A. Detailed description of datasets

We conduct experiments on six challenging and diverse DG datasets to validate the effectiveness of the proposed method. **PACS** [25] contains 7 categories of images from four domains (Photo, Art painting, Cartoon, and Sketch). **OfficeHome** [43] consists of images from four different domains (Art, Clipart, Product, and Real-world). It encompasses 65 object categories that are commonly encountered in office and home environments. **VLCS** [14] comprises images spanning across four domains with 5 categories and has four domains (Caltech, Labelme, SUN, and Pascal). **Digits-DG** [44] includes digit images drawn from MNIST, SVHN, MNIST-M and SYNTH. **Terra Incognita** [5] contains photos of wild animals taken by cameras at different locations (location 38, location 43, location 46, and location 100) with 10 classes. **DomainNet** [32] is a large-scale dataset of common objects in six different domains (clipart, infograph, real, painting, quickdraw, sketch) with 345 categories of objects.

B. Impact of noise perturbation injected

We vary the variance ε^2 of the isotropic Gaussian $\mathcal{N}(0, \varepsilon^2 I)$ and evaluate the impact on our method in Tab. 9. We note that the performance of the method is mostly insensitive to variances 0.1, 0.5, and 1.0. The best performance is achieved at 1.0, however, it decreases upon doubling the variance to 2.0.

ε^2	Average
0.1	59.5
0.5	59.3
1.0	**59.7**
2.0	57.3

Table 9. Results with different values of variance ε^2 of the isotropic Gaussian $\mathcal{N}(0, \varepsilon^2 I)$. Results are shown for the OfficeHome dataset under 10 labels setting.

C. Pseudo-labeling accuracy vs. Confidence threshold

Fig. 5 (left) shows the variation of pseudo-labeling accuracy against the confidence threshold [38] on the OfficeHome dataset under the 10 labels setting for FixMatch [38] and our method. Our proposed method retains a higher pseudo-labeling accuracy than the baseline [38] even when we lower the confidence threshold. Furthermore, we plot the unlabeled data utilization i.e. the percentage of unlabelled data that passes the confidence threshold for both FixMatch and our method as the confidence threshold varies (see Fig. 5 (right)). The weight modulation technique in our method tends to reduce the model's maximum confidence when computing pseudo-labels. As a result, only highly accurate pseudo-labels will make it past the threshold.

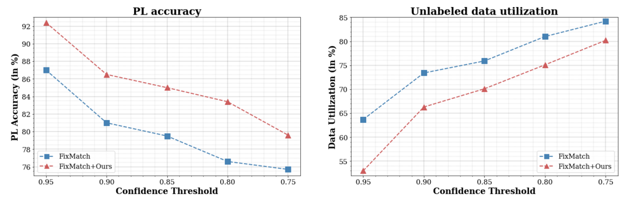

Figure 5. (Left) Pseudo-label accuracy upon varying the confidence threshold in FixMatch and our method. (Right) Unlabelled data utilization i.e. the percentage of unlabelled data that passes the confidence threshold for both FixMatch and our method. These results are shown on the OfficeHome dataset with 10 label settings.

D. Comparison with DG baselines

We show the performance of several DG methods: (ERM [42], MixUp [56], and GroupDRO [35] and also show results after combining these DG methods and pseudo-labelling from FixMatch [38]. Tab. 10 and Tab. 11 report results with the first SSDG setting and the second SSDG setting, respectively.

E. Performance under class-imbalance

VLCS [14] has a significant class imbalance than most of the DG datasets. In Tab. 12 we calculate the ratio between the number of samples for the highest and lowest available classes in each domain. It should be noted that our proposed method shows notable gains of $+5.3\%$ and $+5.2\%$ for 5 and 10 labels settings respectively.

F. Architectural details of encoder-decoder-like pair

For the encoder, we use 3 linear layers each followed by a ReLU [2] activation layer, and reduce the size of the embedding dimension by a factor of 2. Intermediate embedding concatenated with noise will follow a two-linear layer decoder each followed by a ReLU [2] activation layer.

G. t-SNE visualization of domain information vector

The mini-batch mean is a simple way of aggregating the domain-specific information [13, 21] as samples in the same mini-batch are drawn from the same domain. t-SNE visualization (see Fig. 6) of domain information vectors I^k taken during training indicates that these domain information vectors are distinct for each source domain (3 source domains).

Method	5 labels PACS	OfficeHome	VLCS	DigitsDG	TerraInc	DomainNet	10 labels PACS	OfficeHome	VLCS	DigitsDG	TerraInc	DomainNet
ERM	$51.2_{\pm 3.0}$	$51.7_{\pm 0.6}$	$67.2_{\pm 1.8}$	$22.7_{\pm 1.0}$	$22.9_{\pm 3.0}$	$23.5_{\pm 0.2}$	$59.8_{\pm 2.3}$	$56.7_{\pm 0.8}$	$68.0_{\pm 0.3}$	$29.1_{\pm 2.9}$	$23.5_{\pm 1.2}$	$29.4_{\pm 0.1}$
MixUp	$45.3_{\pm 3.8}$	$52.7_{\pm 0.6}$	$69.9_{\pm 1.3}$	$21.7_{\pm 1.9}$	$21.0_{\pm 2.9}$	$23.5_{\pm 0.3}$	$58.5_{\pm 2.2}$	$57.2_{\pm 0.6}$	$69.6_{\pm 1.0}$	$29.7_{\pm 3.1}$	$24.8_{\pm 3.3}$	$28.8_{\pm 0.1}$
GroupDRO	$48.2_{\pm 3.6}$	$53.8_{\pm 0.6}$	$69.8_{\pm 1.2}$	$23.1_{\pm 1.9}$	$22.4_{\pm 3.1}$	$20.2_{\pm 0.2}$	$57.3_{\pm 1.2}$	$57.8_{\pm 0.4}$	$69.4_{\pm 0.9}$	$31.5_{\pm 2.5}$	$25.8_{\pm 3.3}$	$26.5_{\pm 0.5}$
ERM + PL	$62.8_{\pm 3.0}$	$54.2_{\pm 0.6}$	$65.4_{\pm 2.9}$	$43.4_{\pm 2.9}$	$25.4_{\pm 3.2}$	$24.1_{\pm 0.2}$	$63.0_{\pm 1.5}$	$55.5_{\pm 0.3}$	$60.5_{\pm 1.1}$	$55.0_{\pm 2.4}$	$26.8_{\pm 1.5}$	$26.7_{\pm 0.1}$
MixUp + PL	$60.6_{\pm 2.9}$	$51.9_{\pm 0.4}$	$60.8_{\pm 2.8}$	$35.4_{\pm 1.3}$	$24.1_{\pm 3.0}$	$23.3_{\pm 0.2}$	$62.3_{\pm 1.9}$	$55.1_{\pm 0.2}$	$64.4_{\pm 1.1}$	$43.5_{\pm 1.0}$	$27.6_{\pm 2.2}$	$28.5_{\pm 0.3}$
GroupDRO + PL	$62.3_{\pm 1.9}$	$54.5_{\pm 0.5}$	$69.3_{\pm 0.3}$	$39.4_{\pm 1.3}$	$25.1_{\pm 3.2}$	$25.6_{\pm 0.2}$	$62.1_{\pm 2.0}$	$58.5_{\pm 0.3}$	$66.5_{\pm 0.2}$	$49.9_{\pm 1.9}$	$26.9_{\pm 1.2}$	$28.0_{\pm 0.1}$

Table 10. Comparison with the DG methods, DG+PL [38] methods under the first setting i.e only a few instances(5,10) are labeled from each source domain.

Method	PACS	OfficeHome	VLCS	Digits	TerraInc	DomainNet
ERM	69.8±1.8	61.7±0.4	60.8±0.7	36.7±0.7	40.0±2.3	33.1±0.1
MixUp	66.9±1.9	61.6±0.2	61.3±0.5	40.1±1.0	40.1±0.8	33.9±0.1
GroupDRO	71.6±1.3	63.7±0.1	61.5±0.7	38.8±0.7	40.5±1.3	34.1±0.1
ERM+PL	65.2±1.6	60.4±0.4	50.5±0.8	53.4±0.9	41.1±0.8	31.4±0.1
MixUp+PL	66.9±1.4	62.0±0.3	55.9±0.4	49.3±0.3	38.2±1.3	35.5±0.2
GroupDRO+PL	78.6±1.9	64.5±0.1	55.8±0.6	40.5±1.0	42.5±0.4	35.1±0.1

Table 11. Comparison with the DG methods, DG+PL [38] methods under the first setting i.e one source domain is completely labeled and the other completely unlabeled.

Domain	VLCS # of samples Highest	Lowest	Ratio
Caltech	809	62	13.0
LabelMe	1124	39	28.9
Pascal	1394	307	4.6
SUN	1175	19	61.9

Table 12. Num. of samples for highest and lowest available classes for each domain.

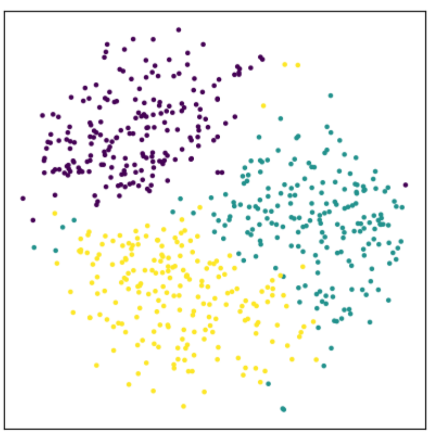

Figure 6. t-SNE visualization of domain information vectors I^k taken during training on OfficeHome dataset.

H. Additional details on the second setting

Under the second SSDG setting, for a given dataset, we select a target domain and keep it fixed while making each source domain labeled and others unlabeled and report the average recognition accuracy. The fixed target domain in each dataset is as follows: ”Photo” in PACS, ”Real-world” in OfficeHome, ”SUN” in VLCS, ”SVHN” in DigitsDG, ”location 100” in TerraIncognita and ”Real” in DomainNet. Each experiment is conducted for 5 independent trials.