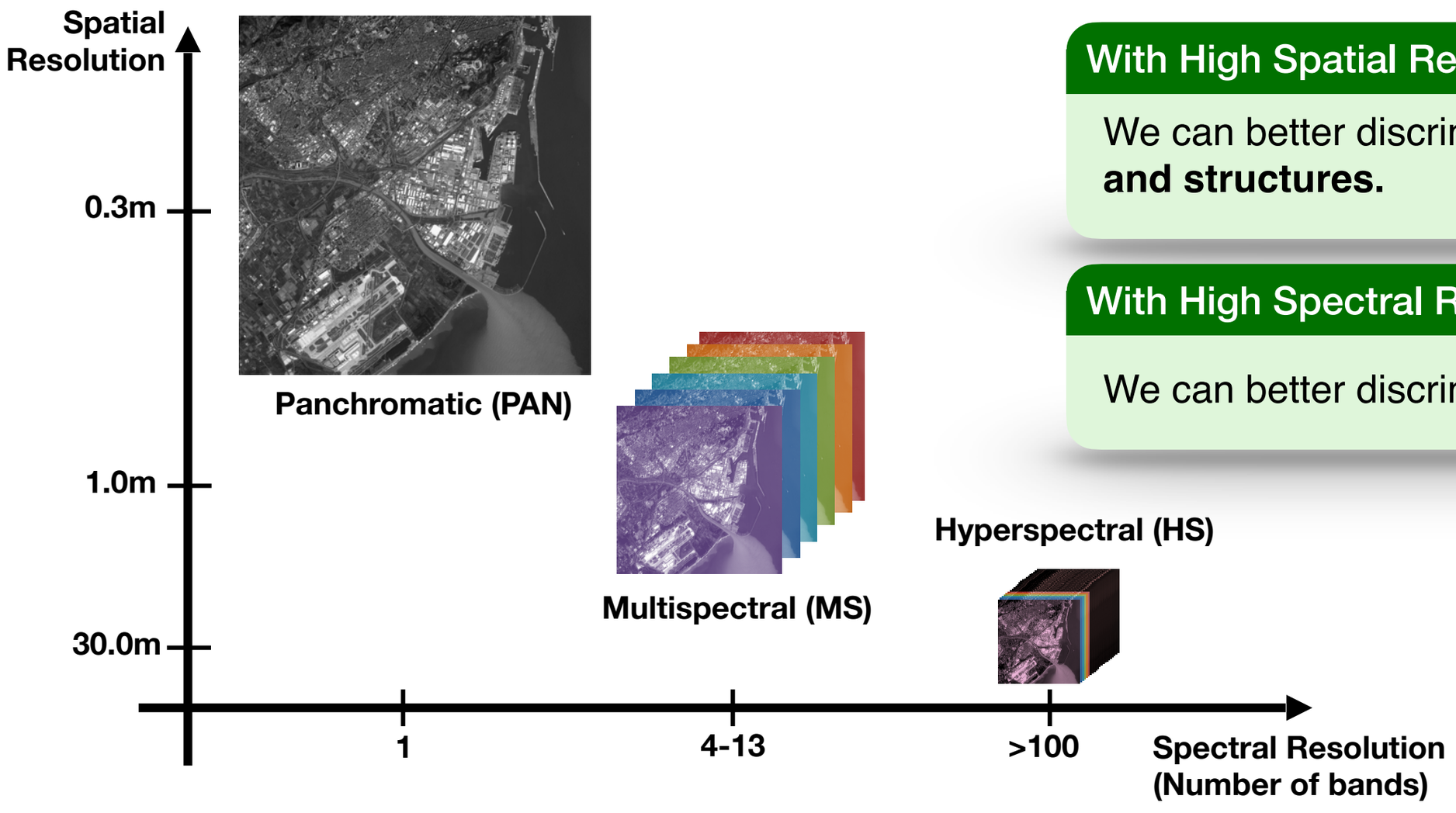




Source: https://natural-resources.canada.ca/sites/nrcan/files/earthsciences/pdf/resource/tutor/fundam/pdf/fundamentals_e.pdf

Types of resolution

1. **Spatial resolution:** size of the smallest object that can be detected (pixel size on the ground)
   - Low: >100m (MODIS); Medium: 10–30m (Sentinel-2, Landsat); High: 1–5m (SPOT); Very High: <1m (Pléiades, WorldView)
2. **Spectral Resolution:** number and width of spectral bands
   - Panchromatic (1 band), Multispectral (4–13 bands), Hyperspectral (>100 bands)



With High Spatial Resolution

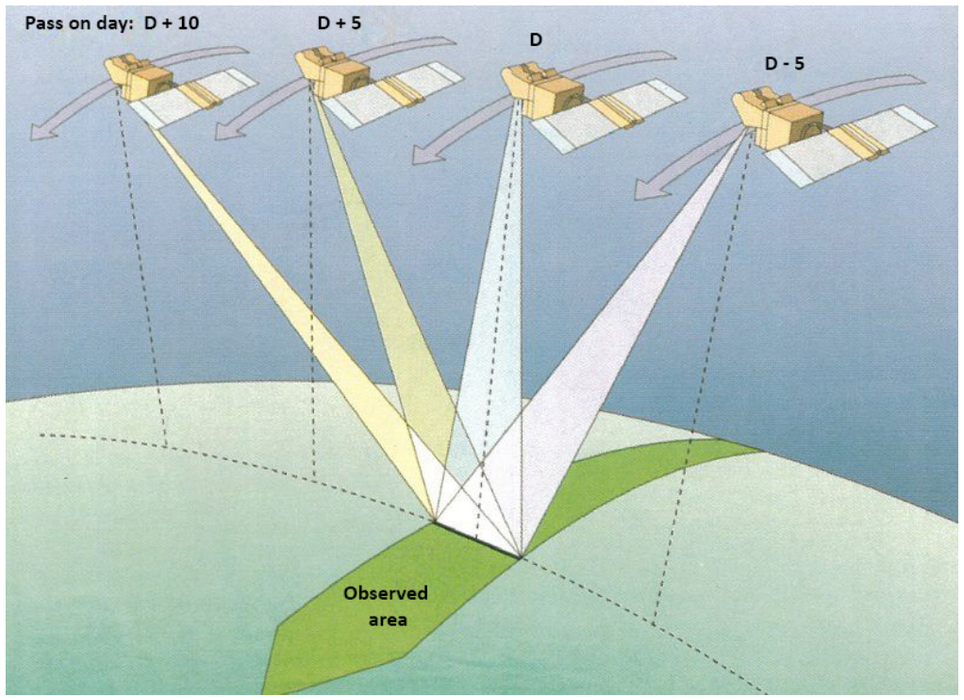
We can better discriminate **objects and structures.**

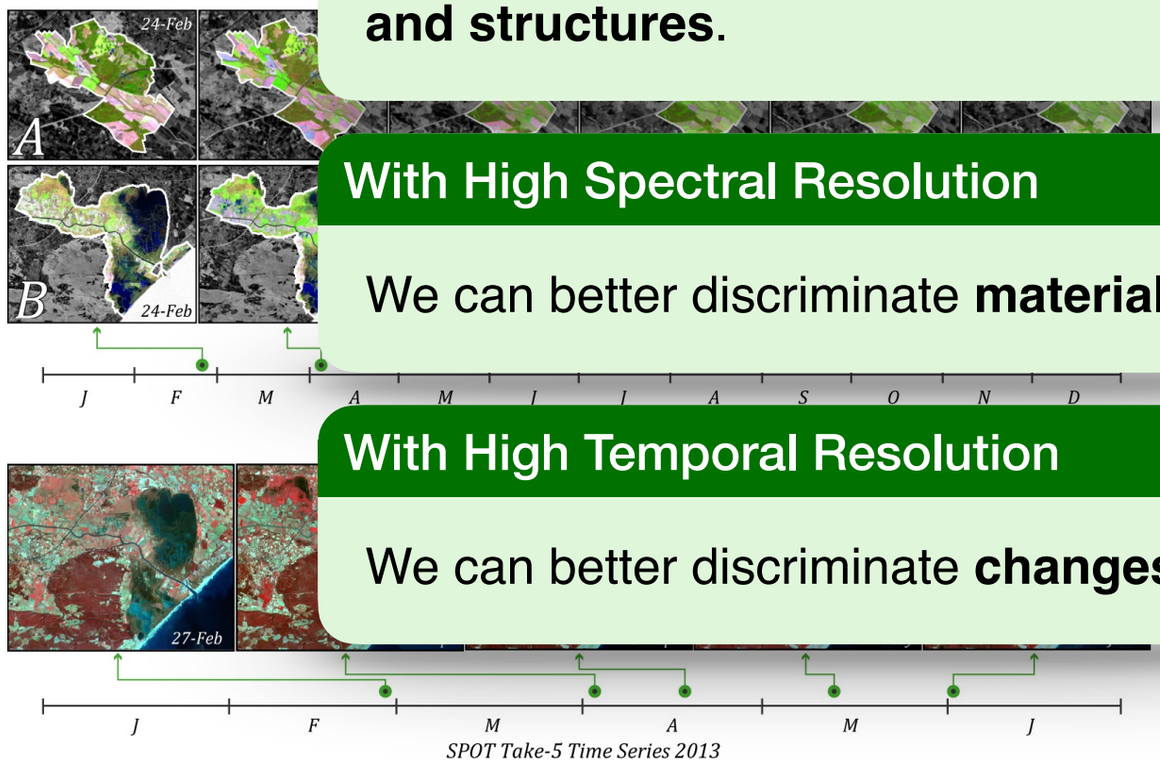
With High Spectral Resolution

We can better discriminate **materials.**

Types of resolution

1. **Spatial resolution:** size of the smallest object that can be detected (pixel size on the ground)
   - Low: >100m (MODIS); Medium: 10–30m (Sentinel-2, Landsat); High: 1–5m (SPOT); Very High: <1m (Pléiades, WorldView)
2. **Spectral Resolution:** number and width of spectral bands
   - Panchromatic (1 band), Multispectral (4–13 bands), Hyperspectral (>100 bands)
3. **Temporal Resolution:** revisit frequency
   - Sentinel-2: 5 days; Landsat: 16 days; MODIS: daily





With High Spatial Resolution

We can better discriminate **objects and structures**.

With High Spectral Resolution

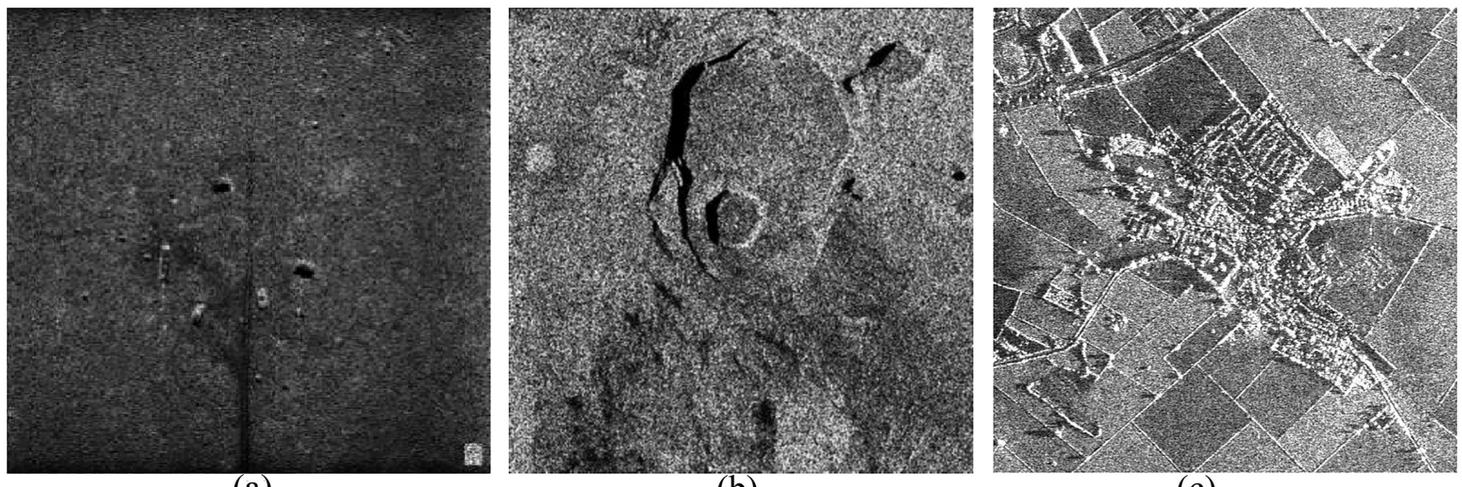
We can better discriminate **materials**.

With High Temporal Resolution

We can better discriminate **changes**.

https://eo.belspo.be/en/iii23-temporal-resolution

SAR (Synthetic Aperture Radar) in Detail



(a) (b) (c)

- SAR transmits microwave pulses and records the **backscatter** signal
- Uses the satellite's motion to synthesize a large antenna aperture, achieving high spatial resolution
- **Polarization:** how the radar wave oscillates
  - VV (vertical send, vertical receive): surface roughness
  - VH (vertical send, horizontal receive): volume scattering (vegetation)
  - Dual-pol (VV+VH) is standard on Sentinel-1
- Sensitive to: surface roughness, dielectric constant (moisture), geometric structure
- **Not sensitive to:** color, chlorophyll content (unlike optical)
- This fundamental difference in what SAR vs. optical "see" is what makes heterogeneous DA so challenging

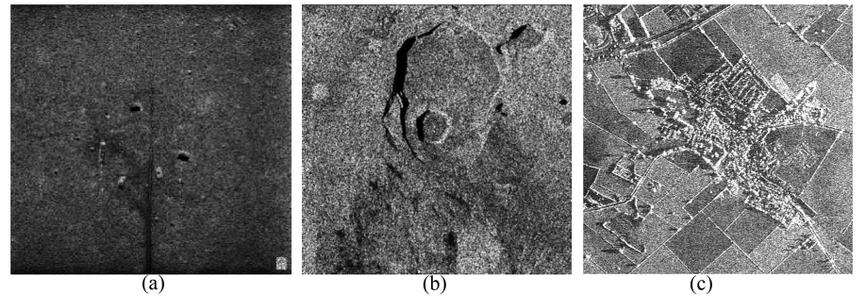
[singh2021] Singh, P., Shree, R., & Diwakar, M. (2021). A new SAR image despeckling using correlation based fusion and method noise thresholding. *Journal of King Saud University-Computer and Information Sciences*, *33*(3), 313-328.

Key Data Types in EO

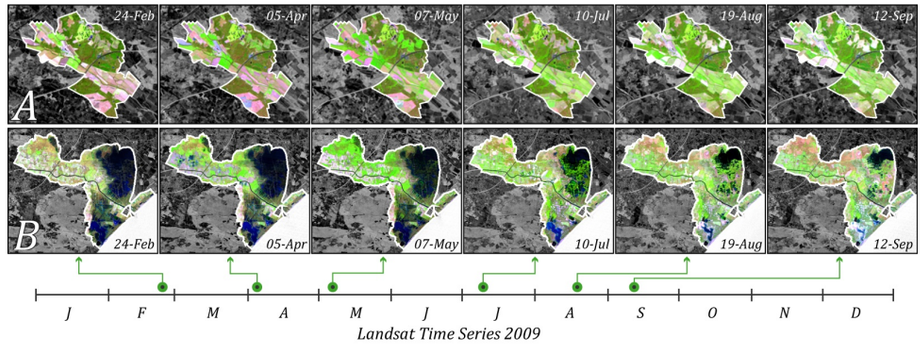
- **Optical / Multispectral imagery**
  - Multiple spectral bands (e.g., Sentinel-2: 13 bands covering visible, NIR, SWIR)
  - Vegetation indices: NDVI $= \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$, EVI, NDWI
  - Widely used for land cover mapping, crop monitoring



- **SAR (Synthetic Aperture Radar)**
  - Backscatter intensity, polarimetry (VV, VH)
  - Sensitive to structure, moisture, surface roughness
  - Complementary to optical (different physical properties)



- **Satellite Image Time Series (SITS)**
  - Sequence of images acquired over the **same location** at different times
  - Captures **phenological dynamics**: vegetation growth, senescence, harvest cycles
  - Essential for crop type classification and change detection



[singh2021] Singh, P., Shree, R., & Diwakar, M. (2021). A new SAR image despeckling using correlation based fusion and method noise thresholding. *Journal of King Saud University-Computer and Information Sciences*, *33*(3), 313-328.

Satellite Image Time Series (SITS)

Why are time series important?

- A single image captures a **snapshot**; a time series captures **dynamics**
- Crop types have distinctive **temporal signatures** (phenological profiles):
  - Winter wheat: green in winter, harvested in summer
  - Corn: planted in spring, peak greenness in summer
  - Rice: distinct flooding + growth + harvest pattern
- NDVI time series show characteristic **growth curves** per crop
- Same crop in different regions may have **shifted** temporal profiles due to climate differences → this is a major source of **domain shift**

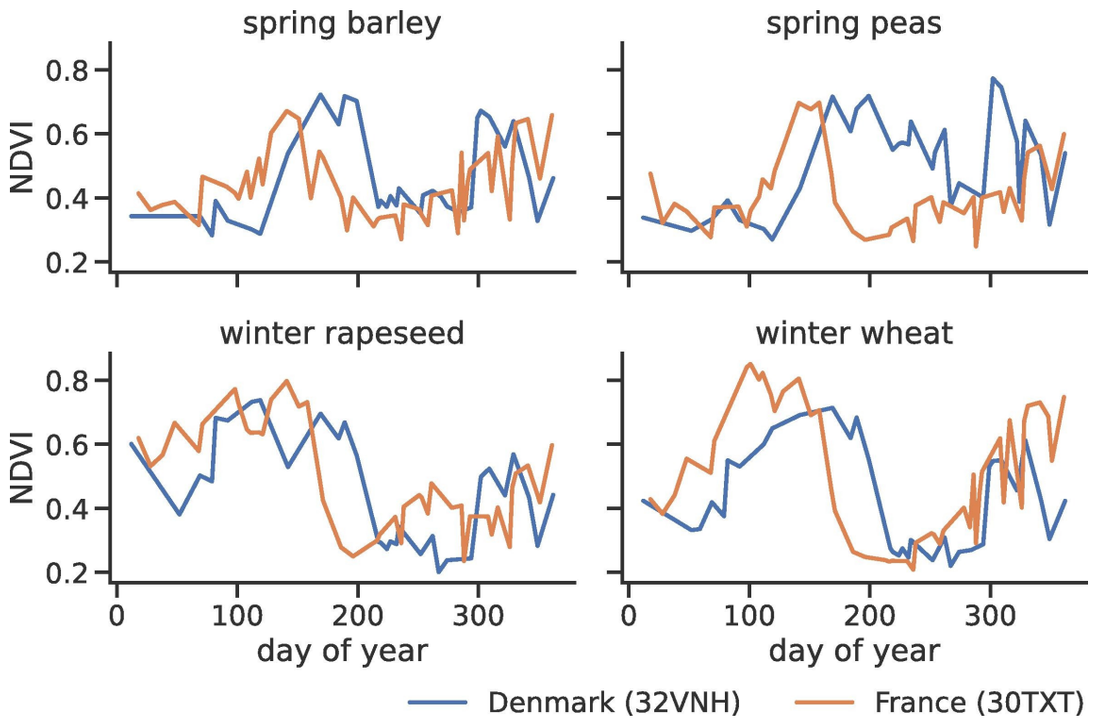


Figure: NDVI profiles for 4 crop types: Denmark (blue) vs. France (orange). The same crop shows temporal shift: peaks occur at different times

Deep Learning for EO: A Brief Overview

Convolutional Neural Networks (CNNs)

- 2D CNNs for spatial feature extraction from image patches
- 1D CNNs (**TempCNN**) for temporal feature extraction from SITS pixel time series
- ResNet-18: common backbone for RS image classification

Recurrent Networks (RNNs/LSTMs)

- Process sequential SITS data
- Capture long-range temporal dependencies
- Being replaced by attention-based models

Attention & Transformers

- **LTAE** (Lightweight Temporal Attention Encoder): efficient attention for SITS
- **Vision Transformers** (ViT): patch-based image classification
- **TinyViT**: compact transformer (5M params)

Common training setup:

- Supervised classification: cross-entropy loss with labeled training data
- Data augmentation: random flips, rotations, spectral jittering
- Problem: requires **large labeled datasets** → annotation bottleneck

The Annotation Bottleneck

The Core Problem

Supervised deep learning models require large amounts of labeled data, but annotations in remote sensing are extremely costly and limited.

Expensive to acquire

- Field campaigns (GPS waypoints, on-site verification)
- Expert photointerpretation (time-consuming, subjective)
- Administrative data (LPIS, cadastre). Not always available or accurate
- Cost: thousands of euros per region per year

Spatially and temporally limited

- Labels available for specific regions/years only

Models fail to generalize because of:

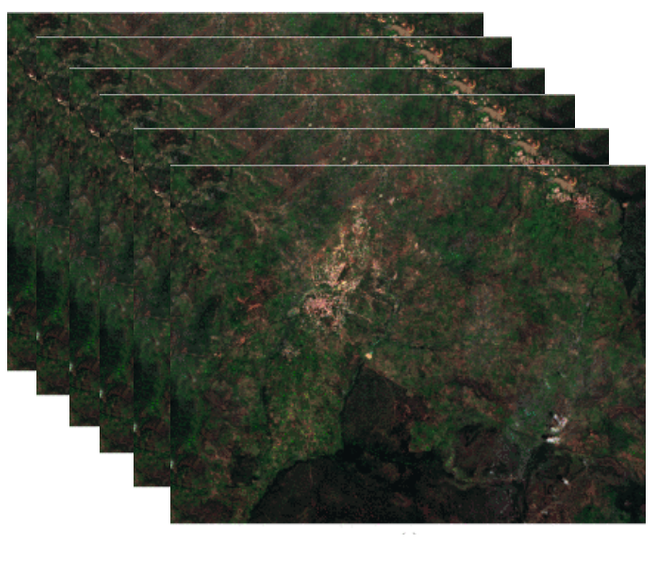
- Different climates → different phenology
- Different sensors → different feature spaces
- Different years → different weather patterns
- Different land use practices across countries
- Different atmospheric conditions

Goal

Train on labeled source domain, deploy on unlabeled target domain ⇒ Domain Adaptation

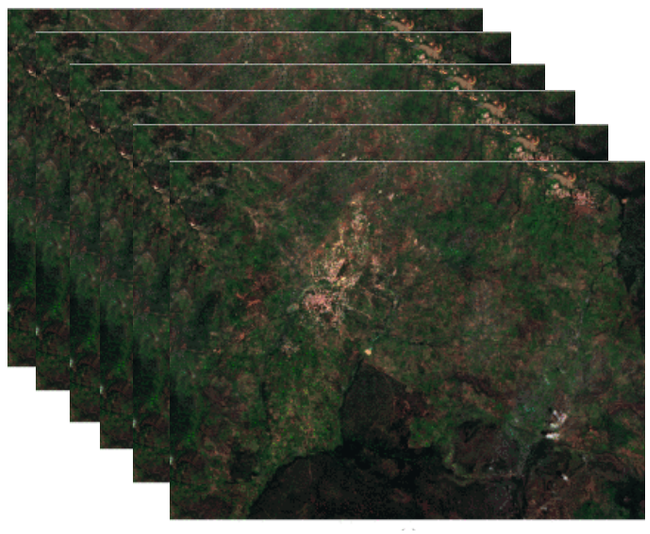
Deep Learning for RS data

Training set



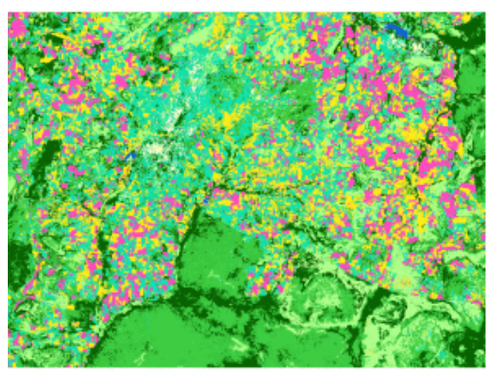
2010

Test set

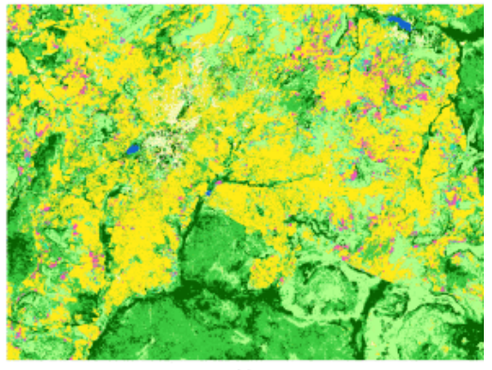


2020

Deep Learning Model



Deep Learning Model



Domain shift

What is Domain Shift?

Definition

A **domain** is defined by a joint distribution $P(\mathbf{X}, Y)$ over input features $\mathbf{X}$ and labels Y.
Domain shift: $P_S(\mathbf{X}, Y) \neq P_T(\mathbf{X}, Y)$ between source (S) and target (T).

Types of domain shift:

- **Covariate shift:** $P_S(\mathbf{X}) \neq P_T(\mathbf{X})$, but $P(Y \mid \mathbf{X})$ stays the same
  - Same crops, different spectral distributions (e.g., different regions, atmospheric conditions)
  - Most common type in RS; addressed by most DA methods
- **Label shift (prior probability shift):** $P_S(Y) \neq P_T(Y)$, different class priors
  - Different crop distributions between countries (e.g., more rice in Asia, more wheat in Europe)
- **Concept shift:** $P_S(Y \mid \mathbf{X}) \neq P_T(Y \mid \mathbf{X})$
  - Same spectral signature → different class interpretation (e.g., “grassland” vs. “pasture”)

Sources of Domain Shift in Remote Sensing

Shift type	Cause	Example
Spatial shift	Different geographic regions with different climate, soil, farming practices	France → Denmark: same crops but different phenological timing
Temporal shift	Different years with different weather patterns, rainfall, temperature	2018 → 2021: drought year vs. normal year changes vegetation profiles
Sensor shift	Different instruments with different spectral/spatial characteristics	Sentinel-2 (10 m, 13 bands) → SPOT-6 (1.5 m, 4 bands)
Modality shift	Fundamentally different physical measurement principles	Optical (reflectance) ↔ SAR (backscatter): different feature spaces

In practice, **multiple shifts often co-occur**: e.g., different region + different year + different sensor. This makes DA in RS particularly challenging compared to standard computer vision benchmarks.

Domain Adaptation in Remote Sensing

Domain Adaptation & Remote Sensing

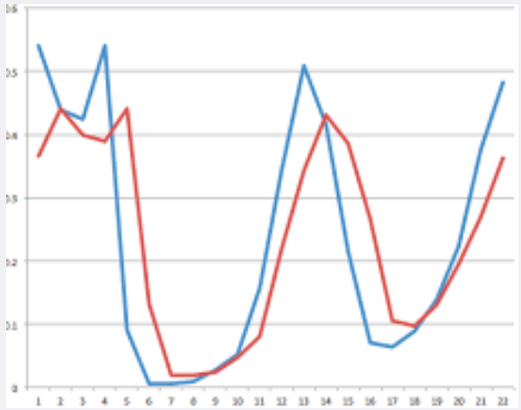
Distribution shifts can be caused by:

Climate changes



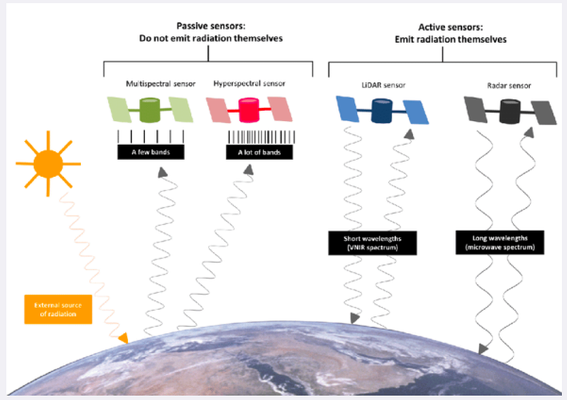
Changes in **acquisition conditions** (i.e. climate/environment).

Temporal changes



Temporal shifts (i.e. same study site from one year to another one or cross-site).

Sensor changes



Sensors differences (i.e. optical sensors with slightly different radiometry).

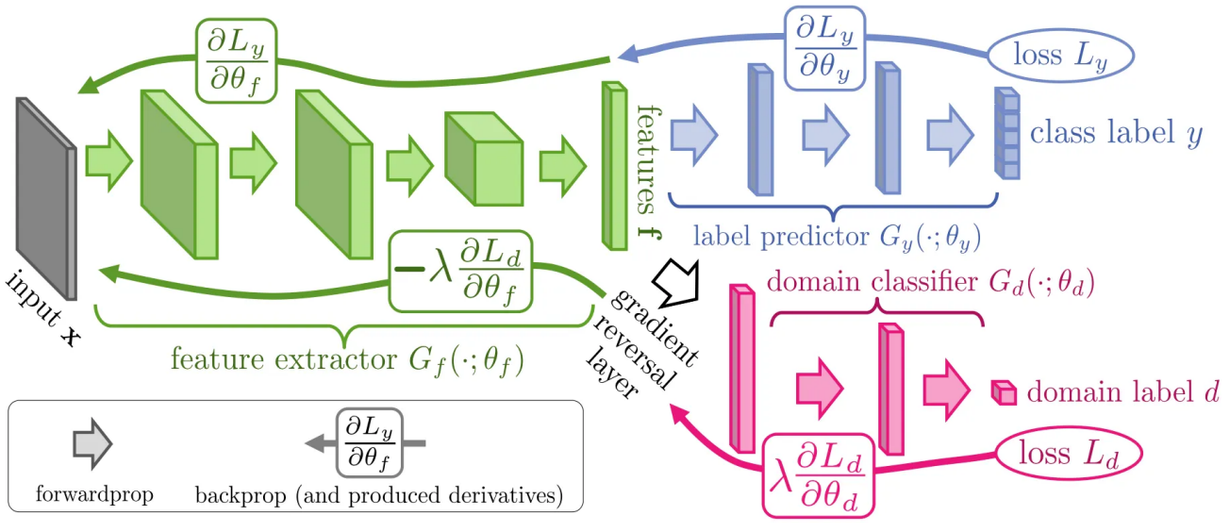
All these points clearly induce data distribution shifts (between training/source and test/target domains) that will **hinder the use of standard supervised DL methods.**

This is why **Domain Adaptation** can be an interesting **research direction** in **Remote Sensing Analysis** to adapt models without costly and time-consuming efforts.

Key Technique: Adversarial Domain Adaptation (DANN)

- **Idea** (Ganin et al., 2016): learn features that are discriminative for the task but invariant across domains
- **Gradient Reversal Layer (GRL)**: reverses gradient sign during backpropagation. The feature extractor learns to confuse the domain classifier
- Three components:
  1. Feature extractor G_f
  2. Task classifier G_y (minimize L_{cl})
  3. Domain classifier G_d (minimize L_{adv})
- Minimax game:

$$\min_{G_f, G_y} \max_{G_d} L_{cl} - \lambda L_{adv}$$



At convergence, features are domain-invariant yet task-discriminative.

Ganin, Yaroslav, et al. "Domain-adversarial training of neural networks." *Journal of machine learning research* 17.59 (2016): 1-35.

DANN: How the Gradient Reversal Layer Works

Forward Pass

- GRL acts as identity: $\text{GRL}(\mathbf{x}) = \mathbf{x}$
- Features flow normally to the domain classifier

Backward pass (gradient computation):

- GRL **multiplies** the gradient by $-\lambda$:

$$\frac{\partial \text{GRL}}{\partial \mathbf{x}} = -\lambda \mathbf{I}$$

- The domain classifier tries to **correctly predict** the domain (source vs. target)
- But the feature extractor receives the opposite gradient, so it learns to **make domains indistinguishable**

Result: the feature extractor produces representations where source and target distributions are **aligned**. The domain classifier cannot tell them apart.

Training dynamics

1. G_d improves at distinguishing domains
2. G_f improves at confusing G_d (via GRL)
3. G_y improves at the task using G_f features
4. At equilibrium: domain-invariant, task-relevant features

Limitation

Standard DANN aligns features globally without considering class structure. Classes that are far apart in the source may get mixed in the aligned space. ⇒ **CDAN** addresses this.

Ganin, Yaroslav, et al. "Domain-adversarial training of neural networks." *Journal of machine learning research* 17.59 (2016): 1-35.

Key Technique: Conditional Domain Adversarial Networks (CDAN)

Improvement over DANN (Long et al., NeurIPS 2018)

Standard DANN conditions domain discrimination **only on features**. CDAN conditions on **both features AND task predictions** via their outer product.

Why is this better?

- Different classes may shift **differently** across domains
- CDAN captures **class-conditional domain shifts**: "align class c in source with class c in target"
- Conditioning signal: $\mathbf{f} \otimes \mathbf{g}(x)$ where $\mathbf{f}$ = features, $\mathbf{g}$ = softmax predictions
- This multilinear map encodes **which class** the sample likely belongs to

Formulation:

- Domain classifier input: $\mathbf{h} = \mathbf{f} \otimes \mathbf{g}$
- Entropy conditioning (CDAN+E): weight samples by prediction entropy, focus adaptation on **hard, uncertain** samples
- Effect: easy-to-classify samples (low entropy) contribute less to adversarial loss

Key Technique: Pseudo-Labeling & Self-Training

Pseudo-Labeling

Use the model's own predictions on **unlabeled target data** as surrogate labels to expand the training set.

Confidence-based (FixMatch style)

- Keep only predictions above threshold τ
- $\hat{y} = \arg\max p(y \mid \mathbf{x})$ if max $\max p > \tau$
- Risk: **confirmation bias** if τ too low (wrong pseudo-labels reinforce errors)
- Risk: too few pseudo-labels if τ too high

Spatial consistency-based

- Exploit co-located source/target pixels (same geographic location, different years)
- No threshold needed, label consistency provides the signal

Self-Training Loop

1. Train model on labeled source data
2. Generate pseudo-labels on target
3. Retrain with source labels + pseudo-labels
4. Repeat iteratively

Teacher–Student (EMA)

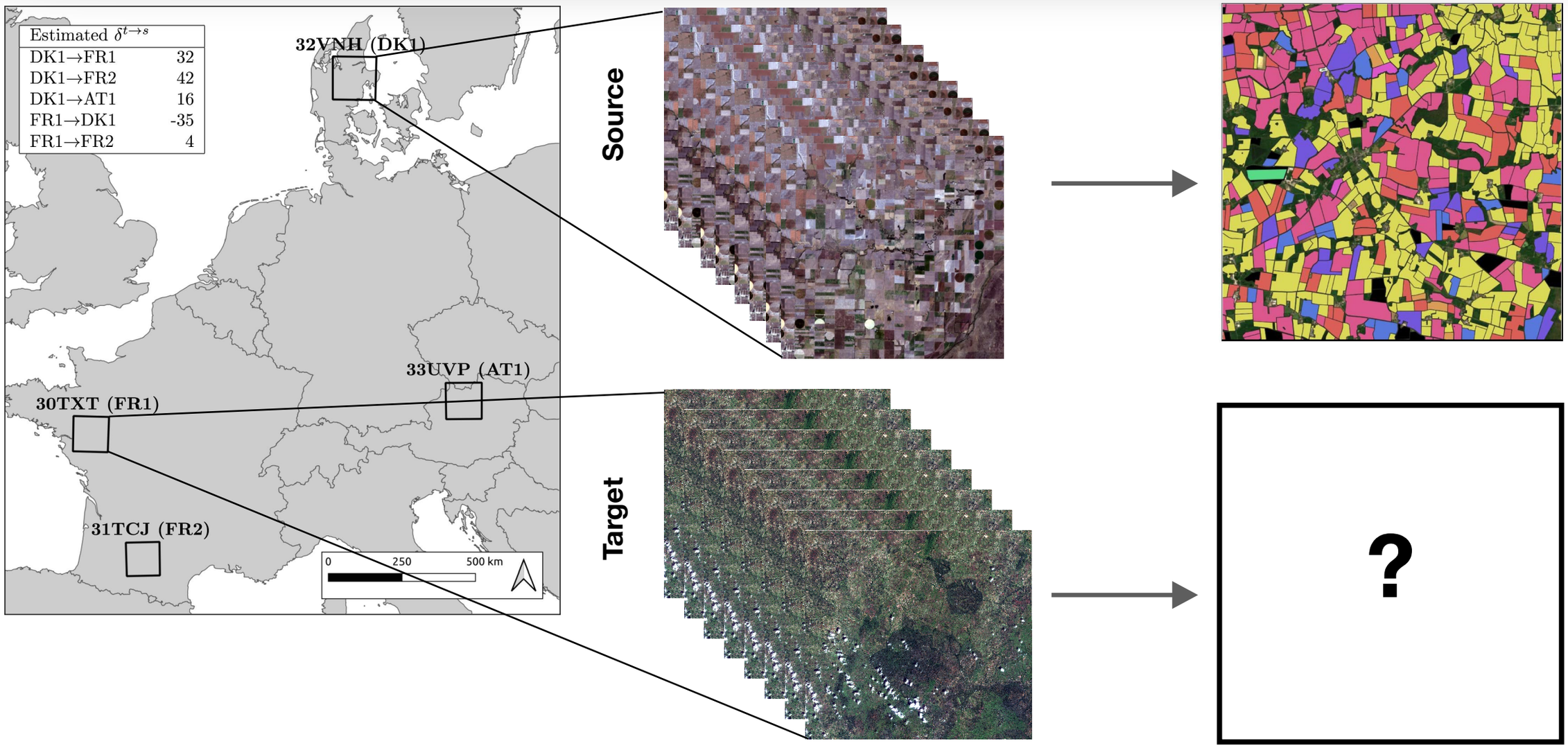
- Teacher: exponential moving average of student weights: $\theta'_t = \alpha\theta'_{t-1} + (1-\alpha)\theta_t$
- Provides more stable, less noisy pseudo-labels
- Used in TimeMatch (α = 0.9999)

Spatial Shift

TimeMatch: Problem Setting

Problem

Crop classifiers trained on SITS in one region fail when applied to a different geographic region, even for the same crop types.

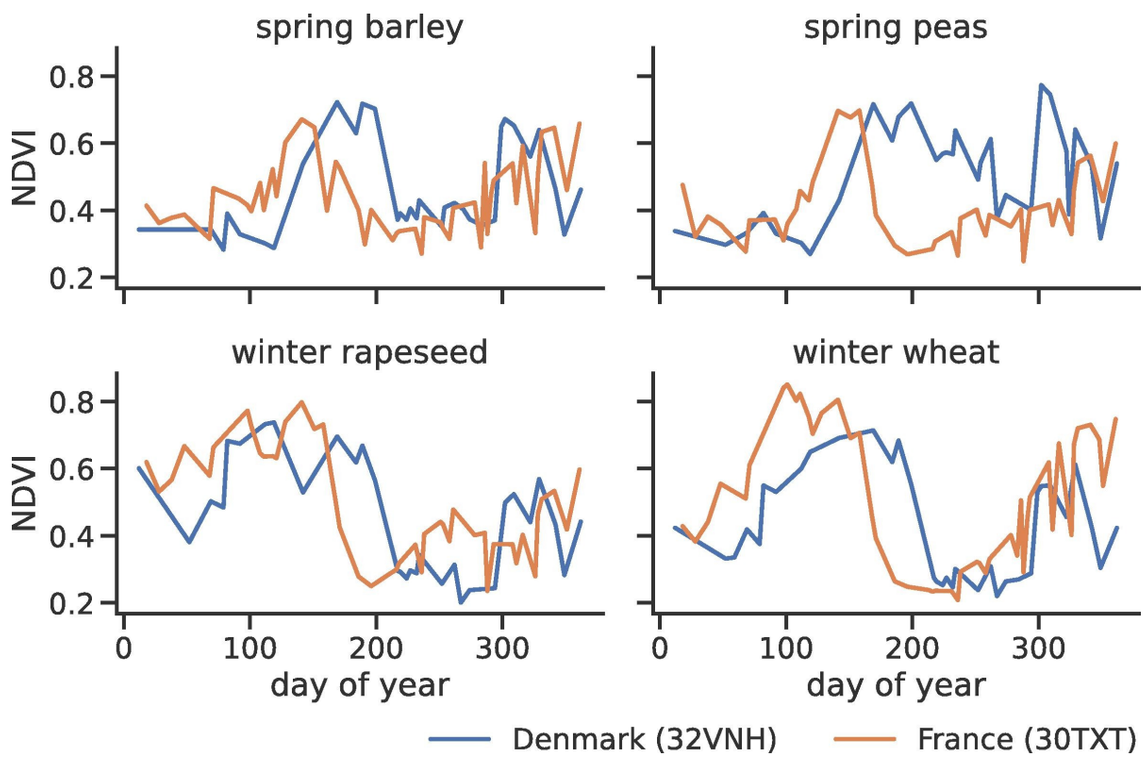


Nyborg, J., Pelletier, C., Lefèvre, S., & Assent, I. (2022). TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *188*, 301-313.

Why Does a Crop Classifier Fail Across Regions?

Root cause: Temporal shift of crop phenology

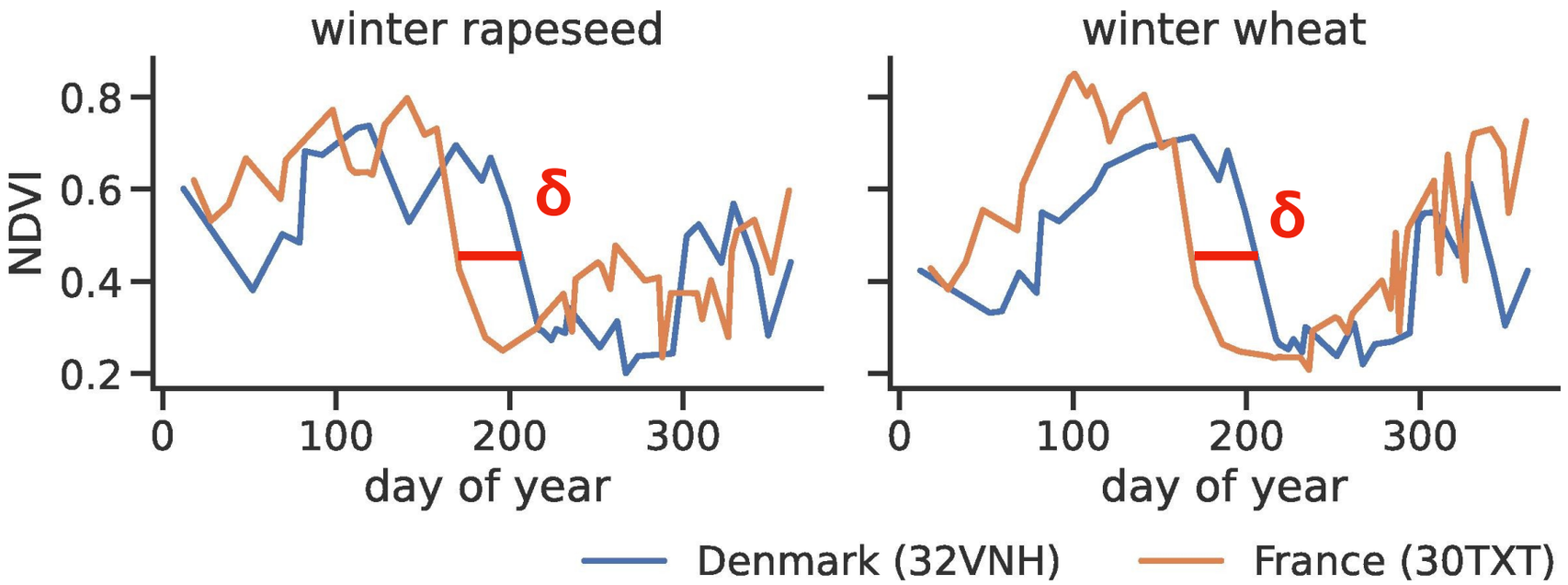
- Same crop species reach growth stages at different calendar dates depending on climate and latitude
- E.g., winter wheat peaks earlier in southern France than in Denmark (up to 42 days difference)
- Standard UDA methods (DANN, CDAN) ignore this temporal structure and treat SITS like generic features
- **Key insight:** the domain shift in SITS is largely temporal, aligning the time axis could solve most of the problem



Setup: model trained on a labeled region (source) → applied to an unlabeled region (target)

Nyborg, J., Pelletier, C., Lefèvre, S., & Assent, I. (2022). TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *188*, 301-313.

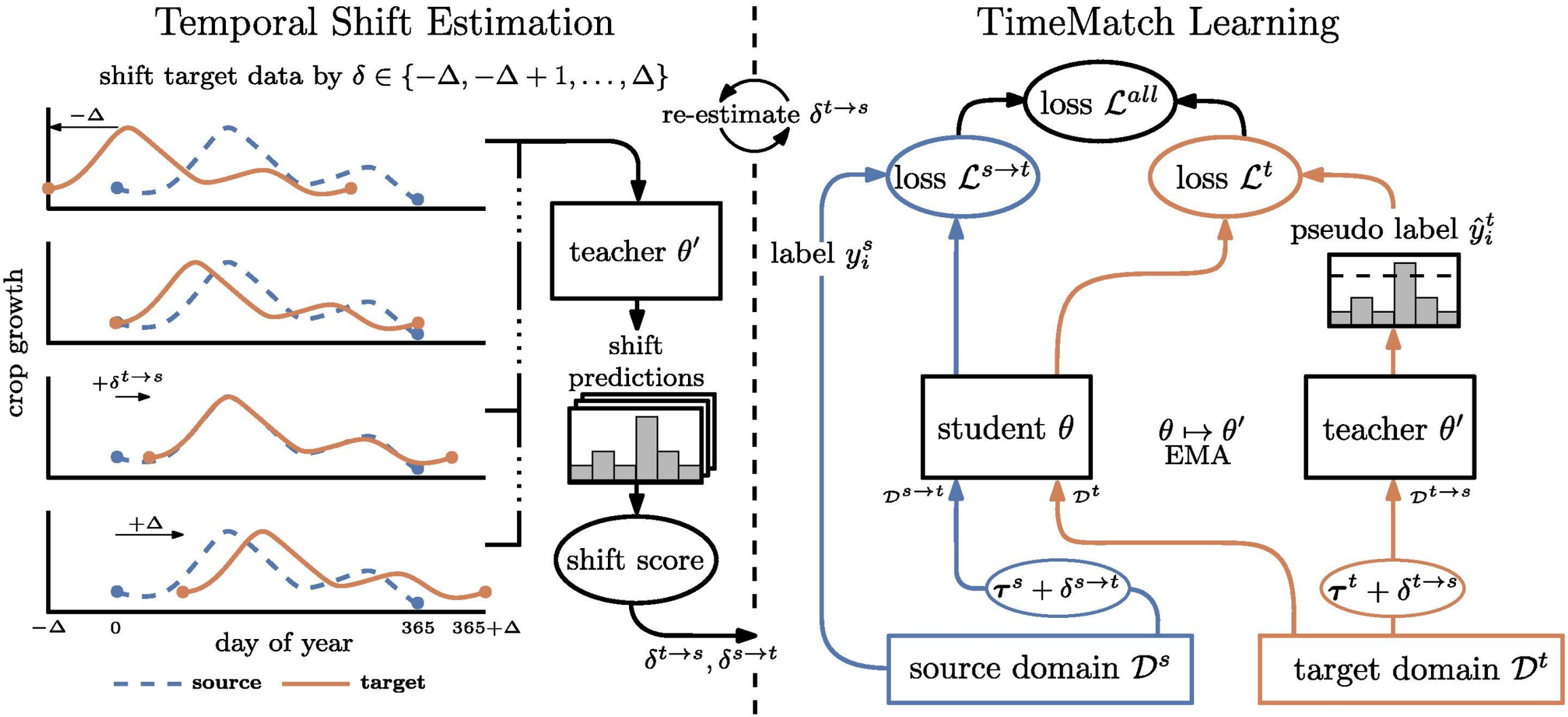
The Key Idea: Estimate the Temporal Shift Explicitly



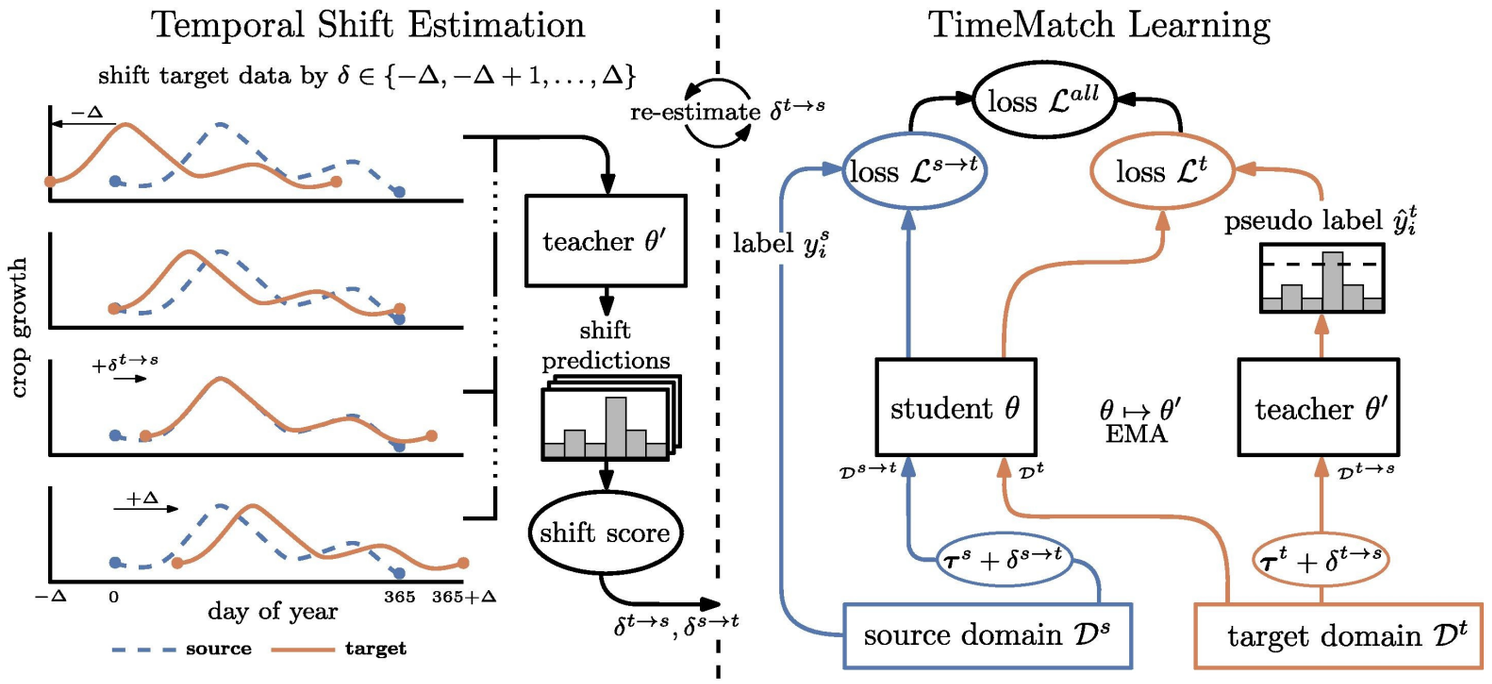
- **Key insight:** if I know by how many days the target's phenology is out of sync with the source, I can temporally realign the data before performing pseudo-labeling
- The temporal shift δ is represented as a scalar (number of days)
- **How to estimate it without labels in the target?** → We use the source-trained model: we test different shifts δ applied to the target data and choose the one that maximizes the confidence and diversity of the predictions
- In other words the model is "queried" with temporally shifted versions of the target data, and the best shift is the one that makes it appear more confident and balanced in its predictions.

Nyborg, J., Pelletier, C., Lefèvre, S., & Assent, I. (2022). TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *188*, 301-313.

The TimeMatch Algorithm: Self-Training with Temporal Alignment



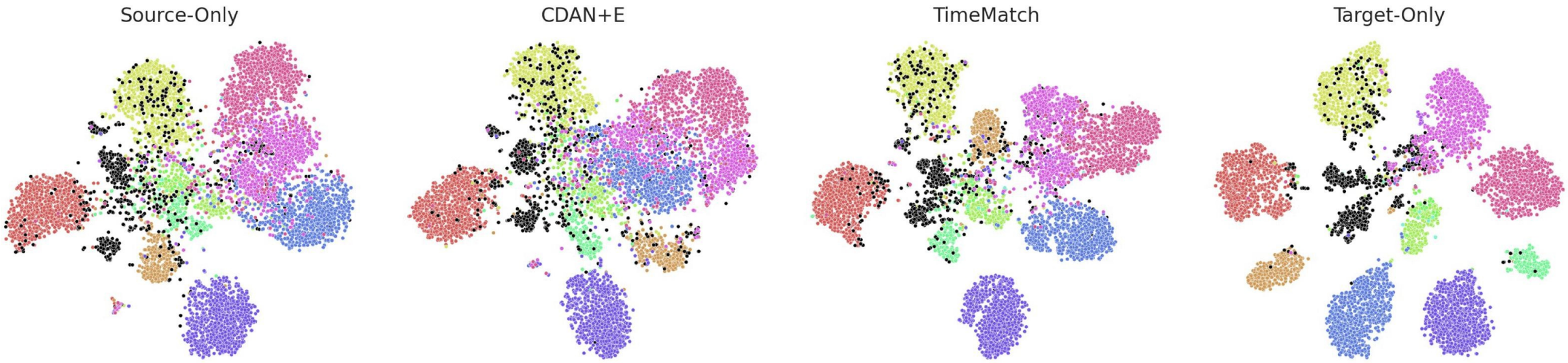
The TimeMatch Algorithm: Self-Training with Temporal Alignment

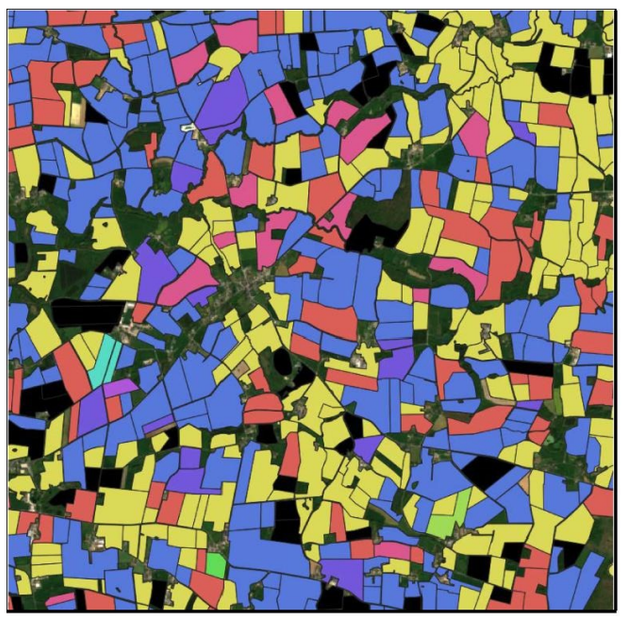


Two-stage approach:

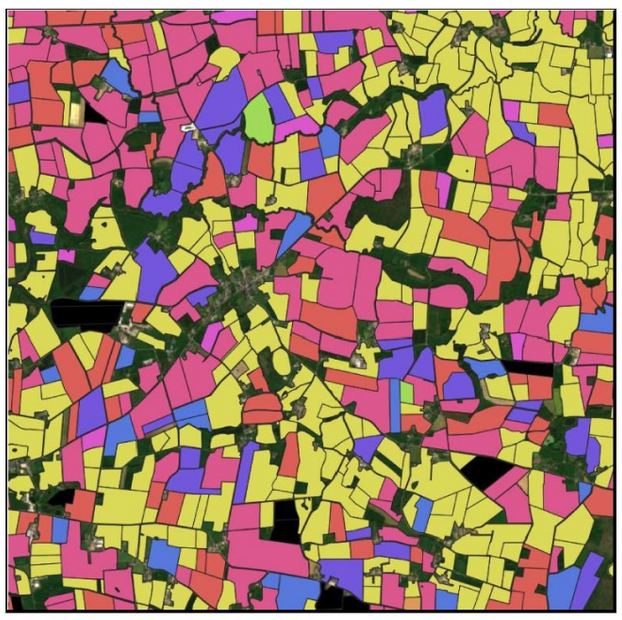
1. **Temporal Shift Estimation**
   - Apply source-trained model to target data shifted by candidate $\delta \in \{-\Delta, \ldots, +\Delta\}$
   - Select $\hat{\delta}$ that **maximizes prediction confidence** (Activation Maximization score)
   - Intuition: when temporal alignment is correct, the model is most confident
   - Refine iteratively using pseudo-label class distribution (per-class δ)
2. **TimeMatch Learning (Teacher–Student self-training)**
   - Teacher (θ') generates pseudo-labels on **temporally shifted** target data ($\delta^{t\to s}$)
   - Student (θ) trains on: source data shifted to target time ($\delta^{s\to t}$) + **confident** pseudo-labels (FixMatch loss with threshold τ)
   - Teacher updated via EMA: $\theta' \leftarrow \alpha\theta' + (1-\alpha)\theta$ with $\alpha = 0.9999$
   - Temporal shift δ is **re-estimated each epoch** as the model improves

Results

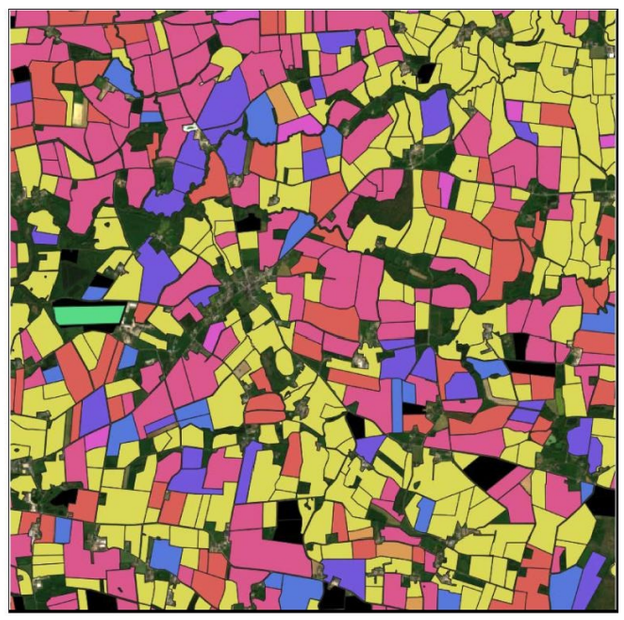




(a) Source-Only



(b) TimeMatch



(c) Ground Truth

Nyborg, J., Pelletier, C., Lefèvre, S., & Assent, I. (2022). TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *188*, 301-313.

Conclusions

Method	DK1→FR1	DK1→FR2	DK1→AT1	FR1→DK1	FR1→FR2	Avg
Source only	28.3	29.0	43.4	24.9	70.3	39.2
DANN	32.1	32.3	46.7	31.7	71.3	42.8
CDAN+E	39.3	37.9	51.5	36.5	71.7	47.4
FixMatch + ShiftAug	48.2	44.2	57.4	51.3	67.7	53.7
TimeMatch	**57.4**	**47.0**	**61.7**	**52.1**	**73.0**	**58.2**
Target trained	*74.6*	*72.4*	*86.9*	*90.6*	*85.7*	*82.0*

F1-Score

TimeMatch addresses the spatial shift (source and target are different regions within the same year)

Key contribution: the temporal shift between regions is the main cause of degradation → explicitly aligning it outperforms the classic feature-invariant approach

What if the problem were the opposite: same geographic area, but different years?

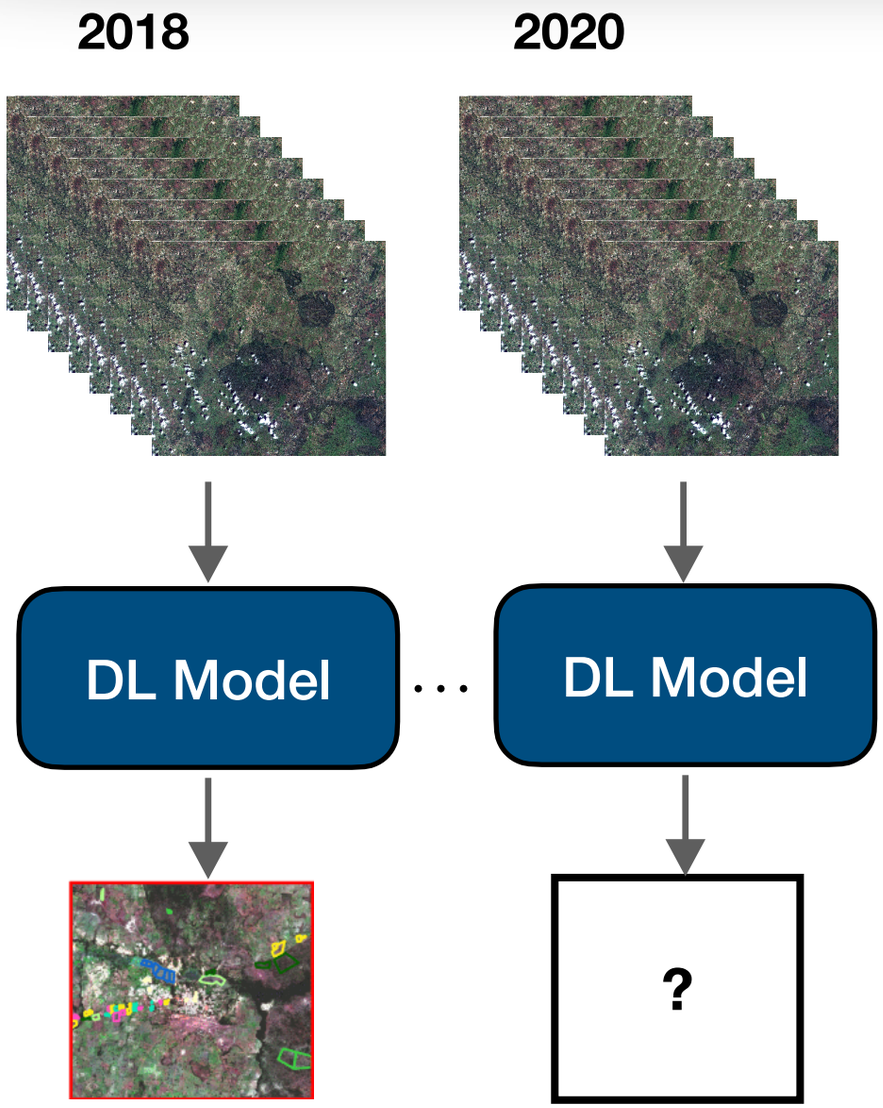
Nyborg, J., Pelletier, C., Lefèvre, S., & Assent, I. (2022). TimeMatch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *188*, 301-313.

Temporal Shift

The Problem: Same Place, Different Year

Problem

A land cover classifier trained on SITS from **year *t*** degrades when applied to **year *t + k*** over the **same area**, due to different inter-annual climate/weather conditions.



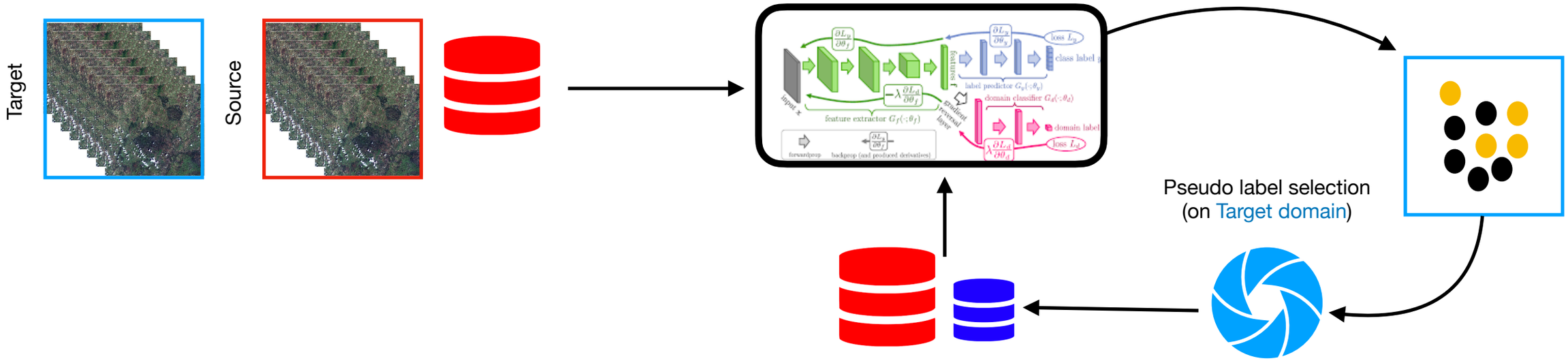
Setting: Temporal UDA (tUDA)

- Source: labeled SITS from year t
- Target: unlabeled SITS from year t + k
- **Same geographic area:** Koumbia, Burkina Faso
- 8 land cover classes: cereals, cotton, oleaginous/ leguminous, grassland, shrubland, forest, bare soil/ built-up, water
- Key difference from TimeMatch: **temporal** (not spatial) shift

Data

- Sentinel-2 (4 bands @ 10m), 24 dates/year
- **Sparse** annotations (real-world constraint — not every pixel is labeled)
- 3 transfer tasks: 2018→2020, 2018→2021, 2020→2021

The SpADANN Framework: Two Combined Strategies



The main ideas behind **SpADANN** is to :

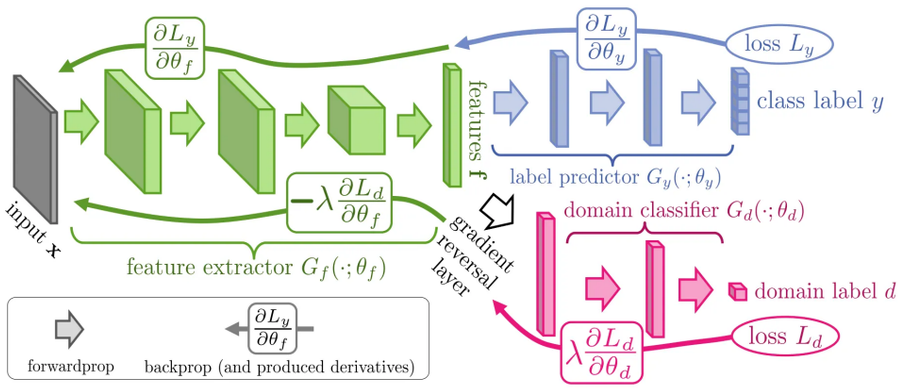
- Learn i**nvariant features w.r.t. distribution shifts** between source and target domain/year
- Identify reliable **areas** (pixels) **that remain stable** (in term of land cover) between domains/years
- **Gradually transfer** the classification model from source to target data

To this end, **SpADANN** combines

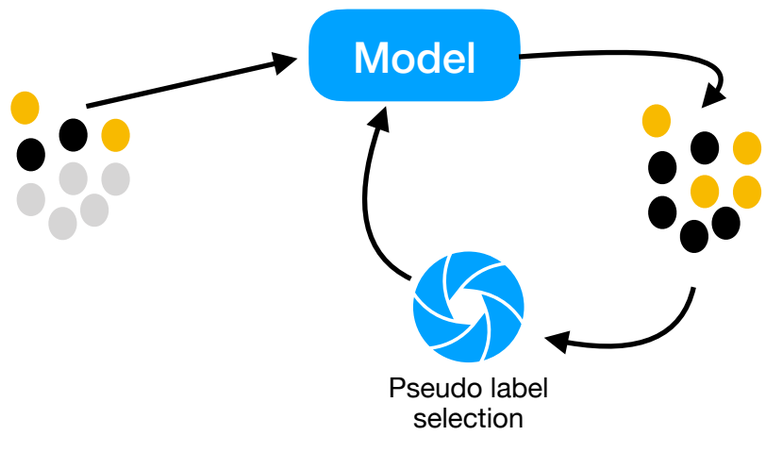
- Adversarial Learning
- Self-Training with Spatial Consistency

The SpADANN Framework: Two Combined Strategies

Adversarial Learning



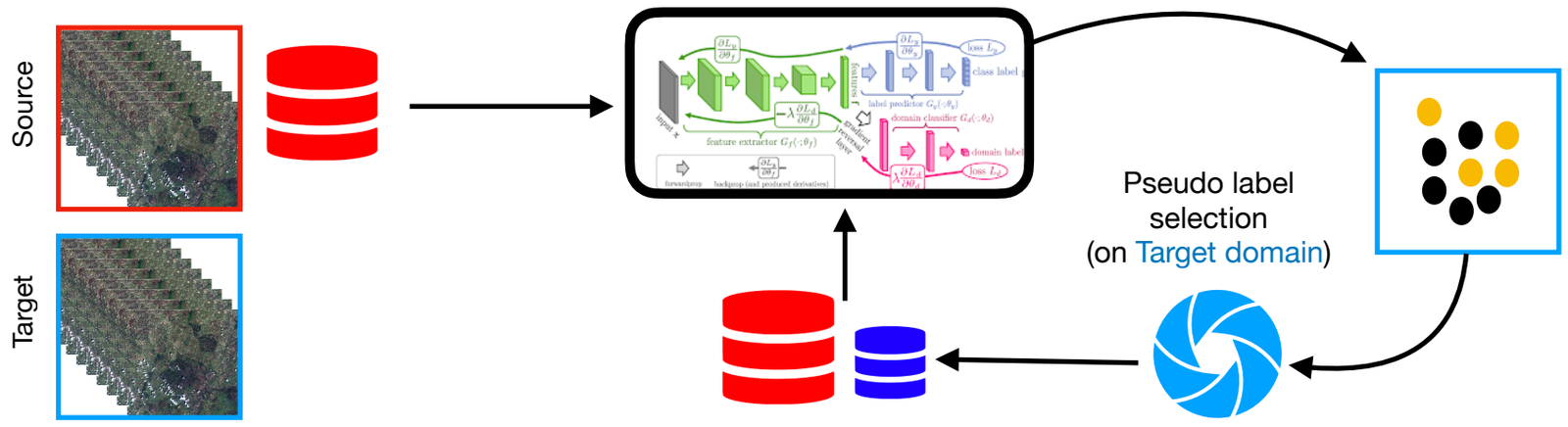
Self-Training



Two combined strategies:

1. **Adversarial learning** (DANN with GRL)
   - Encoder F : **TempCNN** (1D convolutions over time dimension)
   - Domain classifier D with gradient reversal → domain-invariant features
   - **Domain-specific batch normalization**: separate BN statistics for source and target (crucial for
   - handling distribution differences in mean/variance)
2. **Spatially-aware self-training** (novel contribution)
   - Exploit **co-located pixels**: same geographic coordinates observed in both source year and target year
   - A target pixel receives a pseudo-label if:
     1. Classifier predicts the same class for co-located source and target pixels
     2. The source prediction is correct (verified against source ground truth)
   - **No confidence threshold needed!** Spatial consistency replaces confidence filtering

The Key Idea: Spatially-Consistent Pseudo-labeling



Pseudo labels are chosen according to:

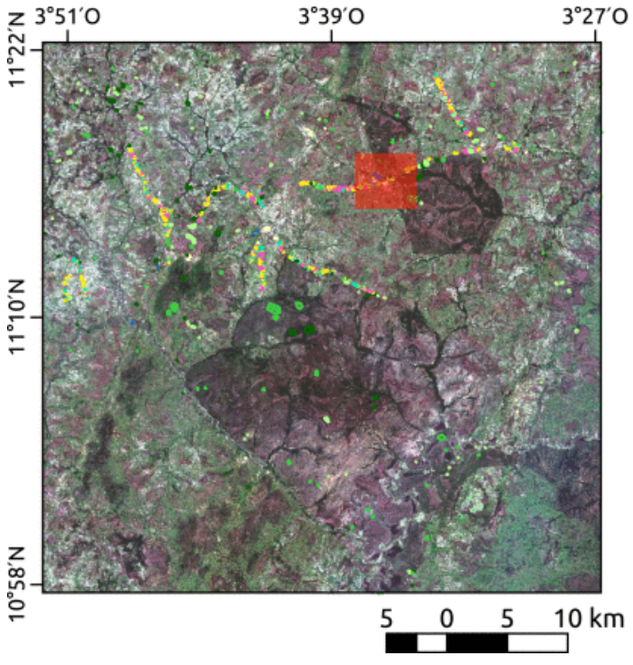
- **Spatial consistency** (the same area considering both source and target domain got the same LC class from the current classifier - tie points)
- **Confidence on source/training area** (Classifier predicts correct class for source time series)

The gradual transfer from Source to Target domain is ensured by the following loss function:

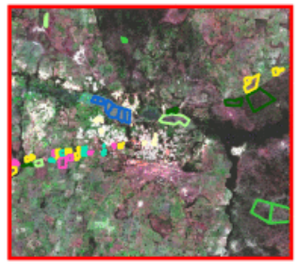
$$L = (1 - \alpha)\, L_{DANN} + \alpha \cdot \beta \cdot L_{pseudo}$$

- α increases linearly from 0 to 1 over training epochs
- β controls the relative weight of pseudo-label loss
- **Early training:** dominated by L_{DANN} (adversarial alignment)
- **Late training:** dominated by L_{pseudo} (self-training with spatial pseudo-labels)
- **Rationale:** adversarial alignment produces a coarse alignment first; then spatial pseudo-labels refine the decision boundary

SpADANN: Study Area
Koumbia, Burkina Faso









Legend:
- Cereals
- Cotton
- Oleag./Legum.
- Grassland
- Shrubland
- Forest
- B. Soil/Built-up
- Water

Why Koumbia?

- Sub-Saharan Africa: high agricultural variability, limited labeled data availability
- Strong **inter-annual variability**: rainfall and temperature patterns differ significantly year to year
- Mix of rainfed agriculture (cereals, cotton) and natural vegetation (forest, shrubland)
- Ground truth collected via field campaigns, expensive and time-consuming
- Representative of the **real-world challenge**: labeled data exists for one year but deploying the model for subsequent years requires DA

Results

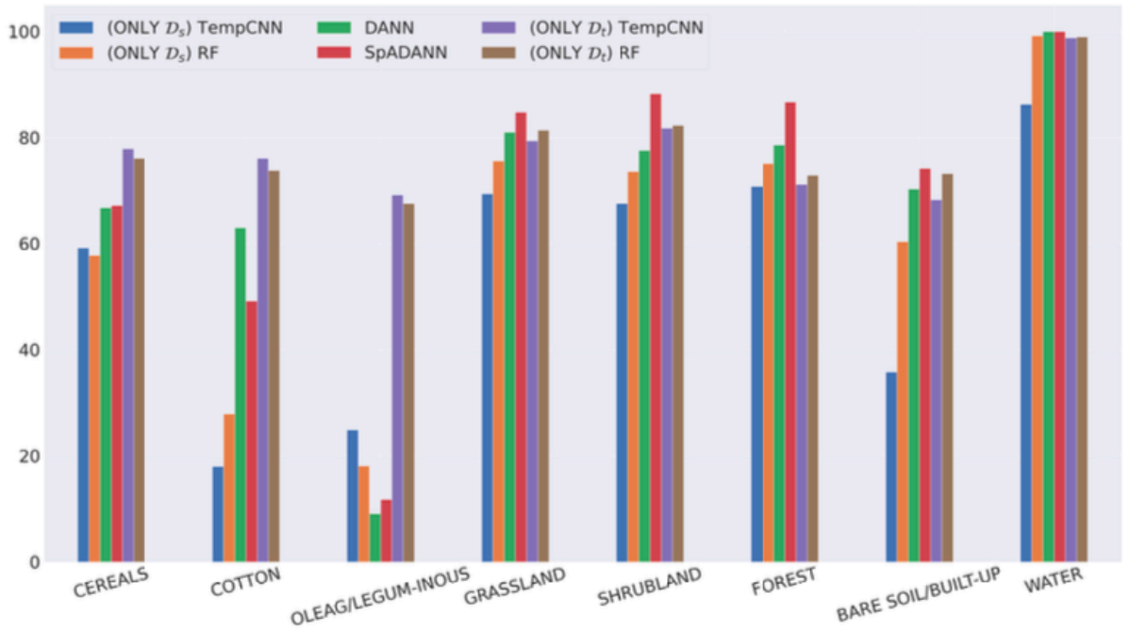


Figure: Per-class F1 scores for 2018→2020. SpADANN (red) outperforms all methods on most classes, especially **Cotton** and **Oleaginous/Leguminous**.

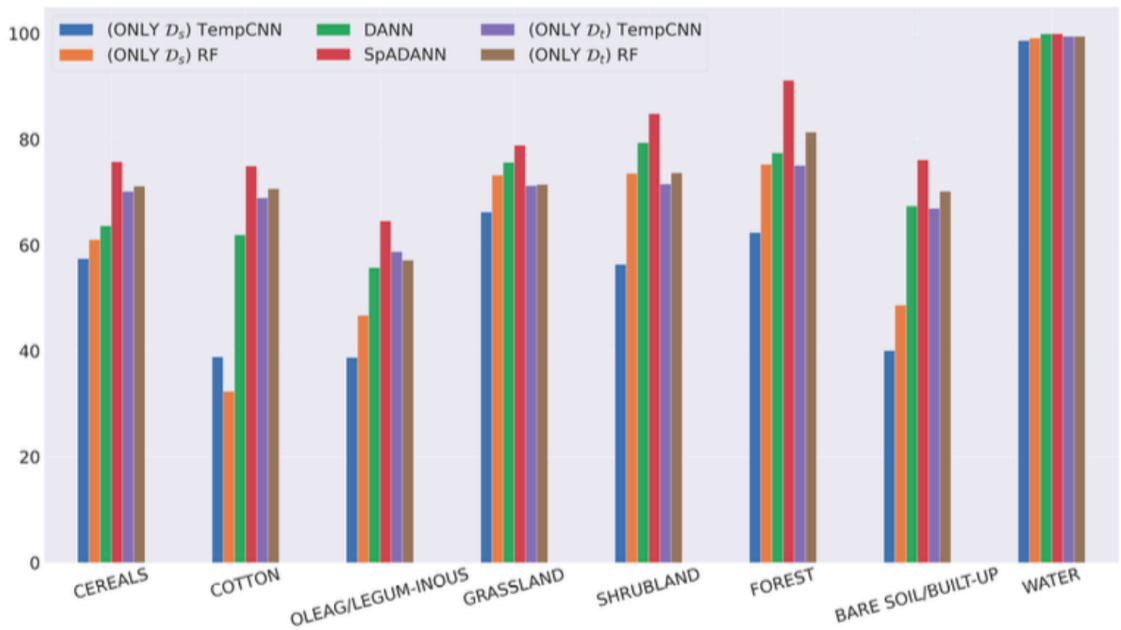


Figure: Per-class F1 scores for 2020→2021. SpADANN consistently improves over source-only (blue) and DANN (green) across all land cover classes.

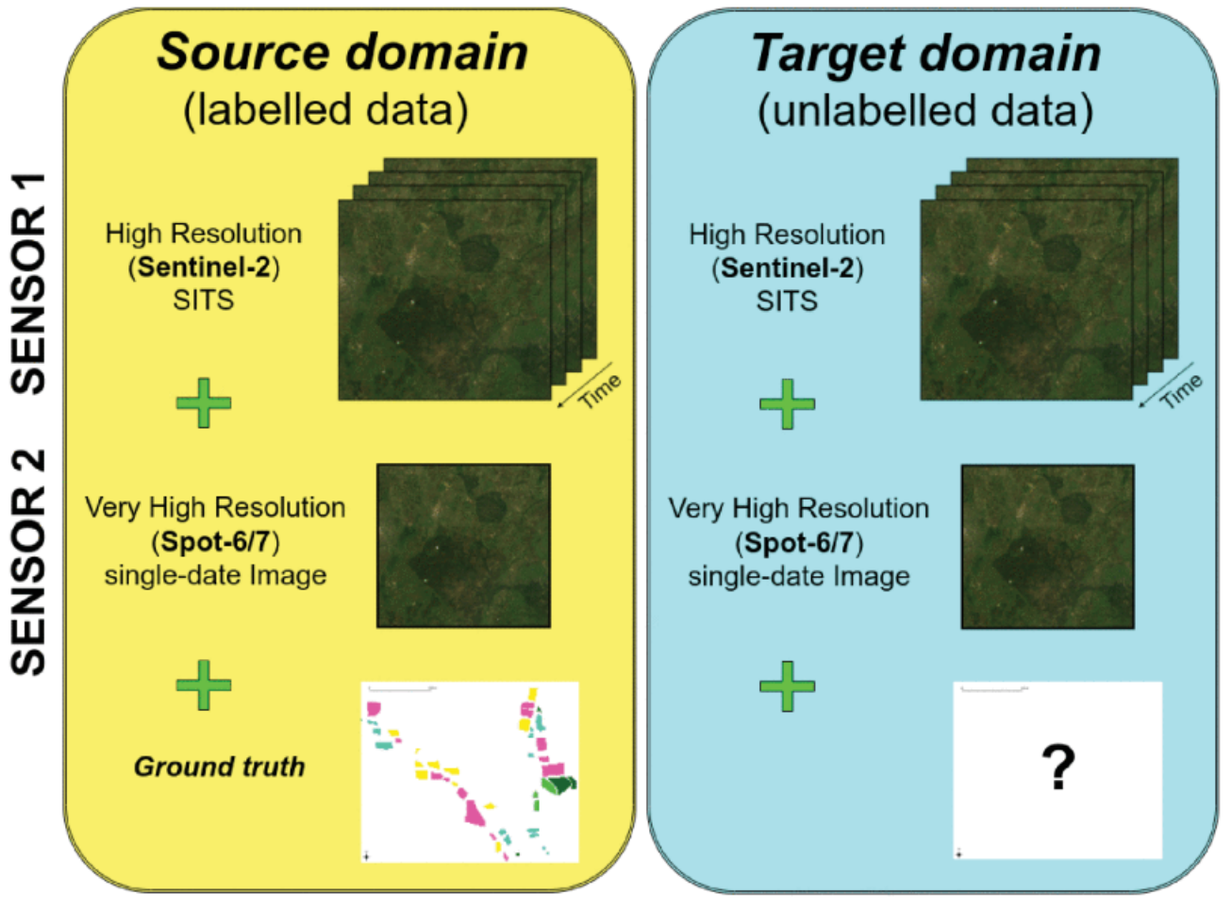
Observation: The hardest classes to transfer are those with the highest inter-annual variability (cotton, oleaginous crops), which are precisely the classes where SpADANN's spatial pseudo-labeling provides the greatest benefit.

Temporal Shift (Multi-Sensors)

From single-sensor to multisensor temporal transfer

Problem

Extend temporal UDA to the multi-sensor case: combine **SITS** (Sentinel-2) + **VHR imagery** (SPOT-6/7) for land cover mapping across years.



Why multi-sensor?

- SITS captures temporal dynamics (phenology, crop growth cycles)
- VHR captures spatial structure (texture, shape, fine-grained spatial patterns)
- Complementary information improves classification accuracy

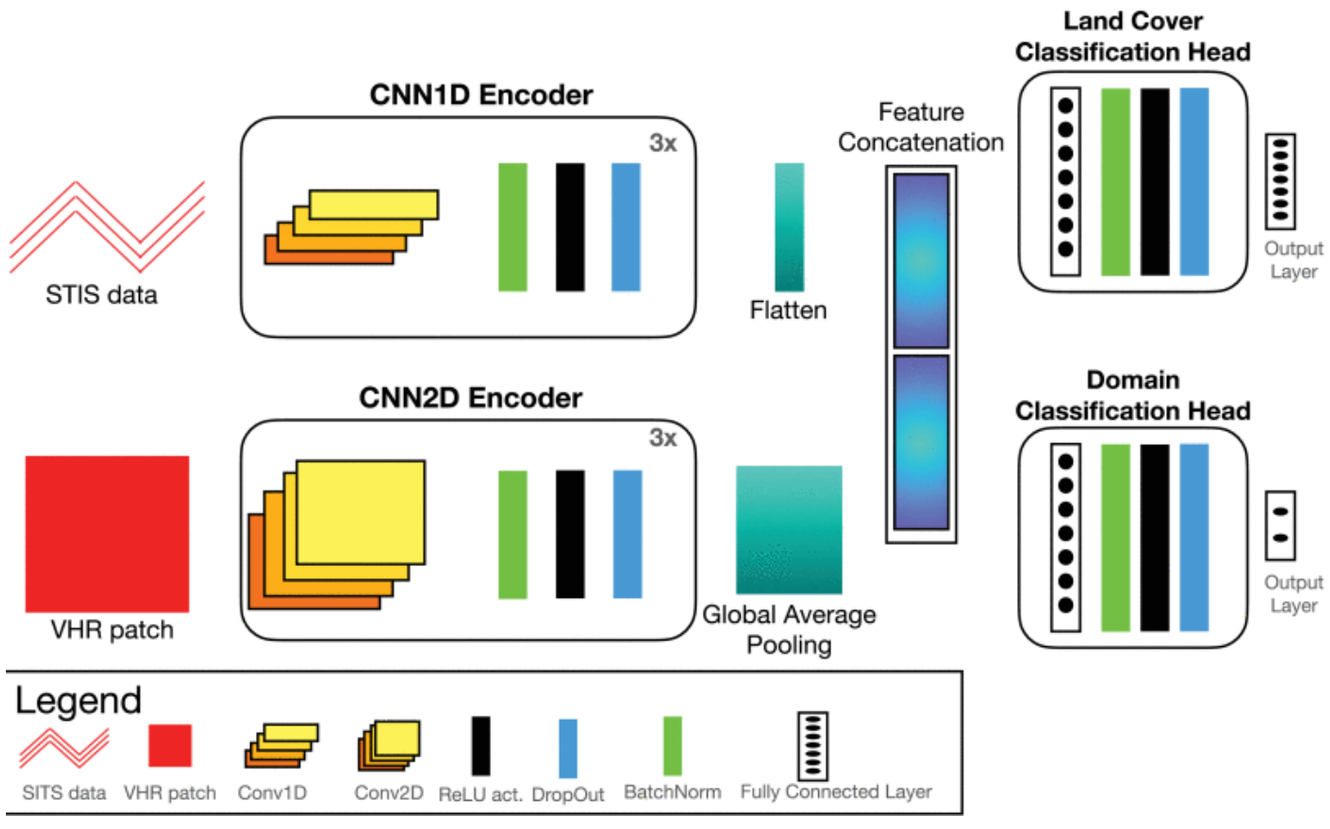
Data

- Same study area: Koumbia, Burkina Faso
- Sentinel-2 SITS: 24 images, 4 bands, 10m
- SPOT-6/7 VHR: single-date, 1.5m, 4 bands (B, G, R, NIR)
- Tasks: 2018→2021, 2021→2018
- 8 land cover classes (same as SpADANN)

Capliez, E., Ienco, D., Gaetano, R., Baghdadi, N., Salah, A. H., Le Goff, M., & Chouteau, F. (2023). Multisensor temporal unsupervised domain adaptation for land cover mapping with spatial pseudo-labeling and adversarial learning. *IEEE Transactions on Geoscience and Remote Sensing*, *61*, 1-16.

The multisensor backbone: two encoders, two heads

M3SPADA = Multisensor, Multitemporal, Multiscale Spatially-Aware Domain Adaptation

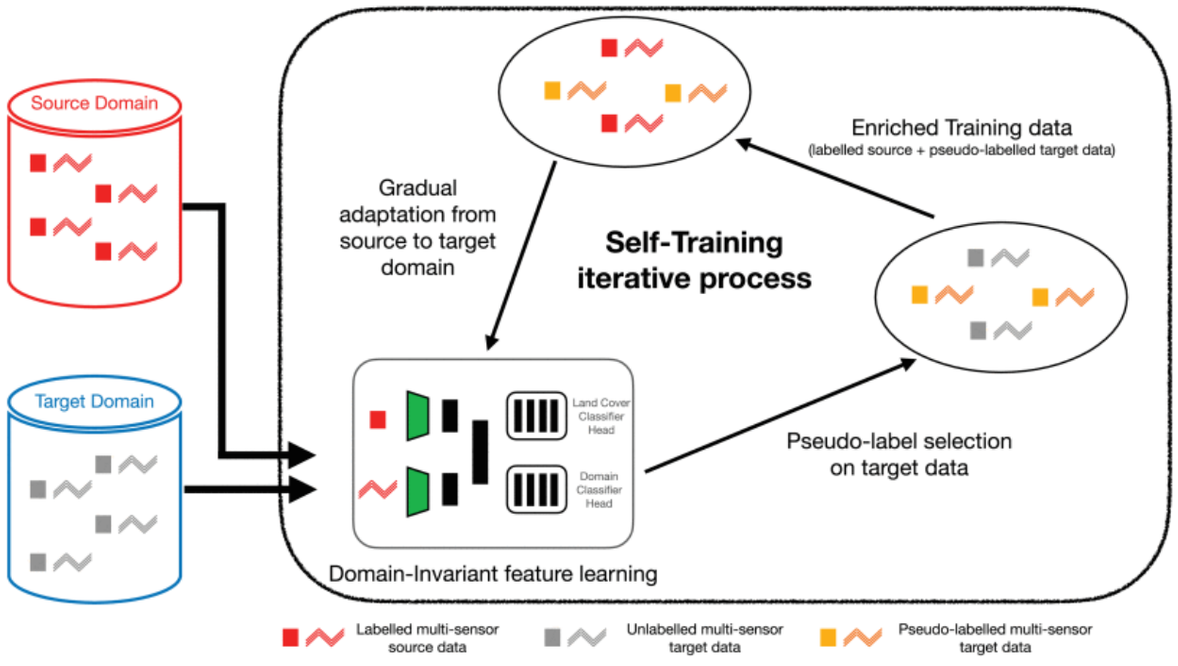


Two-branch late-fusion encoder:

- **SITS branch:** TempCNN (1D convolutions)
  - 3 blocks: Conv1D(64) + BN + ReLU + Dropout
  - Input: T ×B (24 dates ×4 bands)
- **VHR branch:** 2D CNN
  - 3 blocks: Conv2D(128) + BN + ReLU + Dropout
  - Input: 15×15 patches at 1.5m
  - Global Average Pooling
- Feature concate nation → shared classification head
- Auxiliary per-branch classifiers to strengthen individual representations

Note: During inference, the Domain Classification Head is discarded

M³SPADA: adversarial learning + spatially aware self-training



Adaptation strategy (extends SpADANN):

- Adversarial learning (GRL + domain classifier) applied to the **fused features**
- **Spatially-aware pseudo-labeling:** same threshold-free mechanism as SpADANN, now applied to multi-sensor features
- Progressive transfer: ε linear schedule shifts from adversarial to self-training
- Domain-specific batch normalization on **both branches**

Similarly to SpADANN:

$$L = (1 - \alpha)\, L_{DANN} + \alpha \cdot L_{pseudo}$$

Key difference from SpADANN: the addition of VHR spatial information via a second encoder branch provides richer features for both classification and adaptation.

Results

Key findings:

- Multi-sensor fusion (SITS + VHR) consistently outperforms mono-sensor ablations:
  - M3SPADA > M3SPADA$_{SITS}$ (SITS only) > M3SPADA$_{VHR}$ (VHR only)
  - VHR alone is weakest: temporal dynamics are essential for crop classification
  - But VHR adds complementary spatial texture information
- M3SPADA surpasses the supervised target-trained baseline (exploits both labeled source +
- pseudo-labeled target)
- Spatially-aware pseudo-labeling vastly superior to threshold-based variant (M3SPADA$_{Th}$ collapses
- due to confirmation bias)

TABLE III

WEIGHTED $F1$ SCORE AND ACCURACY OF THE COMPETING APPROACHES FOR THE TRANSFER TASK (2021 → 2018). THE BEST SCORE, FOR UDA METHODS, IS HIGHLIGHTED IN BOLD

Strategy	Method	F1-score	Accuracy
Only $\mathcal{D}_s$	$MSLCC_{VHR}$	52.8	55.9
	$MSCC_{SITS}$	60.3	61.2
	$MSLCC$	62.4	64.4
w/ UDA	ALDA	61.4	63.7
	WDGRL	69.1	70.0
	MDD	69.8	71.5
	DANN	72.7	73.1
	CDAN+E	74.1	74.2
	M^3SPADA_{VHR}	77.6	78.8
	M^3SPADA_{SITS}	88.4	88.2
	M^3SPADA	**89.9**	**89.4**
Only $\mathcal{D}_t$	$MSLCC_{VHR}$	72.3 ±1.2	72.8 ±1.2
	$MSCC_{SITS}$	78.3 ±1.3	78.2 ±1.3
	$MSLCC$	80.9 ±1.2	80.9 ±1.2

TABLE IV

$F1$ SCORE AND ACCURACY OF THE M^{3}SPADA ABLATIONS FOR THE TRANSFER TASK (2018 → 2021)

Method	F1-score	Accuracy
DANN	69.8	70.7
M^3SPADA_{Th}	63.2	67.5
M^3SPADA_{OnlyC1}	81.3	83.7
M^3SPADA	**83.0**	**84.5**

Progression from SpADANN

SpADANN (single sensor, temporal) →M3SPADA (multi-sensor, spatio-temporal)
Same core principles (adversarial + spatial pseudo-labels), extended to handle heterogeneous sensor inputs via late fusion.

Temporal Shift (Multi-Source)

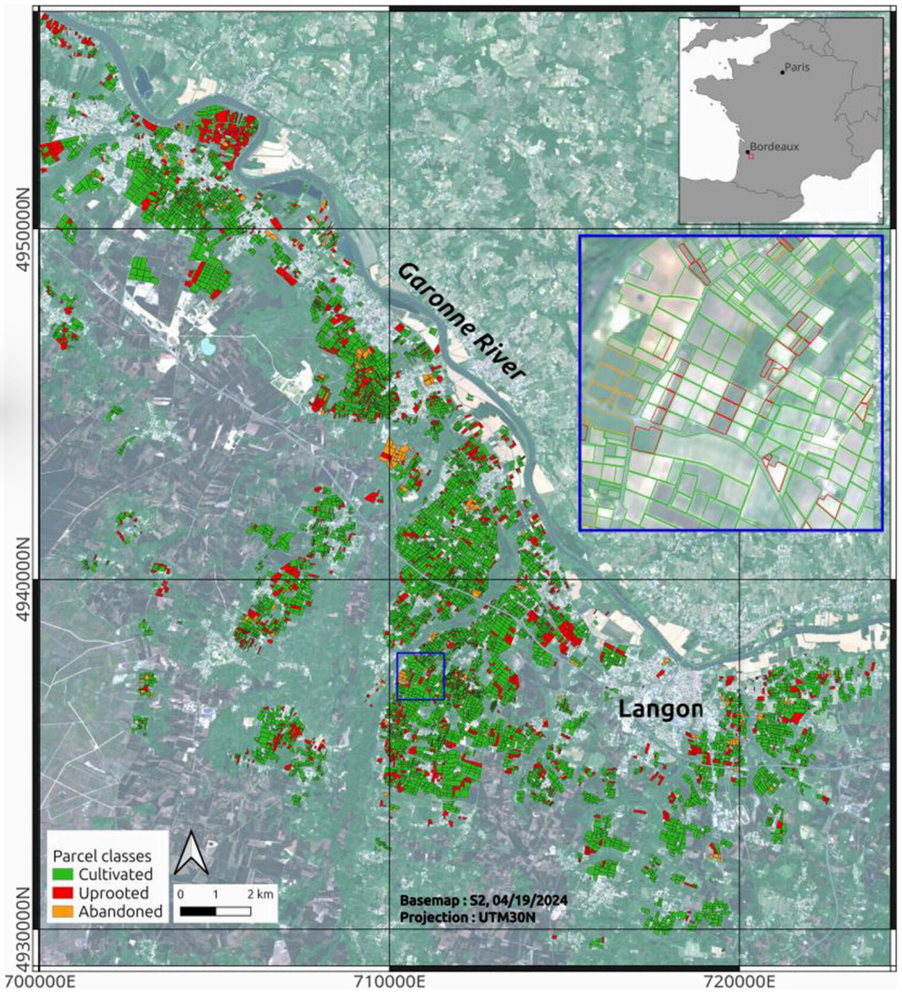
ConDA-Vine: Problem Setting

Problem

Detect abandoned vineyard parcels (disease reservoirs for Flavescence dorée) across years using multi-temporal satellite data.

Multi-source UDA setting:

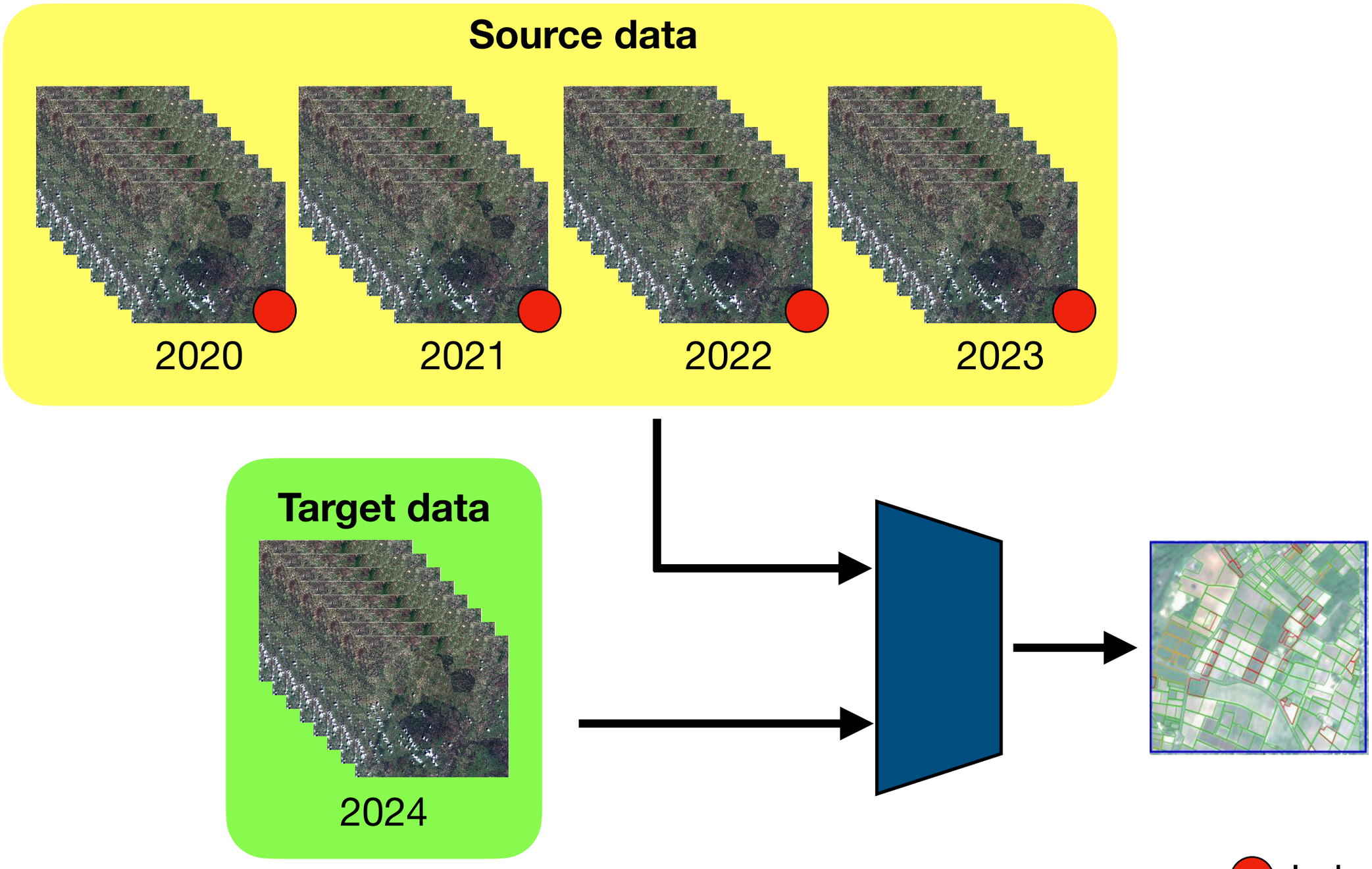
- Multiple source years treated as **separate domains**
- Target: one unseen year (leave-one-year-out protocol)
- Sentinel-2 (optical) + Sentinel-1 (SAR)
- 3 classes: cultivated, uprooted, abandoned
- Severe class imbalance: abandoned vineyards represent only 2–3.5% of parcels
- **Real-world application:** phytosanitary regulation requires identification of abandoned vineyards



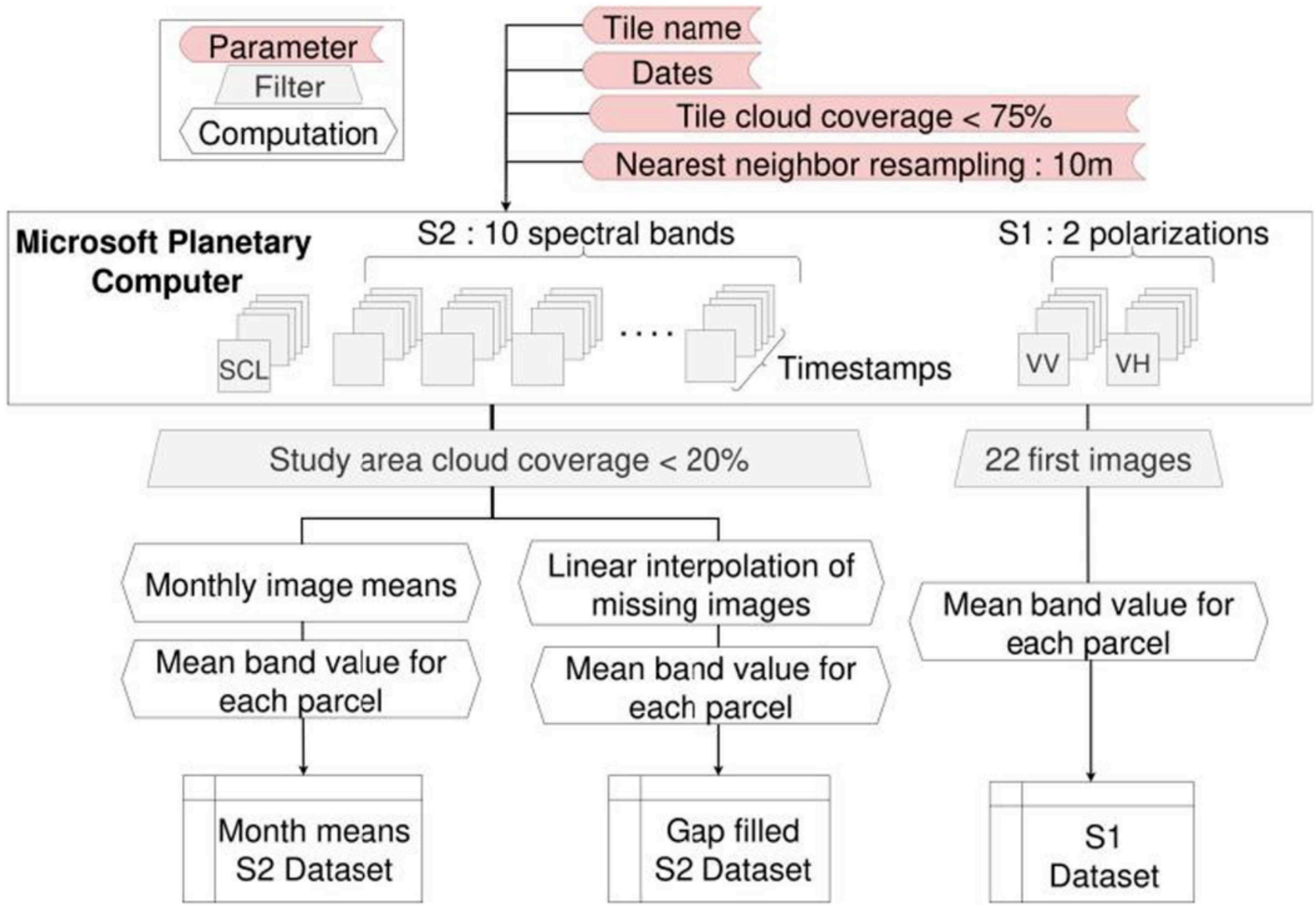
Year	Cultivated	Uprooted	Abandoned	Total
2020	2314	167	56	2537
2021	4850	1039	287	6176
2022	6696	1320	304	8320
2023	8243	1470	305	10018
2024	7964	1833	355	10152

Roussel, A., Alleaume, S., Billotte, A., Fabre, F., & Ienco, D. (2026). Out-of-year parcel identification using adversarial domain adaptation and multi-temporal satellite imagery to enhance vineyard health monitoring. *International Journal of Digital Earth*, *19*(1), 2634468.

Out-of-year classification as multi-source UDA

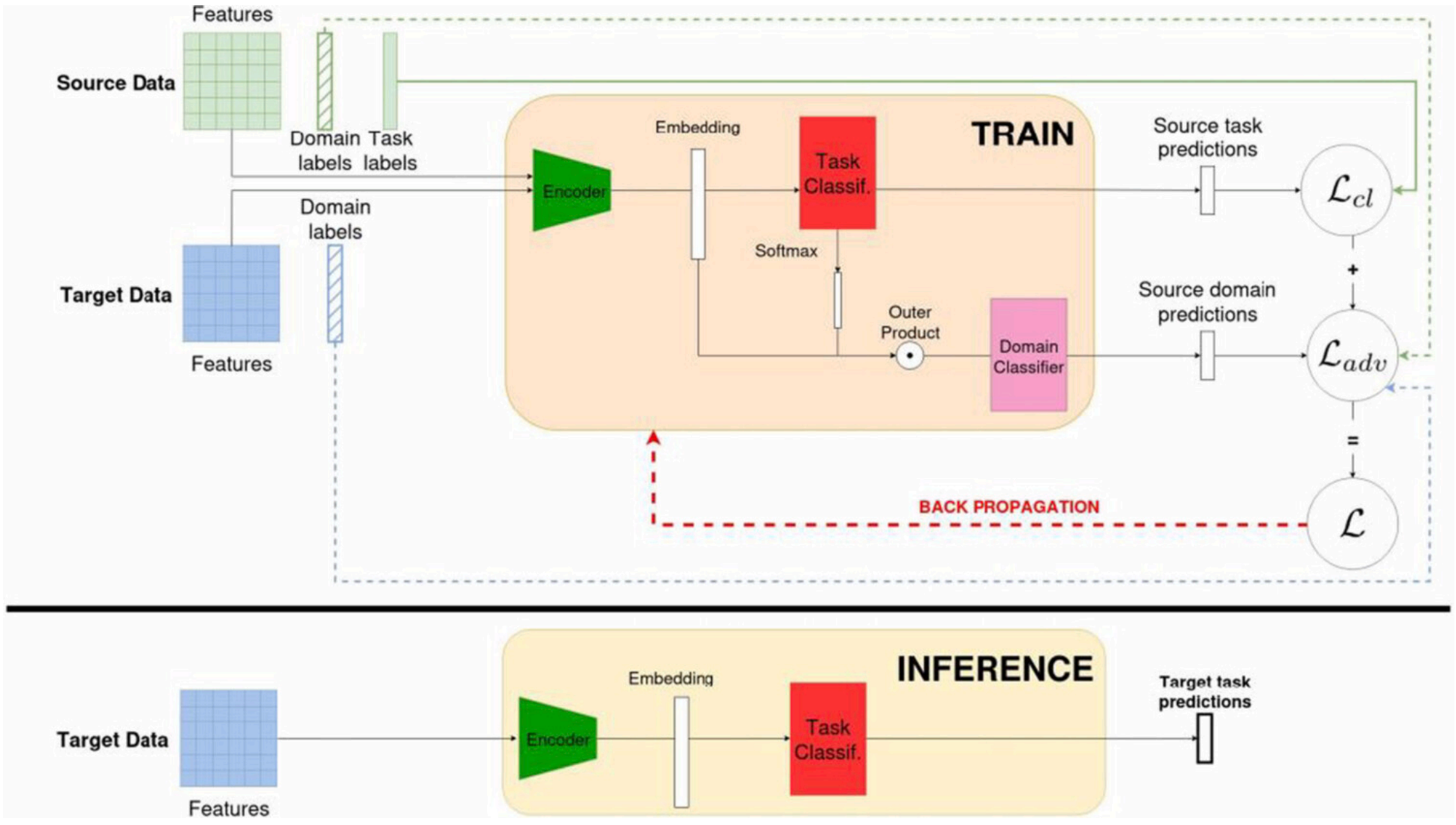


Sentinel-1 + Sentinel-2: multi-source, multi-temporal



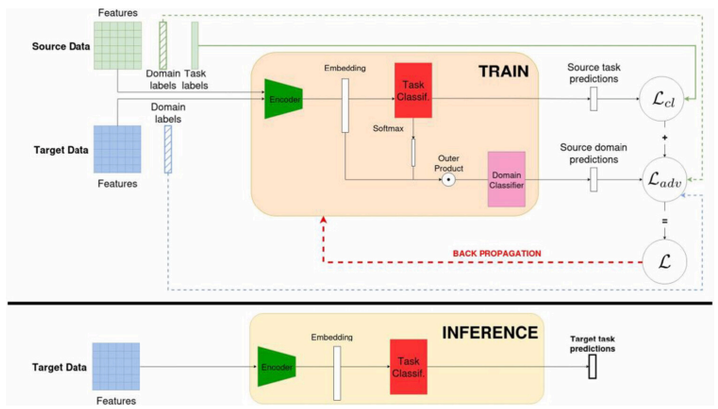
- Sentinel-2: 10 bands, 10 m resolution, cloud filtering, monthly aggregation (6 images/year, April–October)
- Sentinel-1: 22 images/year, VV+VH polarizations

ConDA-Vine: conditional adversarial DA for multi-year classification



[1] Long, M., Cao, Z., Wang, J., & Jordan, M. I. (2018). Conditional adversarial domain adaptation. *Advances in neural information processing systems*, *31*.

ConDA-Vine: conditional adversarial DA for multi-year classification



Based on CDAN (Conditional Domain Adversarial Network, Long et al. 2018):

- Standard DANN conditions domain discrimination only on features
- **CDAN conditions on both features AND task predictions** (outer product)
- Captures **class-conditional domain shifts**: different vineyard statuses shift differently across years

Architecture:

- Encoder: MLP (2 hidden layers, 256 neurons each, BN + ReLU + Dropout)
- Task classifier: 3-class output (cultivated / uprooted / abandoned)
- Domain classifier: conditioned on outer product $\mathbf{f} \otimes \mathbf{g}$ of features and softmax predictions
- GRL for adversarial training
- Loss: $L = L_{cl} + L_{adv}$

Multi-source extension: all source years treated as separate domains; domain classifier has |M| + 1 outputs (|M| source years + 1 target).

[1] Long, M., Cao, Z., Wang, J., & Jordan, M. I. (2018). Conditional adversarial domain adaptation. *Advances in neural information processing systems*, *31*.

Results

Overall performances

Method	Precision	Recall	W-F1-score
RF-*AllData*	**0.78**	0.44	0.46
RF-*Ensemble*	0.57	0.49	0.48
XGBoost-*AllData*	0.63	0.52	0.54
XGBoost-*Ensemble*	0.41	0.41	0.35
MLP-*AllData*	0.61	0.59	0.57
MLP-*Ensemble*	0.47	0.44	0.33
ConDA-Vine-Ensemble	0.60	0.65	0.60
ConDA-Vine	0.67	**0.66**	**0.65**

Per-class performances

Method	Cultivated	Uprooted	Abandoned
RF-*AllData*	**0.92**	0.40	0.05
RF-*Ensemble*	0.85	0.45	0.15
XGBoost-*AllData*	0.91	0.47	0.24
XGBoost-*Ensemble*	0.68	0.29	0.07
MLP-*AllData*	0.88	0.49	0.34
MLP-*Ensemble*	0.60	0.36	0.04
ConDA-Vine-Ensemble	0.89	0.54	0.37
ConDA-Vine	**0.92**	**0.60**	**0.45**

The main gain on the minority classes (uprooted and abandoned), which are essential for this task.

Conclusions



- Multi-source UDA allows for the combination of field data from different years without ignoring inter-annual domain shifts.
- Class imbalance + DA is a challenging but manageable combination: conditional adversarial learning aids class-wise alignment even for rare classes.
- **Operational application:** ConDA-Vine can be used as a pre-screening tool to focus field surveys on at-risk areas, reducing operational costs.

Sensor Shift

The Heterogeneous DA Challenge

New difficulty level

Source and target domains differ not only in distribution, but in **modality**:

optical ↔SAR, RGB ↔multispectral, different spatial resolutions.

The input feature spaces have **different dimensionality** and represent different physical quantities

Why is this harder than homogeneous DA?

- Feature spaces have **different dimensions** (13 optical bands vs. 2 SAR bands)
- No pixel-level correspondence between modalities (optical measures reflectance, SAR measures backscatter)
- Standard DA assumes a **shared feature space**. This assumption breaks
- Requires **separate encoders** per modality to map into a common representation
- The learned common space must be meaningful for **both** modalities

Semi-supervised setting (SSHDA):

- Source: fully labeled (one modality)
- Target: **few labels** + many unlabeled (different modality)
- Realistic assumption: some annotations are always available, even if scarce
- Theory: Fang et al. (TPAMI 2023) provide generalization bounds for SsHeDA, showing that **both** domain alignment and target supervision are needed

The Datasets

Dataset	Volume	Modality	Spatial Res.	Classes
EuroSAT-MS-SAR	13 500 × 13 × 64 × 64 13 500 × 2 × 64 × 64	MS SAR	10m 10m	10
RESISC45-Euro	5 600 × 3 × 256 × 256 24 000 × 13 × 64 × 64	RGB MS	0.2m - 30m 10m	8

New difficulty level

The two modalities contain images captured in different times and zones

→ **Spatio-Temporal-Sensor Shift**

SHeDD (**S**emi-supervised **He**terogeneous **D**omain Adaptation via **D**isentanglement)

Source data
(e.g. Radar)



x_s

Target data
(e.g. Optical)



x_t

x_u

Available datasets (notation)

$$S := \{(\mathbf{x}_s, y_s)^{(i)}\}_{i=1}^{N_s}$$ Labelled source data

$$T := \{(\mathbf{x}_t, y_t)^{(i)}\}_{i=1}^{N_t}$$ Labelled target data

$$U := \{\mathbf{x}_u^{(i)}\}_{i=1}^{N_u}$$ Unlabelled target data

C. F. Dantas, R. Gaetano, and D. Ienco, “Semi-supervised heterogeneous domain adaptation via disentanglement and pseudo-labelling,” in Proc. ECML-PKDD, 2024, pp. 1–17.

SHeDD (Semi-supervised Heterogeneous Domain Adaptation via Disentanglement)

Source data
(e.g. Radar)



x_s

Available datasets (notation)

$$S := \{(\mathbf{x}_s, y_s)^{(i)}\}_{i=1}^{N_s}$$ Labelled source data

$$T := \{(\mathbf{x}_t, y_t)^{(i)}\}_{i=1}^{N_t}$$ Labelled target data

$$U := \{\mathbf{x}_u^{(i)}\}_{i=1}^{N_u}$$ Unlabelled target data

How to leverage these datasets?

Target data
(e.g. Optical)



x_t

Target encoder g_t

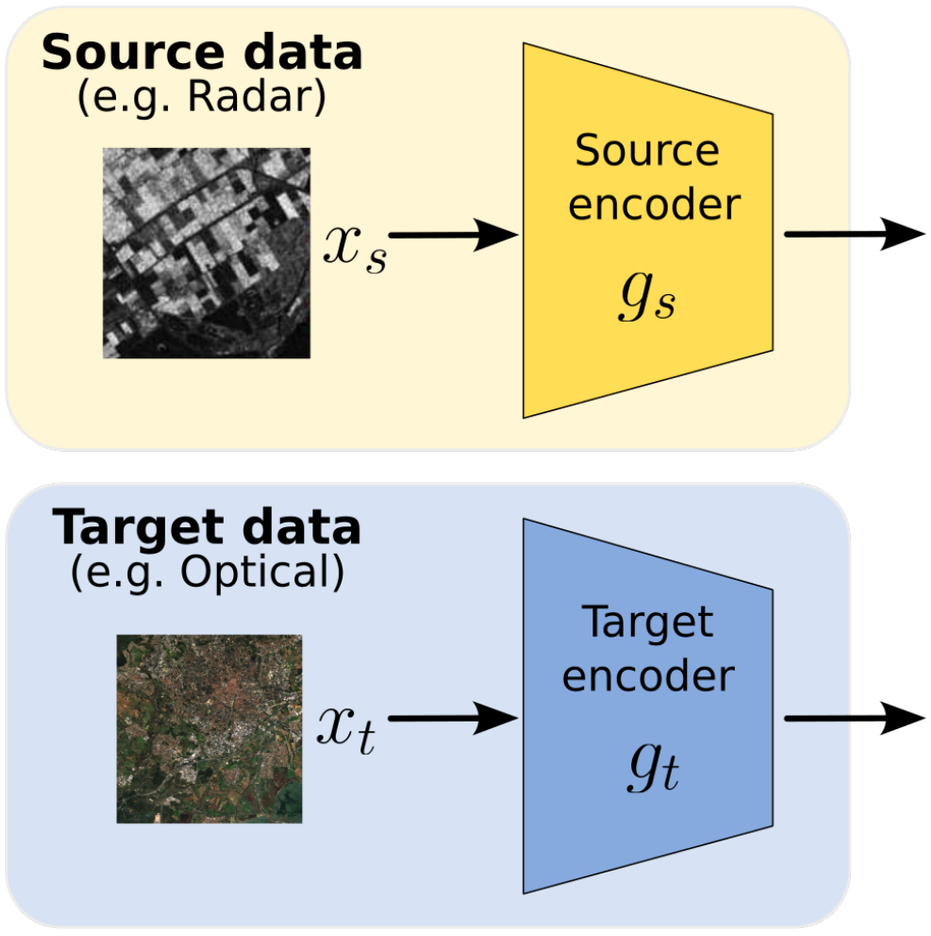
z_t

Task classif. f_{cl}

$\hat{y}$

C. F. Dantas, R. Gaetano, and D. Ienco, “Semi-supervised heterogeneous domain adaptation via disentanglement and pseudo-labelling,” in Proc. ECML-PKDD, 2024, pp. 1–17.

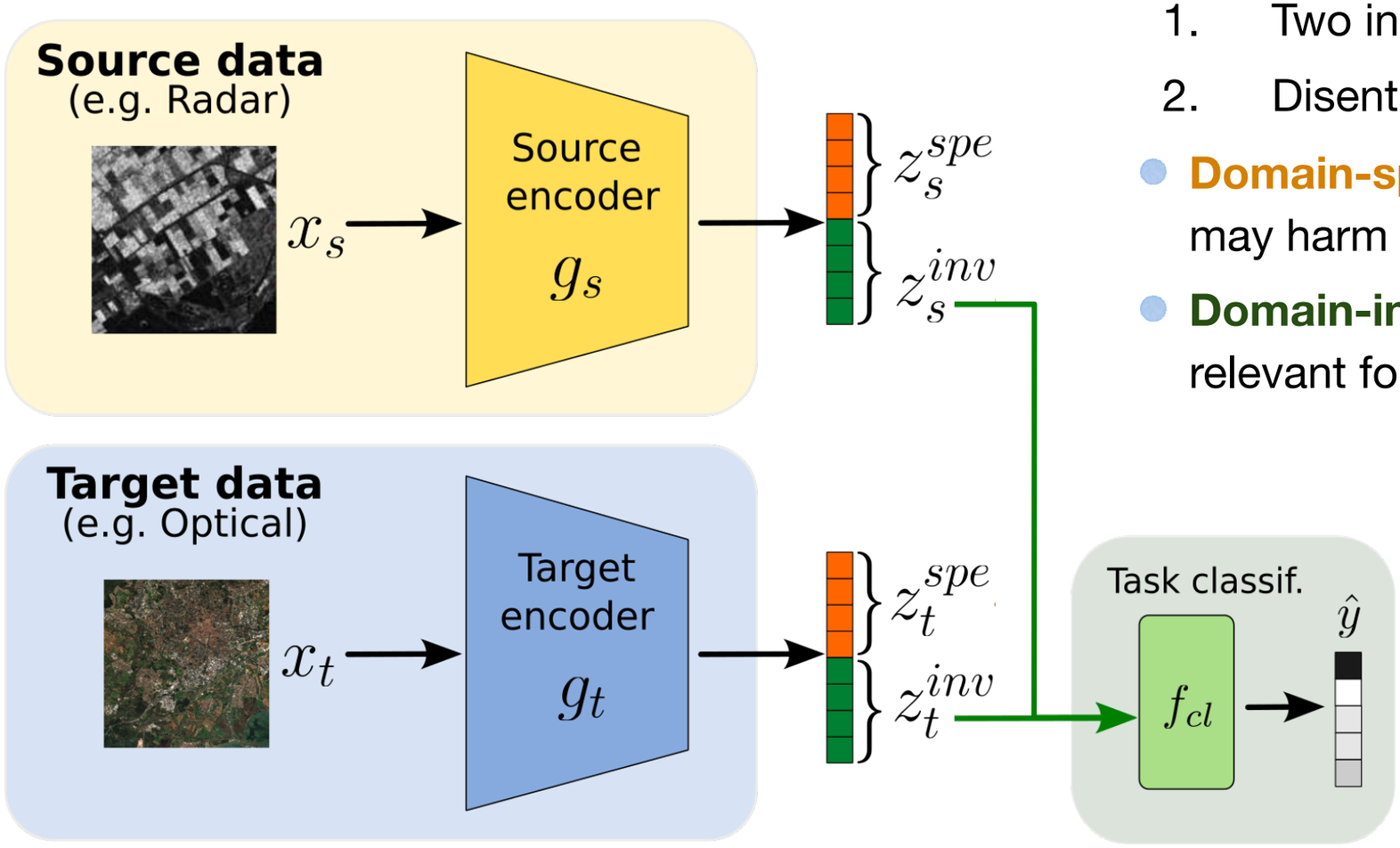
Leveraging source data (heterogeneous)



1. Two independent encoder branches

C. F. Dantas, R. Gaetano, and D. Ienco, “Semi-supervised heterogeneous domain adaptation via disentanglement and pseudo-labelling,” in Proc. ECML-PKDD, 2024, pp. 1–17.

Leveraging source data (heterogeneous)



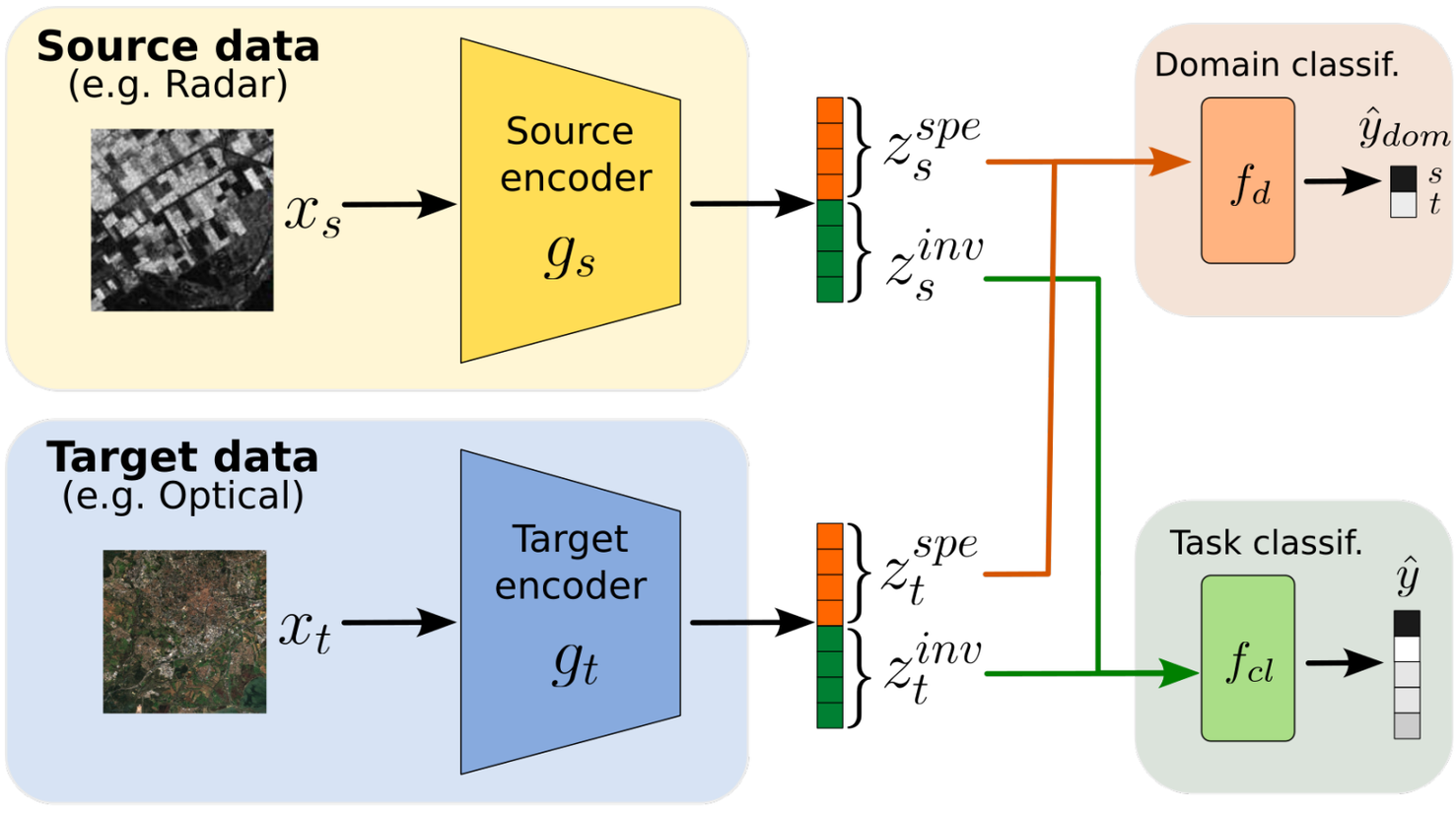
1. Two independent encoder branches
2. Disentangle representations

- **Domain-specific**:
  may harm cross-modality transfer
- **Domain-invariant**:
  relevant for the downstream task

C. F. Dantas, R. Gaetano, and D. Ienco, “Semi-supervised heterogeneous domain adaptation via disentanglement and pseudo-labelling,” in Proc. ECML-PKDD, 2024, pp. 1–17.

Leveraging source data (heterogeneous)

- A **domain classifier** is added to enforce channelling domain-specific information to
  → Predict if embedding comes from source or target domain
- Enforce orthogonality between domain-specific and domain-invariant features



$$\mathcal{L}_{dom} = \mathrm{CE}\left(f_d(\mathbf{z}^{spe}), y_{dom}\right)$$

$$\mathcal{L}_{\perp} = \frac{\langle \mathbf{z}^{inv}, \mathbf{z}^{spe} \rangle}{\|\mathbf{z}^{inv}\|_2 \|\mathbf{z}^{spe}\|_2}$$

$$\mathcal{L}_{cl} = \mathrm{CE}\left(f_{cl}(\mathbf{z}^{inv}), y\right)$$

C. F. Dantas, R. Gaetano, and D. Ienco, “Semi-supervised heterogeneous domain adaptation via disentanglement and pseudo-labelling,” in Proc. ECML-PKDD, 2024, pp. 1–17.

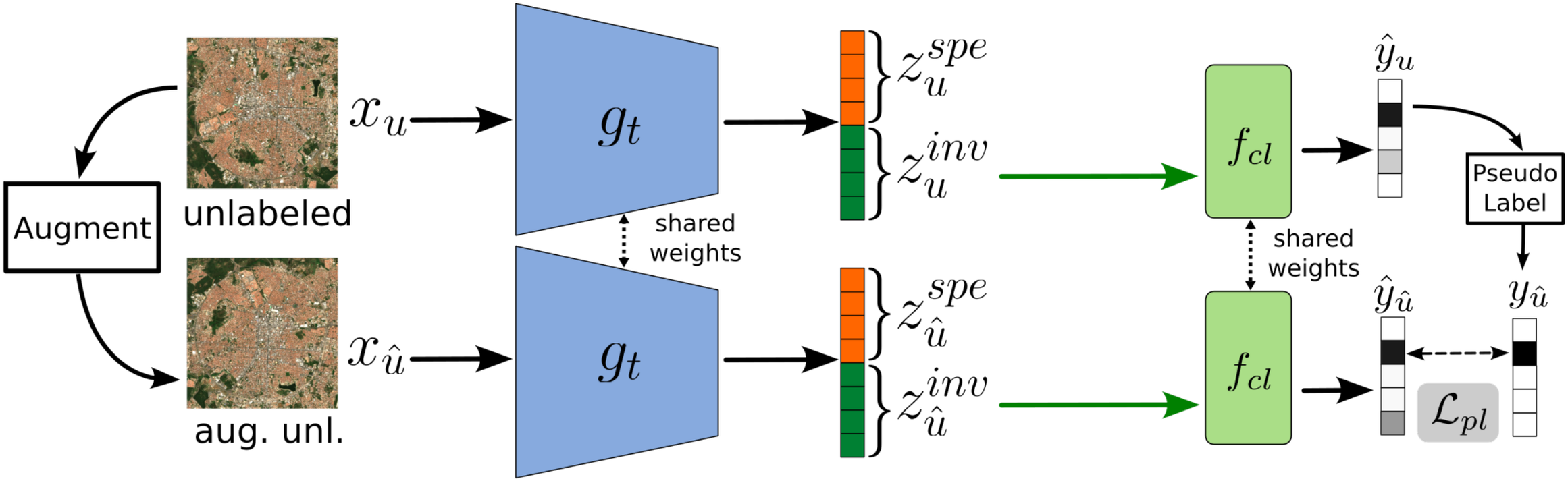
Leveraging unlabelled target data

Inspiration from semi-supervised learning literature [2]

- Confidence-based pseudo-labelling
- Enforce prediction consistency upon augmentations

$$m_u^\tau := \mathbb{1}(\max(f_{cl}(\mathbf{z}_u^{inv})) > \tau)$$

$$\mathcal{L}_{\text{pl}} = m_u^\tau \, \text{CE}(f(\mathbf{z}_{\hat{u}}^{inv}), y_{\hat{u}})$$



C. F. Dantas, R. Gaetano, and D. Ienco, “Semi-supervised heterogeneous domain adaptation via disentanglement and pseudo-labelling,” in Proc. ECML-PKDD, 2024, pp. 1–17.

[2] K. Sohn et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." *NeurIPS,* 2020

Data Augmentation



x_u

Horizontal / vertical flip

Random rotation
$\{0°, 90°, 180°, 270°\}$

Color jitter
(brightness, contrast, saturation, hue)

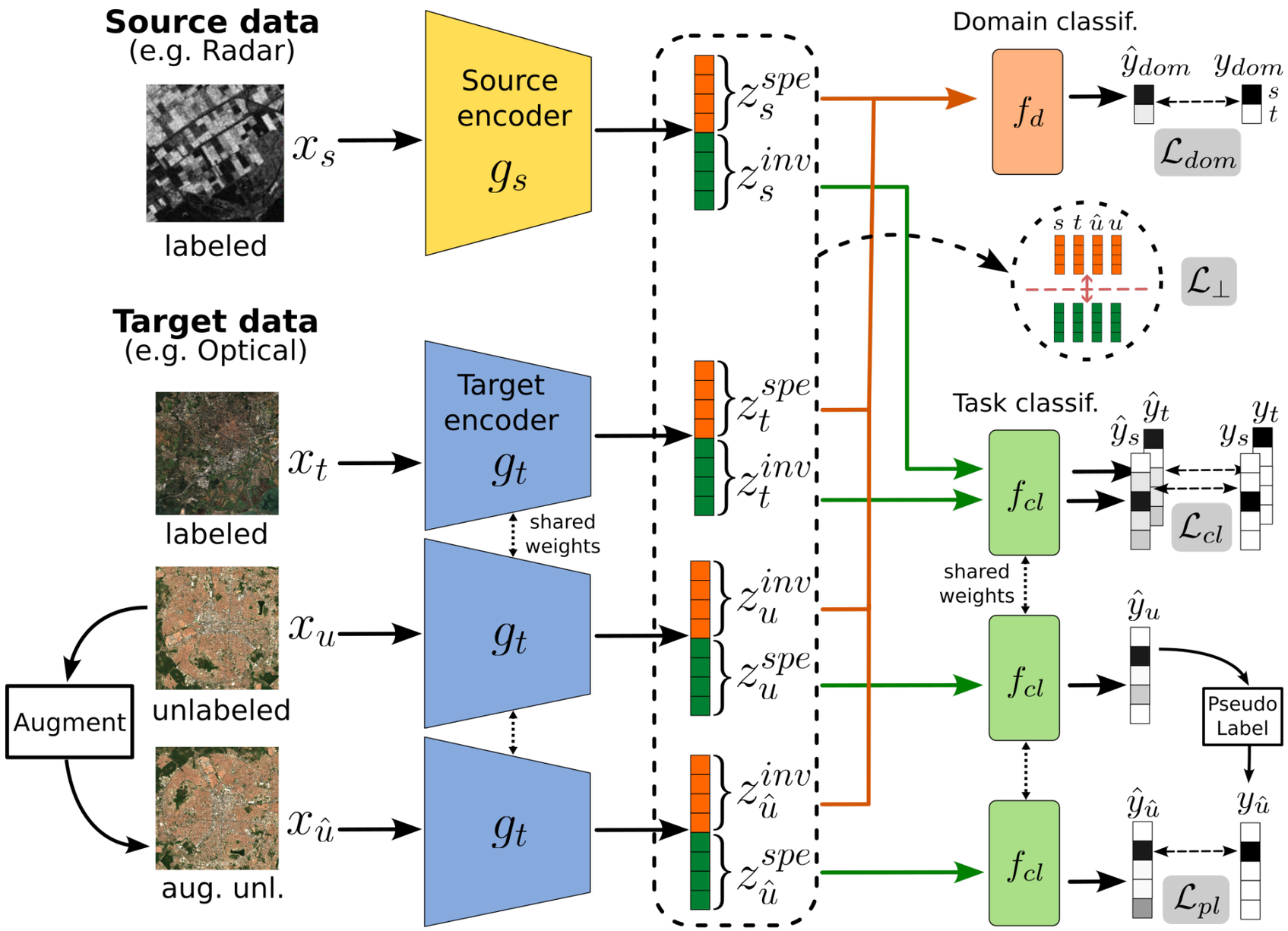
$x_{\hat{u}}$








Training procedure - Summary



C. F. Dantas, R. Gaetano, and D. Ienco, “Semi-supervised heterogeneous domain adaptation via disentanglement and pseudo-labelling,” in Proc. ECML-PKDD, 2024, pp. 1–17.

Quantitative results

F1-scores for EuroSAT MS-SAR dataset. MS (source) → SAR (target)

labelled target samples per class

Method	**25**	**50**	**100**	**200**
Target Only	60.08 ± 1.25	62.52 ± 0.38	64.93 ± 0.31	67.80 ± 0.65
FixMatch	59.07 ± 1.29	64.45 ± 0.35	67.26 ± 0.95	70.38 ± 0.59
SS-HIDA	60.24 ± 1.65	62.96 ± 0.86	66.63 ± 1.00	70.40 ± 0.87
SHeDD	**63.66** ± 1.53	**67.91** ± 1.83	**70.64** ± 1.50	**73.97** ± 0.67

Competitors

- Target Only: fully supervised baseline (exploits **labelled target** data)
- FixMatch [2]: semi-supervised approach (exploits **labelled and unlabelled target** data)
- SS-HIDA [3]: semi-supervised + heterogeneous (exploits **all source and target** data)

2. K. Sohn et al. "Fixmatch: Simplifying semi-supervised learning with consistency and confidence." *NeurIPS,* 2020
3. M. Obrenovic et al. “Learning domain invariant representations of heterogeneous image data.” Mach. Learn., 2023

What if we consider fewer labeled data?

More critical setting

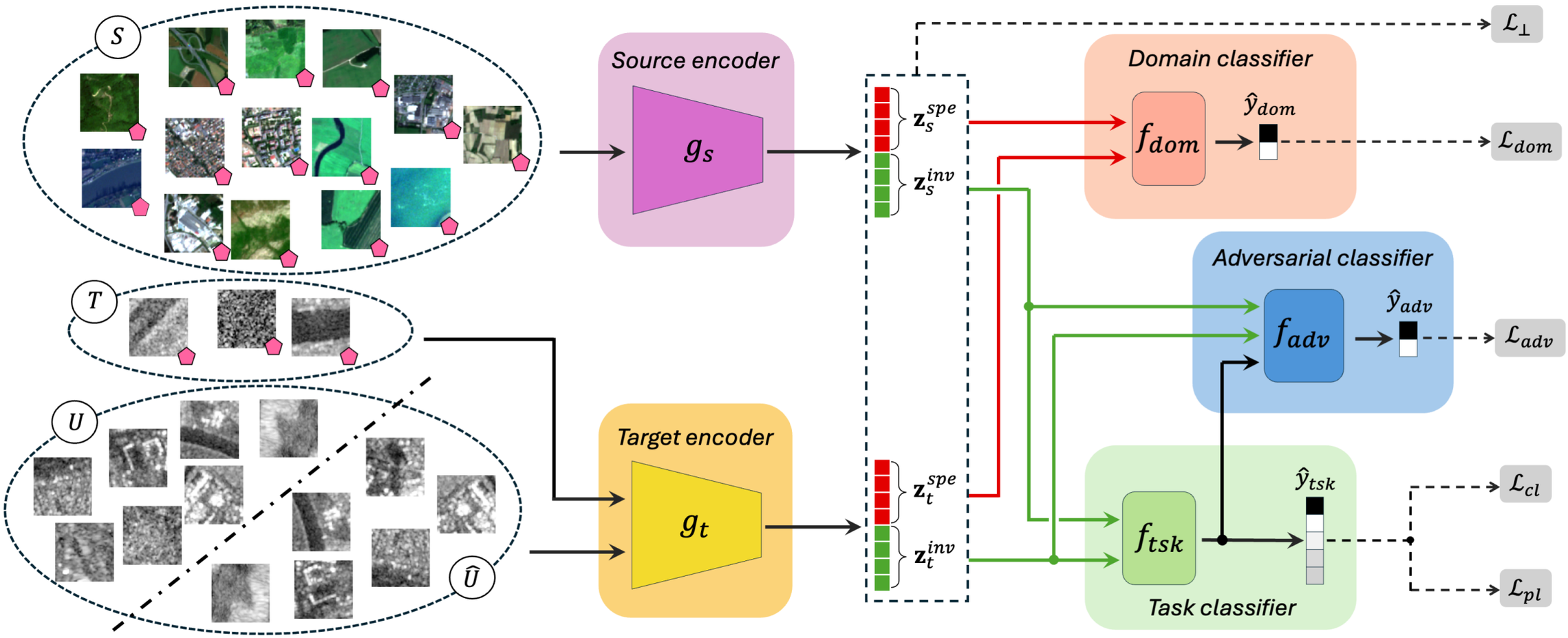
Method	5	10	25	50
Target Only	46.65 ± 3.39	52.71 ± 0.35	60.15 ± 0.79	62.66 ± 0.90
FixMatch [11]	47.00 ± 3.17	53.10 ± 2.07	60.01 ± 1.45	63.53 ± 0.92
FlatMatch [21]	49.01 ± 2.04	51.88 ± 2.46	58.85 ± 2.74	61.74 ± 0.83
SoftMatch [22]	47.16 ± 3.37	54.79 ± 3.31	62.09 ± 0.77	64.70 ± 0.51
SS-HIDA [15]	51.26 ± 3.99	56.15 ± 2.84	60.96 ± 0.81	63.20 ± 0.90
SHeDD [16]	44.57 ± 4.63	55.01 ± 3.14	64.74 ± 1.25	67.26 ± 1.18

Two limitations of SHeDD addressed by SAHARA:

- **No explicit adversarial alignment on $\mathbf{z}^{inv}$**: SHeDD's domain classifier acts only on $\mathbf{z}^{spe}$, but there is no mechanism to actively push $\mathbf{z}^{inv}$ to be domain-invariant
- **Fixed global pseudo-label threshold:** $\tau = 0.95$ for all classes ignores that some classes are harder (lower confidence) and some are easier (higher confidence)

SAHARA framework

Semi-supervised Adaptation in Heterogeneous domains via conditional Adversarial Representation disentanglement and Adaptive pseudo-labeling (SAHARA)



Guarino, G., Dantas, C. F., Ienco, D., Gaetano, R., Vivone, G., Ciotola, M., & Scarpa, G. (2025). SAHARA: Heterogeneous Semi-Supervised Transfer Learning With Adversarial Adaptation and Dynamic Pseudo-Labeling. *IEEE Geoscience and Remote Sensing Letters*, *23*, 1-5.

SAHARA: Key Innovations over SHeDD

1. Conditional Adversarial Adaptation

- SHeDD uses a simple domain classifier on $\mathbf{z}^{spe}$ (no adversarial on $\mathbf{z}^{inv}$)
- SAHARA adds a **CDAN-style adversarial classifier** on $\mathbf{z}^{inv}$
- Conditions on outer product: features ⊗ softmax predictions
- GRL enforces domain invariance **explicitly**. The encoder is actively pushed to produce domain-indistinguishable invariant features
- 5 losses: $L_{cl} + L_{dom} + L_{adv} + L_{\perp} + L_{pl}$

2. Dynamic Classwise Pseudo-Labeling

- SHeDD: fixed global threshold $\tau = 0.95$
- SAHARA: **per-class adaptive thresholds** $\lambda_c(t)$
- Updated via EMA of max prediction confidence per class:
  $\lambda_c(t) = \alpha\lambda_c(t-1) + (1-\alpha)\max_i p_c(x_i)$
- Saturation limit $\lambda_M = 0.95$
- **Effect:** easy classes get high thresholds (strict filtering), hard classes get lower thresholds (more pseudo-labels)
- Avoids **confirmation bias** under extreme label scarcity
- Critical for as few as **5 labels/class**

SAHARA: Results

Method	5	10	25	50
Target Only	55.51 ± 3.08	61.44 ± 3.07	75.27 ± 1.95	82.11 ± 0.72
FixMatch [11]	58.26 ± 4.15	65.06 ± 1.97	75.88 ± 1.07	82.51 ± 0.81
FlatMatch [21]	47.40 ± 5.09	55.63 ± 4.96	74.93 ± 0.70	82.05 ± 0.48
SoftMatch [22]	56.93 ± 5.23	65.31 ± 1.81	75.48 ± 1.46	83.07 ± 1.02
SS-HIDA [15]	58.66 ± 3.04	66.92 ± 1.86	78.40 ± 1.01	85.38 ± 0.40
SHeDD [16]	64.63 ± 3.77	72.77 ± 2.95	82.71 ± 1.45	87.65 ± 0.73
SAHARA	**69.96** ± 4.50	**75.62** ± 3.48	**84.22** ± 0.87	**88.64** ± 0.34

Backbone comparison:

- **ResNet-18** (11.2M params): standard CNN backbone
- **TinyViT** (5.4M params): compact vision transformer
- TinyViT achieves comparable or better results with half the parameters
- Transformers benefit from the attention mechanism which can capture global relationships in the image
- Both backbones are trained from scratch (no ImageNet pretraining), important for non-RGB modalities
- **Key insight:** the SAHARA framework is **backbone-agnostic**. It works with both CNNs and transformers.

SHeDD vs. SAHARA: Summary

Component	SHeDD (ECML 2024)	SAHARA (IEEE GRSL 2026)
Disentanglement	$\mathbf{z}^{inv}$ / $\mathbf{z}^{spe}$ + orthogonality	Same
Domain classifier	CE on $\mathbf{z}^{spe}$	CE on $\mathbf{z}^{spe}$
Adversarial adaptation	—	CDAN + GRL on $\mathbf{z}^{inv}$
Pseudo-label threshold	Fixed global ($\tau = 0.95$)	Dynamic per-class $\lambda_c(t)$
Min labels/class tested	25	**5**
Backbone	ResNet-18	ResNet-18 + TinyViT
# losses	4	5

Evolution

SHeDD introduced the **disentanglement framework** for semi-supervised heterogeneous DA. SAHARA refined it with **adversarial alignment** on invariant features and **adaptive pseudo-labeling**, achieving robustness under extreme label scarcity (down to 5 labels/class).

Y. Li, Z. Li, and Y. Jia, “Complementary label learning with positive label guessing and negative label enhancement,” in Proc. ICLR, 2025, pp. 1–10.

What Varies Across Papers

	Spatial	Temporal	Sensor	Supervision	Key technique
TimeMatch	✓			UDA	Temporal shift est. + self-training
SpADANN		✓		UDA	Adversarial + spatial pseudo-labels
M3SPADA		✓	✓	UDA	Multi-sensor fusion + spatial PL
ConDA-Vine		✓	(✓)	UDA	Conditional adversarial (CDAN)
SHeDD	✓	✓	✓	Semi-sup	Disentanglement + FixMatch PL
SAHARA	✓	✓	✓	Semi-sup	Adv. disentanglement + dynamic PL

Common threads across all papers:

- **Pseudo-labeling** appears in every method (but with different strategies: confidence-based, spatial, dynamic)
- **Adversarial learning** (DANN/CDAN) is the dominant feature alignment strategy
- **Domain-specific batch normalization** is universally adopted when processing different domains
- All methods designed for **operational** RS scenarios with sparse annotations

Future Directions: Foundation Models

Foundation Models for EO

- Large pretrained models trained on massive EO data via self-supervised learning
- Examples: Prithvi (NASA/IBM), SatMAE, CROMA, SSL4EO
- Could **replace hand-crafted DA**: a well-pretrained model may generalize across domains with minimal fine-tuning
- Challenge: adaptation for new sensors/tasks not seen during pretraining
- Opportunity: foundation models as **universal feature extractors** that reduce domain shift by providing robust representations

Multi-Modal Foundation Models

- Joint optical + SAR + text pretraining
- Could **natively handle heterogeneous DA**: the model already understands multiple modalities
- Vision-language models (e.g., RemoteCLIP) for zero-shot classification
- Current limitation: most EO foundation models are optical-only; SAR and multi-modal variants are emerging

Future Directions: Source-Free & Few-Shot DA

Source-Free Domain Adaptation

- Adapt **without access to source data** at adaptation time
- Only the pretrained source model is available (no raw source images)
- **Motivation:** data privacy, storage limitations, data sharing restrictions between organizations
- Techniques: test-time training, batch norm adaptation, entropy minimization on target
- Active research area in CV; emerging in RS
- Challenge: without source data, harder to detect negative transfer

Few-Shot Domain Adaptation

- Bridge between UDA (0 target labels) and full supervision
- SAHARA already pushes to **5 labels/class**, extreme low-label regime
- Meta-learning approaches: learn to adapt quickly from few examples
- Prototypical networks: class representations from just a few labeled target samples
- Promising for RS: it's always possible to collect **a handful** of ground truth samples

Summary

- **Remote sensing** provides massive EO data, **but labels are scarce and localized**. Models trained on one domain fail on another
- **Domain shift** (spatial, temporal, sensor, modality) prevents direct model transfer across regions, years, and sensors
- **UDA methods** bridge this gap:
  - **TimeMatch**: temporal shift estimation + self-training for spatial transfer
  - **SpADANN**: adversarial + spatial pseudo-labels for temporal transfer
  - **M3SPADA**: multi-sensor fusion for spatio-temporal transfer
  - **ConDA-Vine**: conditional adversarial adaptation for multi-source vineyard monitoring
- **Heterogeneous DA** (SHeDD, SAHARA) tackles the hardest problem: cross-modality transfer (optical ↔SAR) with feature disentanglement
- **Key ingredients:** adversarial alignment, pseudo-labeling (various strategies), disentanglement, domain-specific BN
- **Future:** foundation models, source-free DA, few-shot DA, multi-modal pretraining

Key Takeaways

Conceptual takeaways:

- Domain shift is **unavoidable** in real-world RS applications
- DA is not just an academic exercise, it has direct **operational impact** (crop monitoring, vineyard health, land cover mapping)
- The progression from homogeneous UDA to heterogeneous SSDA reflects increasing real-world complexity
- Understanding **sensor physics** (why optical ≠ SAR) is essential to understand why DA is needed

Technical takeaways:

- **Adversarial learning** (DANN, CDAN) is the workhorse of feature alignment
- **Pseudo-labeling** is ubiquitous but must be handled carefully (confirmation bias, class imbalance)
- **Feature disentanglement** ($\mathbf{z}^{inv}$ / $\mathbf{z}^{spe}$) is powerful for heterogeneous DA
- **Domain-specific BN** is a simple but effective technique adopted everywhere
- Always compare against **source-only** and **target-trained** baselines to understand the DA gap

References

1. J. Nyborg, C. Pelletier, S. Lefèvre, I. Assent, “TimeMatch: Unsupervised Cross-Region Adaptation by Temporal Shift Estimation,” ISPRS J. Photogrammetry & Remote Sensing, vol. 188, pp. 301–313, 2022.
2. E. Capliez, D. Ienco, R. Gaetano, N. Baghdadi, A. Hadj Salah, “Temporal-Domain Adaptation for SITS Land-Cover Mapping With Adversarial Learning and Spatially Aware Self-Training,” IEEE JSTARS, vol. 16, 2023.
3. E. Capliez, D. Ienco, R. Gaetano, N. Baghdadi, A. Hadj Salah, M. Le Goff, F. Chouteau, “Multisensor Temporal UDA for Land Cover Mapping With Spatial Pseudo-Labeling and Adversarial Learning,” IEEE TGRS, vol. 61, 2023.
4. A. Roussel, S. Alleaume, A. Billotte, F. Fabre, D. Ienco, “Out-of-year parcel identification using adversarial DA and multi-temporal satellite imagery to enhance vineyard health monitoring,” Int. J. Digital Earth, vol. 19, 2026.
5. C. F. Dantas, R. Gaetano, D. Ienco, “Semi-Supervised Heterogeneous Domain Adaptation via Disentanglement and Pseudo-Labelling,” ECML-PKDD, 2024.
6. G. Guarino, C. F. Dantas, D. Ienco, R. Gaetano, G. Vivone, M. Ciotola, G. Scarpa, “SAHARA: Heterogeneous Semi-Supervised Transfer Learning With Adversarial Adaptation and Dynamic Pseudo-Labeling,” IEEE GRSL, vol. 23, 2026.
7. Z. Fang, J. Lu, F. Liu, G. Zhang, “Semi-Supervised Heterogeneous Domain Adaptation: Theory and Algorithms,” IEEE TPAMI, vol. 45, no. 1, 2023.
8. Y. Ganin et al., “Domain-Adversarial Training of Neural Networks,” JMLR, vol. 17, 2016. M. Long, Z. Cao, J. Wang, M. I. Jordan, “Conditional Adversarial Domain Adaptation,” NeurIPS, 2018.

Thank you