



SAHARA: Heterogeneous Semi-Supervised Transfer Learning with Adversarial Adaptation and Dynamic Pseudo-Labeling

Giuseppe Guarino, Cássio F Dantas, Dino Ienco, Raffaele Gaetano, Gemine Vivone, Matteo Ciotola, Giuseppe Scarpa

► To cite this version:

Giuseppe Guarino, Cássio F Dantas, Dino Ienco, Raffaele Gaetano, Gemine Vivone, et al.. SAHARA: Heterogeneous Semi-Supervised Transfer Learning with Adversarial Adaptation and Dynamic Pseudo-Labeling. IEEE Geoscience and Remote Sensing Letters, 2025, 23, 13 p. <10.1109/LGRS.2025.3635269>. <hal-05370697>

HAL Id: hal-05370697

<https://hal.science/hal-05370697v1>

Submitted on 18 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

SAHARA: Heterogeneous Semi-Supervised Transfer Learning with Adversarial Adaptation and Dynamic Pseudo-Labeling

Giuseppe Guarino^{1,2} (✉), Cássio F. Dantas^{2,4}, Dino Ienco^{2,4},
Raffaele Gaetano^{3,4}, Gemine Vivone^{5,6}, Matteo Ciotola¹, and Giuseppe Scarpa⁷

¹ Department of Electrical Engineering and Information Technology, University Federico II, 80125 Napoli, Italy.

² INRAE, EVERGREEN, UMR TETIS, University of Montpellier, Montpellier, France {giuseppe.guarino}@inrae.fr.

³ CIRAD, INRIA, EVERGREEN, UMR TETIS, University of Montpellier, Montpellier, France

⁴ Inria, University of Montpellier, Montpellier, France.

⁵ National Research Council, Institute of Methodologies for Environmental Analysis, CNR-IMAA, 85050 Tito, Italy.

⁶ National Biodiversity Future Center (NBFC), 90133 Palermo, Italy.

⁷ Department of Engineering, University Parthenope, 80143 Napoli, Italy.

Abstract. Semi-supervised domain adaptation aims to transfer knowledge from a labeled source domain to a scarcely labeled target domain, despite distribution shifts. The challenge becomes greater when source and target data differ in acquisition modality, as in remote sensing where variations in sensor type (e.g., optical vs. radar), spectral properties (e.g., RGB vs. multispectral), or spatial resolution are common. This challenging scenario, known as Semi-Supervised Heterogeneous Domain Adaptation (SSHDA), requires learning across modalities with limited target labels. In this work, we propose SAHARA (Semi-supervised Adaptation in Heterogeneous domains via conditional Adversarial Representation disentanglement and Adaptive pseudo-labeling), a new method for SSHDA that combines conditional adversarial feature adaptation with dynamic pseudo-labeling to learn domain-invariant features and handle extremely scarce target annotations. Experiments on two heterogeneous remote sensing benchmarks for scene classification, conducted with both convolutional and transformer-based backbones, demonstrate that SAHARA consistently outperforms existing SSHDA and semi-supervised methods. The code is available at <https://github.com/giu-guarino/SAHARA>.

Keywords: Domain Adaptation, Heterogeneous data, Feature disentanglement, Pseudo-labeling, Consistency regularization, Conditional adversarial learning.

1 Introduction

The launch of new satellite missions is increasingly frequent, each bringing next-generation sensors providing higher-quality data. These advances create oppor-

tunities to train deep learning models for computer vision tasks, such as biophysical variable estimation [15, 20], semantic segmentation [9, 17], and image super-resolution [1, 6]. However, collecting and labeling data from new sensors remains time-consuming and costly. In many cases, data from the desired sensor may be unavailable, while other satellites with different acquisition modes (e.g., optical vs. radar) are accessible. This motivates using transfer learning to exploit labeled data from one data source with the aim to improve performances on the target data.

Differences in data distributions between a richly labeled source domain and a scarcely labeled target domain hinder knowledge transfer. Domain Adaptation (DA) addresses this by enabling models to generalize across domains despite distribution shifts. While much research has focused on Unsupervised Domain Adaptation (UDA), where no target labels are available [24], real-world cases often include a few labeled target samples. Semi-Supervised Domain Adaptation (SSDA) bridges this gap by combining abundant source labels with limited target labels [11, 26, 19]. Pseudo-labeling is often employed in transfer learning scenarios to exploit unlabeled target data, by assigning labels when model confidence exceeds a fixed threshold [21], or by adapting the confidence criterion on a per-class basis [12].

Yet, most DA methods assume shared modalities between source and target, which rarely holds in remote sensing. Data frequently differ in acquisition mode, spectral range, or spatial resolution [18]. This more complex setting, known as Semi-Supervised Heterogeneous Domain Adaptation (SSHDA) [4], introduces additional challenges due to modality heterogeneity. Recent works [16, 3] explored domain-specific encoders and feature disentanglement with pseudo-label regularization, but their exploration under extreme label scarcity remains limited.

Here, we propose SAHARA (Semi-supervised Adaptation in Heterogeneous domains via conditional Adversarial Representation disentanglement and Adaptive pseudo-labeling), a novel end-to-end framework for SSHDA. It disentangles domain-invariant and domain-specific representations via a conditional adversarial module that enforces invariance by conditioning domain discrimination on both feature embeddings and task predictions [13]. In addition, a class-wise dynamic pseudo-labeling mechanism adapts confidence thresholds per class, improving the exploitation of unlabeled target data. Extensive experiments on two multimodal remote sensing benchmarks show that SAHARA achieves superior accuracy and robustness compared to existing SSHDA and semi-supervised approaches, particularly in low-label regimes.

2 Proposed Solution

The *SAHARA* framework (Fig. 1) includes two encoders: a source encoder (g_s) for the source domain and a target encoder (g_t) for the target domain. Each extracts a data representation $\mathbf{z}$, split into domain-invariant ($\mathbf{z}^{inv}$) and domain-specific ($\mathbf{z}^{spe}$) features. These are fed to a task classifier (f_{tsk}) and a domain

classifier (f_{dom}), while an additional module (f_{adv}) promotes invariant feature extraction. At inference, only the target encoder and task classifier are used.

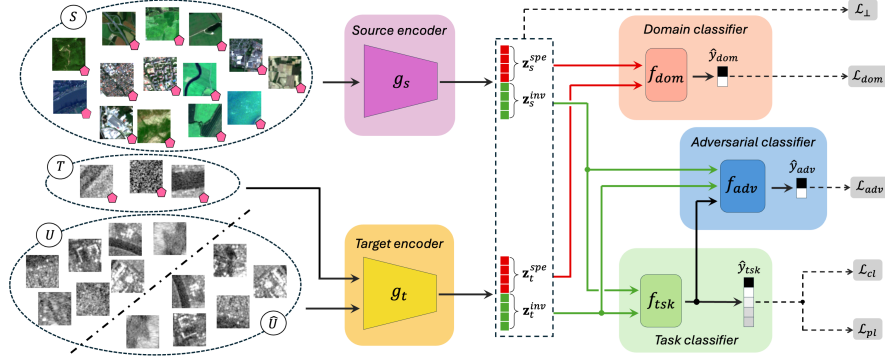


Fig. 1. A schematic overview of *SAHARA*. The source (g_s) and target (g_t) encoders produce features split into domain-specific ($\mathbf{z}_s^{spe}, \mathbf{z}_t^{spe}$) and domain-invariant ($\mathbf{z}_s^{inv}, \mathbf{z}_t^{inv}$) components. The domain classifier (f_{dom}) models domain-specific features, while the task classifier (f_{tsk}) predicts class labels from $\mathbf{z}^{inv}$. A conditional adversarial domain classifier (f_{adv}) enforces feature invariance disentanglement. Training is guided by the losses $\mathcal{L}_\perp$, $\mathcal{L}_{dom}$, $\mathcal{L}_{adv}$, $\mathcal{L}_{cl}$, and $\mathcal{L}_{pl}$ (Eq. 8, 2, 11, 13). At inference, only g_t and f_{tsk} are used for target classification. Labeled samples, belonging to the source (S) and target (T) sets, are marked with $\blacklozenge$, whereas unlabeled data and their augmentations, from sets U and $\hat{U}$, are left unmarked.

The training losses rely on the following sets:

$$\begin{aligned}
 S &:= \{(\mathbf{x}_s, \mathbf{y}_s)^{(i)}\}_{i=1}^{N_s} && \text{labeled source data.} \\
 T &:= \{(\mathbf{x}_t, \mathbf{y}_t)^{(i)}\}_{i=1}^{N_t} && \text{labeled target data.} \\
 U &:= \{(\mathbf{x}_u)^{(i)}\}_{i=1}^{N_u} && \text{unlabeled target data.} \\
 \hat{U} &:= \{(\mathbf{x}_{\hat{u}})^{(i)}\}_{i=1}^{N_{\hat{u}}} && \text{augmented unlabeled target data.}
 \end{aligned}$$

where N_s , N_t , and N_u represent the respective dataset sizes. The set $\hat{U}$ contains the augmented versions of the unlabeled target samples $\mathbf{x}_u$, produced by applying a data augmentation function (details provided in the next sections).

For each labeled sample, whether from the source or the target domain, the output of the task classifier is compared to the ground-truth label $\mathbf{y}$ via the standard cross-entropy loss:

$$\mathcal{L}_{cl} = \text{CE}(f_{tsk}(\mathbf{z}^{inv}), \mathbf{y}) \quad (1)$$

Since the domain label $y_{dom} \in \{s, t\}$ is known for every sample (including unlabeled target data), we also adopt a cross-entropy loss on the domain classifier's

output:

$$\mathcal{L}_{dom} = \text{CE}(f_{dom}(\mathbf{z}^{spe}), y_{dom}) \quad (2)$$

To fully exploit the unlabeled target data, we introduce an advanced dynamic pseudo-labeling mechanism inspired by FixMatch [21]. For each unlabeled target sample $\mathbf{x}_u \in U$, we generate an augmented version $\mathbf{x}_{\hat{u}} = \text{Augment}(\mathbf{x}_u)$. The model is encouraged to produce consistent predictions between $\mathbf{x}_u$ and $\mathbf{x}_{\hat{u}}$ by minimizing the following unsupervised loss:

$$\mathcal{L}_{pl} = m_u^\lambda \text{CE}(f_{tsk}(\mathbf{z}_{\hat{u}}^{inv}), y_{\hat{u}}) \quad (3)$$

where $y_{\hat{u}} := \arg \max \mathbf{p}(\mathbf{z}_u^{inv})$ is a pseudo-label derived from the task classifier prediction $\mathbf{p}(\mathbf{z}_u^{inv}) = \text{softmax}(f_{tsk}(\mathbf{z}_u^{inv}))$ on the original feature $\mathbf{z}_u^{inv}$ and m_u^λ is a binary mask detailed further below.

In the standard FixMatch formulation, a fixed global threshold is used to filter out low-confidence pseudo-labels. However, uniform thresholding fails to account for the varying difficulty of individual instances, leaving hard examples unsupervised for longer and slowing down learning. It also overlooks that the model may learn different classes at different rates, a key issue in real-world imbalanced class settings. To overcome these issues, inspired by [12], we introduce a class-wise, self-adaptive thresholding strategy that dynamically adjusts the confidence threshold for each class as training progresses.

Specifically, given C classes, the threshold $\lambda_c(t)$ for the c -th class is updated as:

$$\lambda_c(t) = \alpha \lambda_c(t-1) + (1-\alpha) f_c(t), \quad \lambda_c(0) = \frac{1}{C} \quad (4)$$

Given the set of all samples predicted as class c at iteration t defined as $\mathcal{S}_c = \{\mathbf{z}_u^{inv} | y_{\hat{u}} = c\}$, we denote with

$$f_c(t) = \max_{\mathcal{S}_c} \mathbf{p}(\mathbf{z}_u^{inv})[c] \quad (5)$$

the maximum prediction confidence obtained for the set $\mathcal{S}_c$. The scalar α controls the trade-off between stability and adaptability of the threshold over time. The binary mask m_u^λ is then defined as:

$$m_u^\lambda := \mathbf{1}(\max \mathbf{p}(\mathbf{z}_u^{inv}) > \lambda_{y_{\hat{u}}}(t)) \quad (6)$$

where $\mathbf{1}(\cdot)$ is the indicator function, activating the loss only for predictions exceeding the corresponding class-specific threshold $\lambda_{y_{\hat{u}}}(t)$ for the class given by the pseudo-label $y_{\hat{u}}$.

To avoid $\lambda_c(t)$ from growing excessively and becoming overly restrictive, we introduce a saturation limit λ_M :

$$\lambda_c(t) = \min(\lambda_c(t), \lambda_M) \quad (7)$$

As a first step in separating domain-invariant and domain-specific features, we introduce a loss that enforces orthogonality between these feature spaces (for both domains):

$$\mathcal{L}_\perp = \frac{\langle \mathbf{z}^{inv}, \mathbf{z}^{spe} \rangle}{\|\mathbf{z}^{inv}\|_2 \|\mathbf{z}^{spe}\|_2} \quad (8)$$

However, this loss alone does not explicitly enforce the features to be domain-invariant or domain-specific; it only encourages orthogonality between them. To address this, we introduce a conditional adversarial domain classifier, inspired by the Conditional Domain Adversarial Network (CDAN) framework [13], designed to explicitly promote domain invariance in the learned features. Unlike standard adversarial domain classifiers that rely solely on feature representations to discriminate between source and target domains, CDAN conditions the domain discriminator on both the features and the task predictions, thereby capturing the cross-covariance between them. Formally, let the task classifier outputs and domain-invariant features for source and target domains be given. The conditioned representation $\mathbf{h}$ is then computed as the outer product between the softmax predictions $\mathbf{p}(\mathbf{z}^{inv}) = \text{softmax}(f_{tsk}(\mathbf{z}^{inv}))$ and the invariant features:

$$\mathbf{h} = \text{vec}(\mathbf{p}(\mathbf{z}^{inv}) \otimes \mathbf{z}^{inv}) \quad (9)$$

where $\otimes$ denotes the outer product, and $\text{vec}(\cdot)$ flattens the resulting matrix into a vector. We then apply a gradient reversal layer (GRL) [5] to the resulting conditioned representations, from both source and target domains, before feeding them to the conditional adversarial domain classifier f_{adv} :

$$\hat{y}_{adv} = f_{adv}(\text{GRL}(\mathbf{h})) \quad (10)$$

The adversarial loss is finally defined as:

$$\mathcal{L}_{adv} = \text{CE}(\hat{y}_{adv}, y_{dom}) \quad (11)$$

Finally, the overall loss function is:

$$\mathcal{L} = \mathcal{L}_{cl} + \mathcal{L}_{dom} + \mathcal{L}_{pl} + \mathcal{L}_{\perp} + \mathcal{L}_{adv} \quad (12)$$

3 Experiments

Dataset The experiments were conducted on two multimodal remote sensing datasets, summarized in TABLE 1: i) EuroSat-MS-SAR [23], which includes 27,000 pairs of multispectral (MS) and synthetic aperture radar (SAR) images. The MS images consist of 13 spectral channels, while the SAR images contain 2 channels. The heterogeneity in this dataset stems from the different acquisition modalities (optical and radar). To reduce potential data biases and spurious correlations, only one modality was retained per sample. Consequently, the benchmark comprises 13,500 MS images and 13,500 SAR images, which are unaligned and covers 10 land cover classes. ii) RESISC45-Euro [16], contains 5,600 RGB images at varying spatial resolutions and 24,000 MS images with 13 spectral channels, spanning 8 land cover classes. In this case, heterogeneity arises from differences in both spatial and spectral resolutions between the domains.

Comparative methods To assess the effectiveness of our framework, we compare it against recent methods specifically designed for SSHDA: SS-HIDA [16] and SHeDD [3], which represent the state-of-the-art in this setting. In addition,

Dataset	Volume	Modality	Spatial Res.	Classes
<i>EuroSAT-MS-SAR</i>	$13\,500 \times 13 \times 64 \times 64$	MS	10m	10
	$13\,500 \times 2 \times 64 \times 64$	SAR	10m	
<i>RESISC45-Euro</i>	$5\,600 \times 3 \times 256 \times 256$	RGB	0.2m - 30m	8
	$24\,000 \times 13 \times 64 \times 64$	MS	10m	

Table 1. Multimodal remote sensing datasets used in this study

given the strong performance of semi-supervised learning techniques in SSDA scenarios [22], we evaluate three state-of-the-art semi-supervised methods: the widely adopted FixMatch [21], and two more recent approaches, FlatMatch [8] which introduces balanced augmentations to reduce over-confident predictions, and SoftMatch [2], that employs soft thresholding to select reliable pseudo-labels and stabilize training. Finally, as a baseline, we consider a fully supervised model, referred as Target Only, which is trained exclusively on the available labeled samples from the target domain.

Experimental setting For each benchmark, we design two transfer tasks (one for each modality) denoted as $\mathcal{D}_s \rightarrow \mathcal{D}_t$, where the source domain $\mathcal{D}_s$ is fully labeled and the target domain $\mathcal{D}_t$ has limited labeled data. We explore different levels of supervision by varying the number of labeled target samples per class in $\{5, 10, 25, 50\}$. All experiments are conducted using two backbone architectures: ResNet-18 [7] and TinyViT 5M version) [25]. The remaining target samples serve both as additional unlabeled data during training and as the test set for evaluation. Model performance is measured on the test set using the weighted F1-score (referred to simply as F1-score). To ensure robustness, each experiment is run five times, reporting the mean and standard deviation.

Implementation details In both SHeDD and *SAHARA*, the task and domain classifiers are implemented as a fully connected layer followed by softmax activation. For SS-HIDA, following its original design, the backbone is split into a domain-specific block (first half of layers) and a shared block (remaining layers). For FixMatch and SHeDD, we set the pseudo-label confidence threshold τ to 0.95. The augmentation function applies random combinations of geometric transformations (flipping and rotation). In *SAHARA*, we use $\alpha = 0.9$, $\lambda_M = 0.95$ and an adversarial classifier composed of three fully connected layers with ReLU activations in the first two. All models are trained for 300 epochs with a batch size of 128 using the AdamW optimizer [14] (learning rate 10^{-4}). To improve stability and performance, we apply exponential moving average (EMA) updates to the weights with a momentum of 0.95, as done in recent work [10]; this is applied uniformly across methods. All the methods are implemented in PyTorch and run on an NVIDIA RTX A4500 GPU (20GB VRAM).

Results discussion Tables 2 and 3 summarize the F1-scores achieved by all methods across varying levels of target supervision (5, 10, 25, 50 labeled samples) and both backbone architectures (ResNet-18 and TinyViT). In the tables, the best performance is shown in bold, and the second-best is underlined. Taking Target Only as a baseline, *SAHARA* consistently demonstrates its ability to

MS $\rightarrow$ SAR					
	Method	5	10	25	50
ResNet-18	Target Only	46.65 $\pm$ 3.39	52.71 $\pm$ 0.35	60.15 $\pm$ 0.79	62.66 $\pm$ 0.90
	FixMatch [21]	47.00 $\pm$ 3.17	53.10 $\pm$ 2.07	60.01 $\pm$ 1.45	63.53 $\pm$ 0.92
	FlatMatch [8]	49.01 $\pm$ 2.04	51.88 $\pm$ 2.46	58.85 $\pm$ 2.74	61.74 $\pm$ 0.83
	SoftMatch [2]	47.16 $\pm$ 3.37	54.79 $\pm$ 3.31	62.09 $\pm$ 0.77	64.70 $\pm$ 0.51
	SS-HIDA [16]	<u>51.26</u> $\pm$ 3.99	<u>56.15</u> $\pm$ 2.84	60.96 $\pm$ 0.81	63.20 $\pm$ 0.90
	SHeDD [3]	44.57 $\pm$ 4.63	55.01 $\pm$ 3.14	<u>64.74</u> $\pm$ 1.25	<u>67.26</u> $\pm$ 1.18
	<i>SAHARA</i>	56.62 $\pm$ 4.36	61.48 $\pm$ 1.95	67.02 $\pm$ 1.28	69.47 $\pm$ 0.98
TinyViT	Target Only	36.70 $\pm$ 3.07	39.39 $\pm$ 2.14	49.90 $\pm$ 1.27	54.70 $\pm$ 1.43
	FixMatch [21]	38.29 $\pm$ 1.64	43.00 $\pm$ 1.89	49.98 $\pm$ 2.09	55.94 $\pm$ 1.63
	FlatMatch [8]	34.74 $\pm$ 3.58	39.19 $\pm$ 2.61	48.85 $\pm$ 2.05	55.44 $\pm$ 1.17
	SoftMatch [2]	38.28 $\pm$ 2.65	44.49 $\pm$ 1.34	52.92 $\pm$ 0.84	58.37 $\pm$ 2.26
	SS-HIDA [16]	<u>44.78</u> $\pm$ 2.71	<u>50.39</u> $\pm$ 1.30	<u>53.10</u> $\pm$ 0.75	56.65 $\pm$ 1.04
	SHeDD [3]	20.30 $\pm$ 11.61	33.21 $\pm$ 10.08	45.33 $\pm$ 7.73	60.25 $\pm$ 3.53
	<i>SAHARA</i>	48.28 $\pm$ 3.31	54.54 $\pm$ 2.07	62.48 $\pm$ 1.73	66.39 $\pm$ 1.33
SAR $\rightarrow$ MS					
	Method	5	10	25	50
ResNet-18	Target Only	55.81 $\pm$ 3.37	65.50 $\pm$ 1.53	76.54 $\pm$ 1.46	83.05 $\pm$ 0.14
	FixMatch [21]	57.95 $\pm$ 1.71	66.06 $\pm$ 2.21	77.11 $\pm$ 1.45	84.12 $\pm$ 1.19
	FlatMatch [8]	55.70 $\pm$ 2.51	65.48 $\pm$ 0.61	76.44 $\pm$ 1.11	82.45 $\pm$ 0.46
	SoftMatch [2]	57.59 $\pm$ 3.30	66.32 $\pm$ 2.00	76.97 $\pm$ 0.67	83.63 $\pm$ 1.16
	SS-HIDA [16]	<u>64.81</u> $\pm$ 2.51	<u>71.75</u> $\pm$ 1.36	81.32 $\pm$ 1.19	86.79 $\pm$ 2.04
	SHeDD [3]	56.53 $\pm$ 4.28	70.21 $\pm$ 1.19	<u>82.95</u> $\pm$ 1.29	<u>88.81</u> $\pm$ 1.14
	<i>SAHARA</i>	64.95 $\pm$ 1.51	74.23 $\pm$ 1.08	83.99 $\pm$ 1.79	89.82 $\pm$ 0.64
TinyViT	Target Only	50.63 $\pm$ 3.15	58.86 $\pm$ 1.38	69.88 $\pm$ 1.50	76.86 $\pm$ 0.77
	FixMatch [21]	53.99 $\pm$ 0.73	60.53 $\pm$ 0.59	68.76 $\pm$ 2.12	77.65 $\pm$ 1.25
	FlatMatch [8]	51.83 $\pm$ 3.12	58.19 $\pm$ 2.26	68.14 $\pm$ 1.44	76.44 $\pm$ 1.18
	SoftMatch [2]	54.23 $\pm$ 3.46	61.77 $\pm$ 2.38	70.30 $\pm$ 1.27	78.34 $\pm$ 0.78
	SS-HIDA [16]	34.61 $\pm$ 2.31	52.66 $\pm$ 1.79	63.39 $\pm$ 0.66	73.01 $\pm$ 0.61
	SHeDD [3]	<u>59.86</u> $\pm$ 3.62	<u>66.80</u> $\pm$ 2.21	<u>77.45</u> $\pm$ 2.70	<u>86.85</u> $\pm$ 0.92
	<i>SAHARA</i>	63.64 $\pm$ 2.38	70.84 $\pm$ 2.76	81.22 $\pm$ 2.84	87.16 $\pm$ 0.77

Table 2. F1-score mean and standard deviation (5 runs) on EuroSAT MS-SAR, varying per class target labeled samples in {5, 10, 25, 50}. Best scores are bolded, second-best underlined.

EURO → RESISC45					
	Method	5	10	25	50
ResNet-18	Target Only	55.51 ± 3.08	61.44 ± 3.07	75.27 ± 1.95	82.11 ± 0.72
	FixMatch [21]	58.26 ± 4.15	65.06 ± 1.97	75.88 ± 1.07	82.51 ± 0.81
	FlatMatch [8]	47.40 ± 5.09	55.63 ± 4.96	74.93 ± 0.70	82.05 ± 0.48
	SoftMatch [2]	56.93 ± 5.23	65.31 ± 1.81	75.48 ± 1.46	83.07 ± 1.02
	SS-HIDA [16]	58.66 ± 3.04	66.92 ± 1.86	78.40 ± 1.01	85.38 ± 0.40
	SHeDD [3]	<u>64.63</u> ± 3.77	<u>72.77</u> ± 2.95	<u>82.71</u> ± 1.45	<u>87.65</u> ± 0.73
	SAHARA	69.96 ± 4.50	75.62 ± 3.48	84.22 ± 0.87	88.64 ± 0.34
TinyViT	Target Only	37.77 ± 2.78	47.06 ± 1.50	66.22 ± 2.50	71.86 ± 1.73
	FixMatch [21]	47.06 ± 2.60	54.00 ± 2.94	67.29 ± 1.52	73.13 ± 0.59
	FlatMatch [8]	39.83 ± 4.05	46.69 ± 2.29	66.38 ± 2.48	70.99 ± 1.08
	SoftMatch [2]	46.28 ± 6.06	52.64 ± 1.12	67.50 ± 2.63	73.60 ± 1.13
	SS-HIDA [16]	36.93 ± 7.48	45.62 ± 2.19	56.37 ± 0.67	62.12 ± 3.53
	SHeDD [3]	<u>60.73</u> ± 2.84	<u>65.47</u> ± 3.48	77.70 ± 1.77	82.26 ± 1.00
	SAHARA	61.17 ± 2.59	68.11 ± 2.18	<u>75.50</u> ± 1.45	<u>81.74</u> ± 1.20
RESISC45 → EURO					
	Method	5	10	25	50
ResNet-18	Target Only	60.57 ± 3.51	70.49 ± 1.71	79.67 ± 0.97	83.96 ± 0.95
	FixMatch [21]	65.31 ± 1.40	72.61 ± 2.76	80.14 ± 1.85	84.17 ± 0.22
	FlatMatch [8]	58.70 ± 2.50	71.82 ± 0.87	79.21 ± 1.40	83.34 ± 0.91
	SoftMatch [2]	63.26 ± 4.26	71.11 ± 2.50	79.02 ± 1.65	84.28 ± 1.32
	SS-HIDA [16]	67.31 ± 2.91	76.43 ± 1.45	81.67 ± 2.16	88.25 ± 0.99
	SHeDD [3]	<u>69.38</u> ± 2.40	<u>77.00</u> ± 2.01	<u>85.11</u> ± 0.91	89.69 ± 1.01
	SAHARA	69.92 ± 1.84	79.69 ± 1.48	85.56 ± 0.93	<u>89.65</u> ± 0.59
TinyViT	Target Only	52.22 ± 3.97	61.32 ± 4.83	74.10 ± 0.85	78.12 ± 1.02
	FixMatch [21]	56.22 ± 2.97	64.36 ± 1.91	73.13 ± 1.37	77.66 ± 0.61
	FlatMatch [8]	52.25 ± 2.36	62.92 ± 2.09	72.51 ± 2.49	77.15 ± 1.47
	SoftMatch [2]	55.75 ± 3.27	64.93 ± 1.77	73.76 ± 1.87	78.85 ± 1.26
	SS-HIDA [16]	54.09 ± 1.97	64.35 ± 1.68	70.73 ± 1.21	77.01 ± 1.05
	SHeDD [3]	<u>64.34</u> ± 3.44	<u>71.17</u> ± 3.39	<u>79.92</u> ± 2.59	<u>86.50</u> ± 1.36
	SAHARA	68.85 ± 1.59	75.20 ± 0.96	81.93 ± 1.53	86.95 ± 0.91

Table 3. F1-score mean and standard deviation (5 runs) on RESISC45-Euro, varying per class target labeled samples in {5, 10, 25, 50}. Best scores are bolded, second-best underlined.

exploit source-domain knowledge to improve target performance, regardless of the backbone or the amount of labeled target data. It secures the top score in 29 out of 32 experiments and ranks second in the few remaining cases. In contrast, semi-supervised approaches that rely solely on target data fail to capitalize on the information from the source domain. As a result, their performance gains are modest at best, and in several instances, they fall below the baseline. Among the SSHDA competitors, SS-HIDA displays limited robustness: it offers only marginal improvements and occasionally performs worse than Target Only (e.g., in EuroSAT – SAR $\rightarrow$ MS and RESISC45-Euro – EURO $\rightarrow$ RESISC45 with TinyViT). Turning to SHeDD, the method achieves second-best results in several cases and the highest score in three experiments. This outcome is likely due to its high-confidence pseudo-labeling strategy, which helps improve performance when the underlying F1-score is already strong. However, when this condition is not met, the approach becomes prone to confirmation bias, causing the model to reinforce incorrect predictions. This leads to a collapse of the pseudo-labeling mechanism, resulting in unstable and degraded performance (such as EuroSAT – MS $\rightarrow$ SAR with TinyViT).

Additional evaluation tables based on overall accuracy confirm the consistency of these results and are available at <https://github.com/giu-guarino/SAHARA>.

Fig. 2 illustrates the per-class F1-scores for the RESISC45-Euro benchmark (EURO $\rightarrow$ RESISC45) using ResNet-18 with 5 labeled target samples per class. *SAHARA* consistently secures the top score across all classes, confirming its robustness in challenging SSHDA scenarios. In contrast, competing methods show more fluctuating behavior, with F1-scores that, for certain classes, even fall below the Target Only baseline.

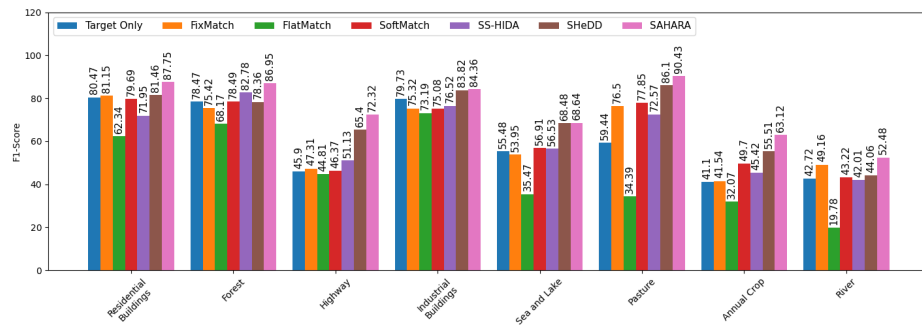


Fig. 2. Per-class F1-score on RESISC45-Euro (EURO $\rightarrow$ RESISC45, 5 labels per class) using ResNet-18 backbone.

Ablation study To assess the individual contribution of each component of *SAHARA*, we conducted an ablation study. TABLE 4 reports the results on the EuroSAT benchmark (MS $\rightarrow$ SAR) using only 5 labeled samples per class,

which represents the most challenging and practically relevant scenario. In our analysis, we progressively disabled key components of the framework: i) the dynamic pseudo-labeling strategy (*DPL*), replacing it with a global fixed threshold (as in FixMatch), ii) the orthogonality loss ($\mathcal{L}_\perp$), and iii) the conditional adversarial domain classifier ($\mathcal{L}_{adv}$). The results clearly demonstrate that the dynamic pseudo-labeling mechanism plays a pivotal role, regardless of the chosen backbone. In particular, replacing it with a global fixed threshold leads to severe performance degradation and, in some cases, model collapse, highlighting the limitations of static confidence thresholds when only a few labeled target samples are available. Once the dynamic pseudo-labeling strategy is introduced, adding the orthogonality loss and the conditional adversarial domain classifier leads to more marginal refinements. Nevertheless, their combined integration consistently results in the best overall performance.

<i>DPL</i>	$\mathcal{L}_\perp$	$\mathcal{L}_{adv}$	F1-score	
			ResNet-18	TinyViT
		✓	24.13 ± 13.73	2.10 ± 0.29
	✓		42.26 ± 8.22	11.49 ± 11.55
	✓	✓	34.48 ± 12.84	4.29 ± 6.05
✓			54.95 ± 4.79	43.01 ± 7.79
✓		✓	54.94 ± 5.09	45.11 ± 3.31
✓	✓		55.61 ± 4.24	47.33 ± 5.95
✓	✓	✓	56.62 ± 4.36	48.28 ± 3.31

Table 4. F1-score mean and standard deviation (5 runs) for ablation study of *SAHARA* on EuroSAT (MS → SAR, 5 labels per class).

4 Conclusions

This paper introduced *SAHARA*, a novel framework for SSHDA designed for scenarios with minimal labeled target data. By combining conditional adversarial feature disentanglement with class-wise dynamic pseudo-labeling, *SAHARA* effectively learns domain-invariant representations across heterogeneous domains, addressing the complex distribution shifts associated with cross-sensor adaptation tasks. Extensive experiments on multimodal remote sensing benchmarks show that *SAHARA* consistently outperforms both baselines and state-of-the-art SSHDA methods, regardless of the feature extractor backbone or the level of target label supervision, with particularly strong gains in low-label settings.

Acknowledgments. The work is financed by the European Union – NextGenerationEU within the National Biodiversity Future Center project (NBFC, CN00000033, CUP B83C22002930006), the French Space Study Center (CNES, TOSCA 2024 project)

and the National Research Agency (ANR) under the ANR-23-IAS1-0002 (GEO-ReSeT) project.

Bibliography

- [1] Carbone, A., Restaino, R., Vivone, G., Chanussot, J.: Model-based super-resolution for sentinel-5p data. *IEEE Trans. Geosci. Remote Sens.* (2024)
- [2] Chen, H., Tao, R., Fan, Y., Wang, Y., Wang, J., Schiele, B., Xie, X., Raj, B., Savvides, M.: Softmatch: Addressing the quantity-quality trade-off in semi-supervised learning. *arXiv* (2023)
- [3] Dantas, C.F., Gaetano, R., Ienco, D.: Semi-supervised heterogeneous domain adaptation via disentanglement and pseudo-labelling. In: *ECML-PKDD* (2024)
- [4] Fang, Z., Lu, J., Liu, F., Zhang, G.: Semi-supervised heterogeneous domain adaptation: Theory and algorithms. *IEEE Trans. on Patt. Analysis and Machine Intell.* (2022)
- [5] Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: *ICML* (2015)
- [6] Guarino, G., Ciotola, M., Vivone, G., Poggi, G., Scarpa, G.: Zero-shot hyperspectral pansharpening using hysteresis-based tuning for spectral quality control. *IEEE Trans. Geosci. Remote Sens.* (2025)
- [7] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR* (2016)
- [8] Huang, Z., Shen, L., Yu, J., Han, B., Liu, T.: Flatmatch: Bridging labeled data and unlabeled data with cross-sharpness for semi-supervised learning. *NeurIPS* (2023)
- [9] Interdonato, R., Ienco, D., Gaetano, R., Ose, K.: Duplo: A dual view point deep learning architecture for time series classification. *ISPRS J. Photogramm. and Remote Sens.* (2019)
- [10] Lakkapragada, A., Sleiman, E., Surabhi, S., Wall, D.P.: Mitigating negative transfer in multi-task learning with exponential moving average loss weighting strategies (student abstract). In: *AAAI* (2023)
- [11] Li, J., Li, G., Shi, Y., Yu, Y.: Cross-domain adaptive clustering for semi-supervised domain adaptation. In: *CVPR* (2021)
- [12] Li, Y., Li, Z., Jia, Y.: Complementary label learning with positive label guessing and negative label enhancement. In: *ICLR* (2025)
- [13] Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. *NeurIPS* **31** (2018)
- [14] Loshchilov, I.: Decoupled weight decay regularization. *arXiv* (2017)
- [15] Mazza, A., Guarino, G., Scarpa, G., Yuan, Q., Vivone, G.: PM2.5 retrieval with sentinel-5p data over europe exploiting deep learning. *IEEE Trans. Geosci. Remote Sens.* (2025)
- [16] Obrenović, M., Lampert, T., Ivanović, M., Gañarski, P.: Learning domain invariant representations of heterogeneous image data. *Machine Learning* (2023)
- [17] Pastorino, M., Moser, G., Serpico, S.B., Zerubia, J.: CRFNet: A deep convolutional network to learn the potentials of a CRF for the semantic seg-

- mentation of remote sensing images. *IEEE Trans. Geosci. Remote Sens.* (2024)
- [18] Peng, J., Huang, Y., Sun, W., Chen, N., Ning, Y., Du, Q.: Domain adaptation in remote sensing image classification: A survey. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* (2022)
 - [19] Qin, C., Wang, L., Ma, Q., Yin, Y., Wang, H., Fu, Y.: Semi-supervised domain adaptive structure learning. *IEEE Trans. Image Process.* (2022)
 - [20] Razzano, F., Di Stasio, P., Mauro, F., Meoni, G., Esposito, M., Schirinzi, G., Ullo, S.L.: AI techniques for near real-time monitoring of contaminants in coastal waters on board future ϕ sat-2 mission. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* (2024)
 - [21] Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *NeurIPS* (2020)
 - [22] Song, J., Kim, T.S., Kim, J., Nam, G., Kooi, T., Choo, J.: Is user feedback always informative? retrieval latent defending for semi-supervised domain adaptation without source data. In: *ECCV* (2024)
 - [23] Wang, Y., Hernández, H.H., Albrecht, C.M., Zhu, X.X.: Feature guided masked autoencoder for self-supervised learning in remote sensing. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* (2024)
 - [24] Wilson, G., Cook, D.J.: A survey of unsupervised deep domain adaptation. *ACM Trans. on Intell. Sys. and Tech. (TIST)* (2020)
 - [25] Wu, K., Zhang, J., Peng, H., Liu, M., Xiao, B., Fu, J., Yuan, L.: Tinyvit: Fast pretraining distillation for small vision transformers. In: *ECCV* (2022)
 - [26] Yang, L., Wang, Y., Gao, M., Shrivastava, A., Weinberger, K.Q., Chao, W.L., Lim, S.N.: Deep co-training with task decomposition for semi-supervised domain adaptation. In: *ICCV* (2021)