Unlocking the Potential of Unlabeled Data in Semi-Supervised Domain Generalization

Supplementary Material

A. Dataset Description

We conduct experiments on four SSDG benchmarks. The PACS dataset comprises four domains: art-painting, cartoon, photo, and sketch, and includes seven classes: dog, elephant, giraffe, guitar, horse, house, and person. The OfficeHome dataset consists of four domains: art, clipart, product, and real-world, with a total of 65 classes. The miniDomainNet dataset includes four domains: clipart, painting, real, and sketch, and contains 345 classes. Lastly, the DigitsDG dataset is composed of four domains: MNIST, MNIST-M, SVHN, and SYN, and features 10 classes, ranging from 0 to 9.

B. Psuedo Code Algorithm

To illustrate the detailed implementation of the UPCSC method, we provide a pseudo code in Algorithm A. For simplicity, the domain label index d, weak augmentation α, and strong augmentation $\mathcal{A}$ are omitted.

C. Loss for SSL-based baseline

In this section, we give a detailed description on loss for SSL-based baselines, such as FixMatch. SSL-based baselines comprise a supervised loss (L_{sup}) using labeled data x^l, and an unsupervised consistency loss (L_{unsup}) which leverages *confident-unlabeled samples* x^{uc}. L_{sup} is computed for labeled data as follows:

$$L_{sup} = \text{CE}(\text{softmax}((h \circ f)(x^l)), y^l). \tag{6}$$

The unsupervised consistency loss, L_{unsup}, utilizes *confident-unlabeled samples* x^{uc} with pseudo-labels $\hat{y} = \text{argmax}(\text{softmax}((h \circ f)(\alpha(x^{uc}))))$ generated via weak augmentation α. This loss ensures that the predictions for strongly augmented samples $\mathcal{A}(x^{uc})$ are aligned with their pseudo-labels. Formally, the unsupervised consistency loss is computed as follows:

$$L_{unsup} = \text{CE}(\text{softmax}((h \circ f)(\mathcal{A}(x^{uc}))), \hat{y}). \tag{7}$$

D. T-SNE Visualization by Domain

In Fig. A, we provide the visualizations of t-SNE for each domain, a more specific version of t-SNE represented in the analysis section, Fig. 6. To explain further the visualization procedure, the t-SNE visualization for a specific domain is performed on a model trained with that domain as the target domain. Since the model is not accessible to the target domain data during training, the labeled and unlabeled train data (source data) are drawn from the remaining domains, excluding the target domain. The data corresponding to the target domain can only be found in the test data.

E. Detailed Visualization on Accuracy of unlabeled train data

In Fig. B, we provide the plot on the accuracy of unlabeled train data for each domain rather than the average which was represented at Fig. 5.

F. More Ablation Studies on our proposed modules

Table A summarizes the contributions of our proposed modules to overall performance in various settings. These results demonstrate that the proposed UPC and SC modules substantially boost performance.

Table A. Ablation study of how the proposed UPC and SC modules contribute to performance across various experimental settings.

	PACS		OH	
Labels per class	10	5	10	5
FixMatch	76.8	73.6	57.7	55.0
+UPC	79.2	76.8	58.4	55.9
+SC	77.0	77.8	58.4	55.5
+UPCSC	79.6	78.9	58.6	56.1

G. Ablation study on UPC

Table B illustrates the importance of incorporating *unconfident-unlabeled samples* x^{uu} within the UPC module. As shown in the table, excluding these samples yields a 1.2%p improvement over the baseline, while their inclusion raises the margin to 2.4%p. This result clearly demonstrates the benefit of leveraging such *unconfident-unlabeled samples* for performance enhancement.

H. Variants of SC module

Table C presents an ablation study on strategies for generating positive pairs from *unconfident-unlabeled samples* in the SC module. Here, **Top-1** denotes the results obtained by using only the proxy of the highest-confidence class, while **Avg. Proxy** represents those obtained by averaging all candidate class proxies. Finally, **Ours** is based on a weighted average of the candidate class proxies.

Table B. Ablation study on UPC without unconfident-unlabeled samples x^{uu} on PACS 10 labels per class setting. Each result represents the average accuracy.

Method	Accuracy (%)
FixMatch	76.8
+UPC (w/o x^{uu})	78.0
+UPC	79.2
+UPCSC	79.6

Table C. Comparison of different positive pair selection strategies for SC in PACS 10 labels per class setting.

Method	Top-1 Proxy	Avg. Proxy	Ours
FixMatch	78.9	80.5	79.6
StyleMatch	78.3	80.2	81.5

I. Additional Experiment Results

Table D shows the results of applying the plug-and-play methods to FreeMatch as a complement to Table 4, which presents the results of applying these plug-and-play methods to FixMatch and StyleMatch. Notably, when applied to FreeMatch, our method demonstrates superior performance compared to other plug-and-play approaches across all datasets. Due to space constraints in the main text, the results for FreeMatch are included in the supplementary material.

J. Code Asset

In Sec. 5.2, we used the benchmark introduced in StyleMatch [32]. The code of this work is also built upon this work. Authors thank to their open sourcing.

Table D. Comparison of various plug-and-play methods incorporated on FreeMatch [26] in SSDG under 10 labels and 5 labels per class settings. Each result represents the average over five different random seeds.

Model	Labels per class = 10				Labels per class = 5			
	PACS	OH	DigitsDG	DN	PACS	OH	DigitsDG	DN
FreeMatch [26]	73.5 ± 1.1	57.7 ± 0.4	74.2 ± 2.1	54.8 ± 0.2	71.6 ± 1.8	55.9 ± 0.5	63.3 ± 2.0	52.0 ± 0.7
FreeMatch + FBCSA [9]	73.7 ± 2.3	58.6 ± 0.4	78.7 ± 0.9	55.5 ± 0.3	69.2 ± 1.4	55.8 ± 0.3	76.2 ± 1.0	51.0 ± 0.7
FreeMatch + DGWM [10]	73.3 ± 1.3	57.6 ± 0.4	74.0 ± 0.7	54.7 ± 0.3	72.2 ± 1.9	55.8 ± 0.6	62.2 ± 4.3	52.0 ± 0.5
FreeMatch + Ours	**77.8 ± 1.4**	**59.1 ± 0.5**	**80.4 ± 0.7**	**56.5 ± 0.3**	**73.5 ± 2.1**	**56.8 ± 0.8**	**76.4 ± 0.6**	**53.7 ± 0.4**

Algorithm A Pseudo Code of UPCSC

Require: Labeled data (x^l, y^l), unlabeled data (x^u), confidence threshold τ, number of classes C, total epochs E, feature projector p_f, classifier projector p_c, featurizer f, classifier $h = [h_1, h_2, \ldots, h_C]$, normalization operation $\|\cdot\|$, indexing operation $[\cdot]_i$ for selecting i-th element.

for epoch = 1 to E **do**
  Step 1: Divide *Confident-Unlabeled* and *Unconfident-Unlabeled Samples*
  $c(x) = \text{softmax}((h \circ f)(x))$
  $x^{uc} = \{x \mid \max(c(x)) \geq \tau,\ x \in x^u\}, \quad N^{uc} = |x^{uc}|$
  $x^{uu} = \{x \mid \max(c(x)) < \tau,\ x \in x^u\}, \quad N^{uu} = |x^{uu}|$
  Step 2: Compute Pseudo Labels for *Confident-Unlabeled Samples*
  $\hat{y} = \text{argmax}(c(x^{uc}))$
  Step 3: Define Candidate and Excluded Class Sets for *Unconfident-Unlabeled Samples*
  for x^{uu} index i = 1 to N^{uu} **do**
    $\mathcal{C}_i = \{y \mid [c(x_i^{uu})]_y > 1/C\}, \quad \mathcal{E}_i = \{y \mid y \notin \mathcal{C}_i\}$
  end for
  Step 4: Compute Supervised Loss for Labeled Data
  $L_{\text{sup}} = \text{CrossEntropy}(c(x^l)), y^l)$
  Step 5: Compute Unsupervised Consistency Loss for *Confident-Unlabeled Samples*
  $L_{\text{unsup}} = \text{CrossEntropy}(c(x^{uc}), \hat{y})$
  Step 6: Compute Unlabeled Proxy-based Contrastive Loss
  for x^{uc} index i = 1 to N^{uc} **do**
    $z_i^{uc}, w_i^{uc} = \|p_f(f(x_i^{uc}))\|,\ \|p_c(h_{\hat{y}_i})\|$
    $\mathcal{L}_{\text{UPC}} \mathrel{+}= -\frac{1}{N^{uc}} \log \frac{\exp(z_i^{uc} \cdot w_{\hat{y}_i}^{uc})}{\exp(z_i^{uc} \cdot w_{\hat{y}_i}^{uc}) + \sum_{\{j \mid \hat{y}_j \neq \hat{y}_i\}} \exp(z_i^{uc} \cdot z_j^{uc}) + \sum_{\{j \mid \hat{y}_i \in \mathcal{E}_j\}} \exp(z_i^{uc} \cdot z_j^{uu})}$ ▷ Eq. (2)
  end for
  Step 7: Compute Surrogate Class Learning
  for x^{uu} index i = 1 to N^{uu} **do**
    $z_i^{uu}, w_i^{uu} = \|p_f(f(x_i^{uu}))\|, \{\|p_c(h_j)\| \mid j \in \mathcal{C}_i\}$
    $SC(x_i^{uu}) = \sum_{j=1}^{C} \mathbb{1}(j \in \mathcal{C}_i) \cdot [c(x_i^{uu})]_j \cdot [w_i^{uu}]_j$
    $\mathcal{L}_{\text{SC}} \mathrel{+}= -\frac{1}{N^{uu}} \log \frac{\exp(z_i^{uu} \cdot SC(x_i^{uu}))}{\exp(z_i^{uu} \cdot SC(x_i^{uu})) + \sum_{\{j \mid \hat{y}_j \in \mathcal{E}_i\}} \exp(z_i^{uu} \cdot z_j^{uc}) + \sum_{\{j \mid \mathcal{C}_j \cap \mathcal{C}_i = \emptyset\}} \exp(z_i^{uu} \cdot z_j^{uu})}$ ▷ Eq. (4)
  end for
  Step 8: Compute Total Objective and Update Parameters
  $L_{\text{total}} = L_{\text{sup}} + L_{\text{unsup}} + \mathcal{L}_{\text{UPC}} + \mathcal{L}_{\text{SC}}$
  $\texttt{update}(f, h, p_c, p_f; L_{\text{total}})$
end for

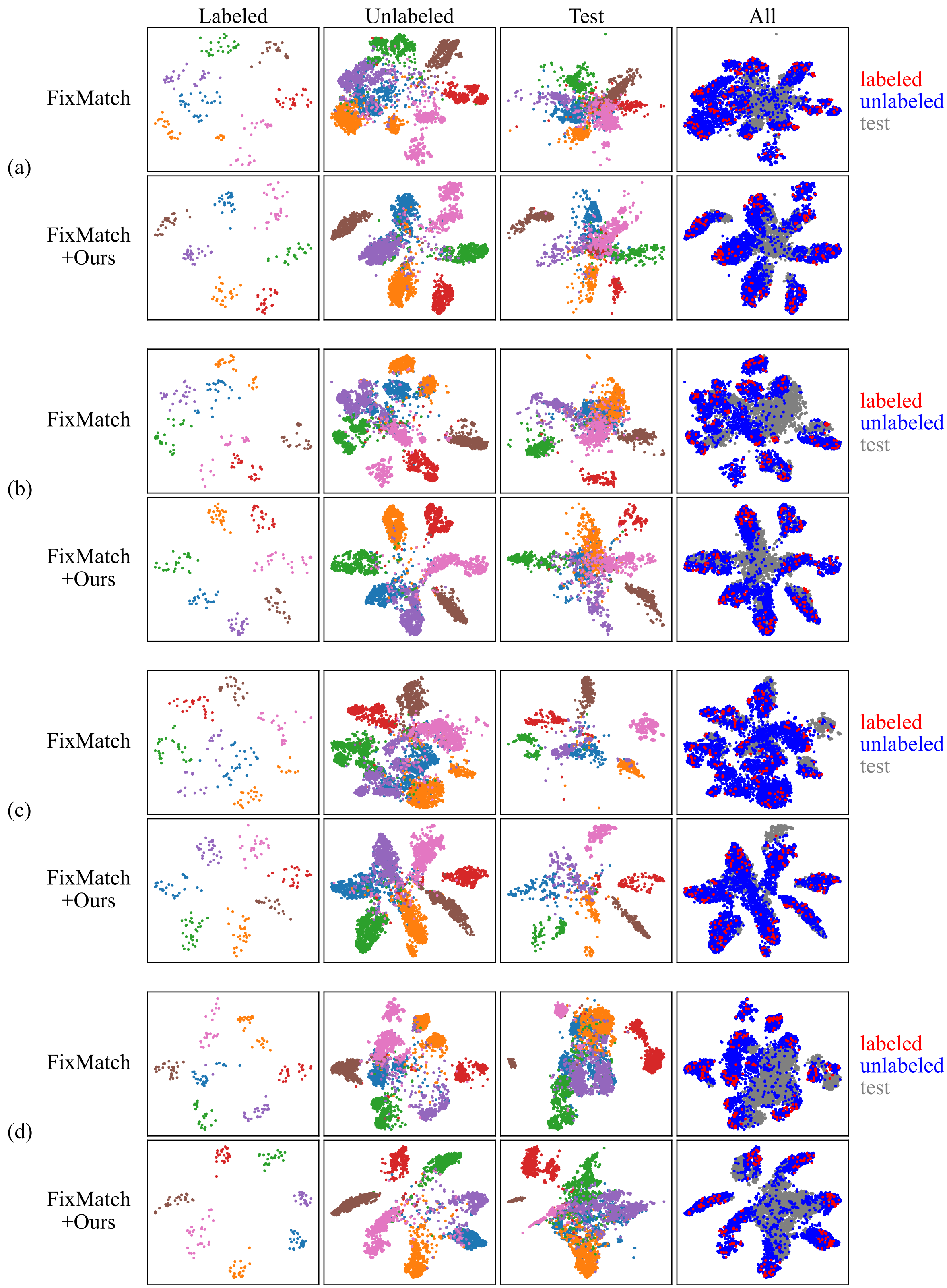

Figure A. T-SNE visualization of FixMatch and FixMatch + Ours in the PACS dataset under the 10 labels per class setting are presented for each domain. Each sub-figure corresponds to the target domain being (a) art-painting, (b) cartoon, (c) photo, and (d) sketch, respectively.

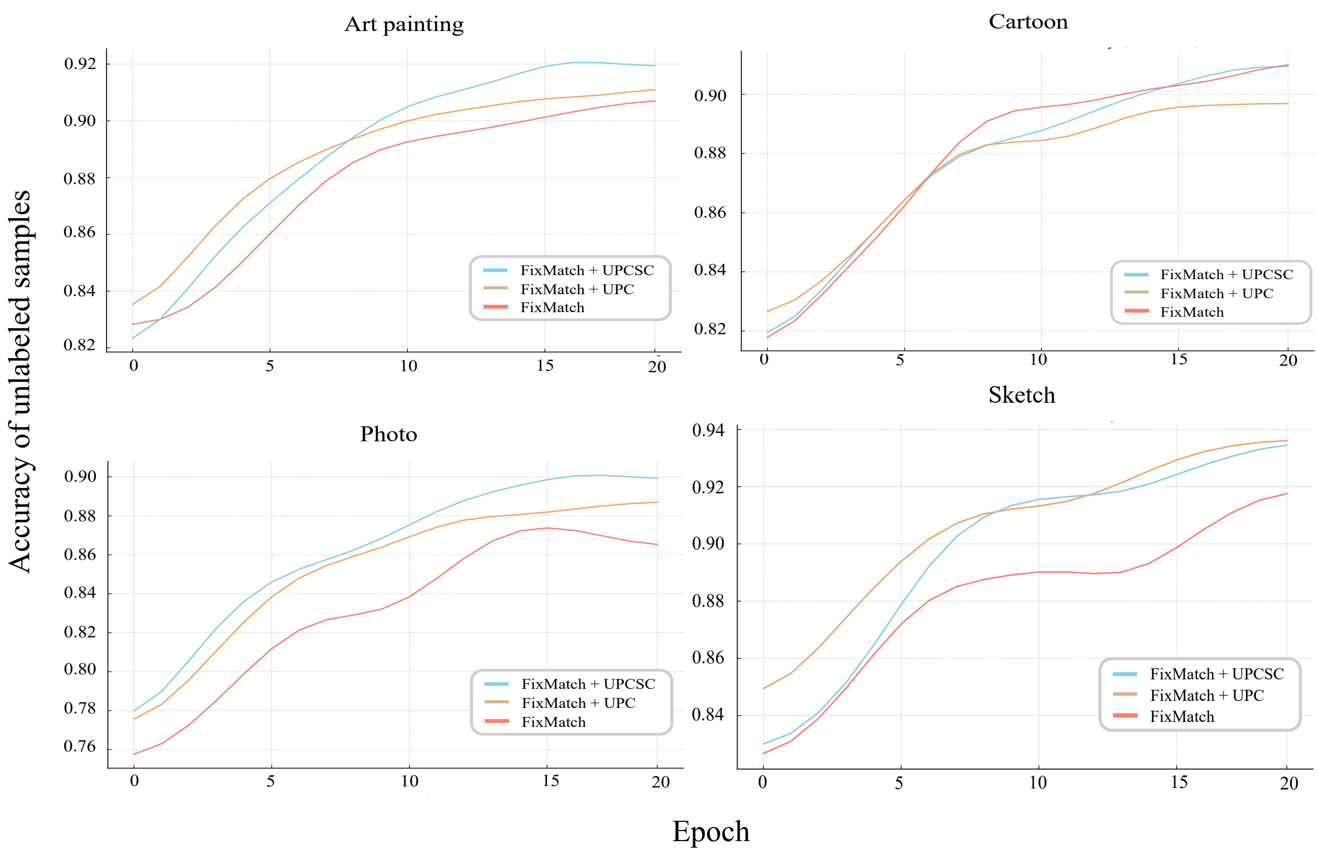

Figure B. Accuracy of unlabeled samples from the source domain in the PACS 10 labels per class setting for each domain.