Towards Generalizing to Unseen Domains with Few Labels

Chamuditha Jayanga Galappaththige[1,2*], Sanoojan Baliah[1*], Malitha Gunawardhana[1,3],
Muhammad Haris Khan[1]

[1]Mohamed Bin Zayed University of AI [2]Queensland University of Technology [3]University of Auckland

{chamuditha.jayanga,sanoojan.baliah,malitha.gunawardhana,muhammad.haris}@mbzuai.ac.ae

Abstract

We approach the challenge of addressing semi-supervised domain generalization (SSDG). Specifically, our aim is to obtain a model that learns domain-generalizable features by leveraging a limited subset of labelled data alongside a substantially larger pool of unlabeled data. Existing domain generalization (DG) methods which are unable to exploit unlabeled data perform poorly compared to semi-supervised learning (SSL) methods under SSDG setting. Nevertheless, SSL methods have considerable room for performance improvement when compared to fully-supervised DG training. To tackle this underexplored, yet highly practical problem of SSDG, we make the following core contributions. First, we propose a feature-based conformity technique that matches the posterior distributions from the feature space with the pseudo-label from the model's output space. Second, we develop a semantics alignment loss to learn semantically-compatible representations by regularizing the semantic structure in the feature space. Our method is plug-and-play and can be readily integrated with different SSL-based SSDG baselines without introducing any additional parameters. Extensive experimental results across five challenging DG benchmarks with four strong SSL baselines suggest that our method provides consistent and notable gains in two different SSDG settings. Our code is available at FBC-SA.

1. Introduction

Top-performing visual object recognition models [13, 18, 31] tend to sacrifice performance when there is a difference between the training and testing data distributions, also termed as the domain shift problem [49, 52]. To handle the domain shift problem, the research line of domain generalization (DG), amongst others, has been explored with greater interest. As such, DG places fewer assumptions and so it is potentially more widely applicable than the other alternatives e.g., unsupervised domain adaptation (UDA)

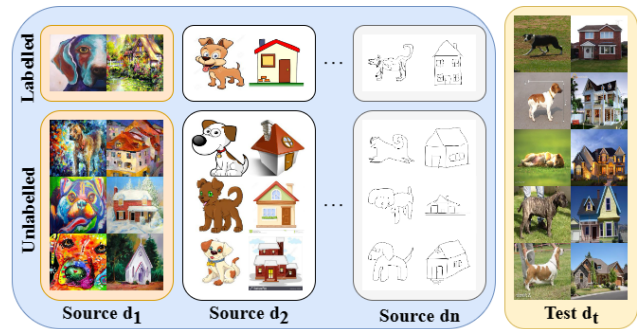*Equally contributing authors

Figure 1. Visual illustration of Semi-supervised Domain Generalization (SSDG) setting.

[53]. In DG, the goal is to learn a generalizable model, relying on data from multiple source domains for training, that is capable of performing well on data from an unseen target domain [7, 23, 28]. In the recent past, several promising DG methods have been proposed that leverage different feature-level constraints [28], adapt meta-learning frameworks [4, 24], leverage proxy tasks [8, 40], or develop data augmentation mechanisms [20, 56]. Note that almost all existing DG methods operate under fully supervised settings i.e. the data from all source domains are completely labeled.

In many real-world use-cases, for instance, in healthcare, it is often hard, if not impossible, to acquire a sufficiently large set of labeled data from all source domains [46]. Usually, only a small subset of data is labelled, and the remaining large fraction of data is unlabelled. Therefore, besides, aiming to achieve cross-domain generalization, a model should be able to rely on limited labels [48]. The topic of semi-supervised learning (SSL) is quite relevant in this case [6, 22, 32, 33], which aims to leverage abundantly available unlabeled data with a small fraction of labeled data to achieve learning.

In our work, we aim to study the problem of semi-supervised domain generalization (SSDG), which unifies domain generalization and data efficiency under the same framework [57]. SSDG is similar to DG in terms of core objective i.e. learning a generalizable model by leverag-

ing multiple source domains. However, DG assumes that the data from all source domains are fully-labelled. On the contrary, SSDG operates under SSL settings where only a handful of labeled data is available while a relatively large chunk of data is unlabeled [57]. Fig. 1 draws a visual illustration of the SSDG setting. It is shown that the DG methods tend to perform poorly under the limited labels setting of SSDG as they are not developed to exploit the unlabeled data. On the other hand, SSL methods, in particular, [32], perform relatively better than the DG methods [57], but still their obtained performance is considerably lower than the fully supervised training. Fig. 2 shows the comparison of results in SSDG problem setting among the following methods: DG, DG combined with pseudo-labeling (following [32]) and SSL. Note that, the naive combination of DG and SSL methods also performs poorly [57].

We propose a new approach for tackling semi-supervised domain generalization (SSDG) by observing the key limitations in best-performing SSL-based baselines for SSDG. A dominant challenge is how to achieve accurate pseudo-labels (PLs) when the unlabeled data exhibit various domain shifts. It is further exacerbated by the scarcity of labeled data, which possibly increases the chances of model overfitting. Towards addressing them, we resort to the feature space of the model and propose a feature-based conformity (FBC) module and a semantics alignment (SA) loss. The feature-based conformity module aims at aligning the posteriors from the feature space with the pseudo-labels from the model's output space via two different constraints. To learn semantically harmonious features under unlabelled data from multiple heterogeneous sources, we present a semantics alignment loss that attempts to regularize the semantic structure in the feature space by domain-aware similarity-guided cohesion and repulsion of examples. In summary, we make the following key contributions:

- We study the relatively unexplored yet highly practical problem of semi-supervised domain generalization (SSDG) and propose a new approach, consisting of feature-based conformity and semantics alignment loss, for addressing the important challenges in SSDG.
- Our approach is plug-and-play and as such it can be seamlessly applied to different SSL-based SSDG baselines without adding any learnable parameters. We show the adaptability and effectiveness of our method with four strong baselines.
- We perform extensive experiments on five different DG datasets: PACS [23], OfficeHome [36], DigitsDG [55], TerraIncognita [5] and VLCS [34] with four strong baselines: FixMatch [32], FlexMatch [50] and FreeMatch [41] StyleMatch[57]. Our approach delivers consistent and visible gains across all datasets with four baselines in two variants of SSDG settings.
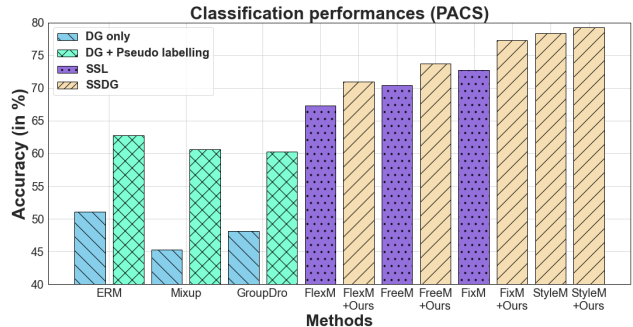

Figure 2. Recognition performance comparison between different DG, DG combined with pseudo-labeling, SSL methods and ours in SSDG settings. Here, GD - GroupDro, FlexM - FlexMatch, FreeM - FreeMatch, FixM - FixMatch and StyleM - StyleMatch.

2. Related Work

Domain generalization: Several studies have been conducted towards improving domain generalization (DG) performance. Most methods aim to learn domain-invariant features from available source domain data [15, 28]. Empirical risk minimization (ERM) can be regarded as the earliest attempt that aims to reduce the sum of errors across data aggregated from multiple source domains [35]. Following DG works, utilized maximum mean discrepancy (MMD) constraint [28], developed multi-task autoencoder [15], achieved adversarial feature learning with MMD[25], learned invariant predictors [2], introduced low-rank regularization [23, 44] in pursuit of extracting domain-invariant features. Another line of work adapted the meta-learning framework to simulate domain drifts during training [24]. Furthermore, some work leveraged proxy tasks [8, 40] to promote domain-generalizable features. Some used domain-specific masks [10] and domain-specific normalizations [29] to strike the balance between domain-specific and domain-invariant features. Inspired by the contrastive learning paradigm, some DG works adapted self-supervision and different variants of ranking losses [14, 21, 27]. [9] proposed stochastic weight averaging in a dense manner to achieve flatter minima for DG. Many DG approaches proposed new techniques to synthesize examples from novel domains to increase the diversity in source domains. [30] devised crossgrad training [30], [37] imposed a wassertein constraint in semantic space, [56] adapted a CNN-based generator. [20] exploited class-conditional covariance to augment novel source domain features. [45] used mixup for creating new images by mixing source images.

However, most of the existing DG works expect fully labelled data from all source domains. Their performance degrades substantially upon reducing the amount of labeled data (see Fig. 2), which limits their applicability to several important application domains, including medical imaging and autonomous vehicles, where sufficiently labelled data
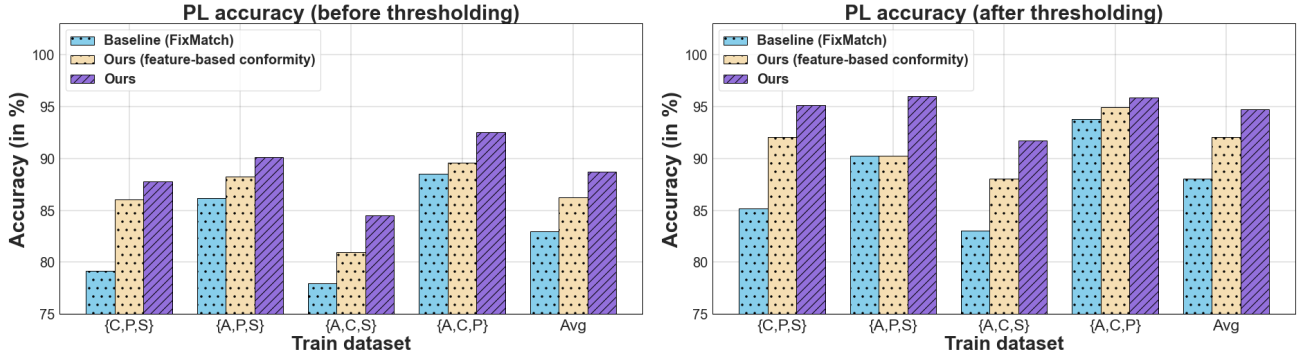
Figure 3. PL accuracy in the training samples without thresholding (left) and for selected PL after thresholding (right) for the baseline (FixMatch [32]), ours with only feature-based conformity, and ours for PACS dataset in 5 labels per class setting. Here A, C, P, and S denote Art-painting, Cartoon, Photos, and Sketch domains, respectively.

is scarce. To this end, in this paper, we study the relatively underexplored problem of semi-supervised domain generalization (SSDG), which unifies the domain generalization and data efficiency under a common framework, and propose a new SSDG approach based on feature-based conformity and semantics alignment constraint.

Semi-Supervised Learning: The research direction of Semi-Supervised Learning (SSL) has seen numerous methods developed in the literature, with entropy minimization [16] consistency learning [26, 33, 42] and pseudo-labelling [22, 32] being the most prominent approaches. Consistency learning involves making predictions of a model on two different views of the same input similar to each other [54] by imposing a consistency loss on penultimate features [1] or output probabilities [32]. Recently, [33] found that using a model's exponential moving average to generate the target for consistency learning improves performances. On the other hand, pseudo-labeling [22] generates either soft or hard pseudo-labels for unlabeled data using a pre-trained model [42] or the model under training [32]. Further, [6, 32, 42, 43] demonstrated that inducing strong noise such as the strong augmentation or dropout to the student model can significantly boost performance. [11, 41, 50] focus on improving pseudo-labelling building upon [32]. To address distribution shifts between labeled and unlabeled data caused by sampling bias, some studies [1, 38] have adopted ideas from domain adaptation [19] to minimize the feature distance.

While semi-supervised domain generalization (SSDG) and SSL both deal with unlabelled data, SSDG poses a greater challenge as the data is collected from heterogeneous sources with potentially different underlying data distributions. SSDG is a relatively underexplored problem and so very little research has been done. [47] proposed active exploration, which queries labels of examples with higher ranks in class uncertainty, domain representativeness and information diversity, and combines inter

and intra-domain knowledge with mixup [51]. [48] proposed a method that relies on a generated similarity graph and a graph Laplacian regularizer. [39] proposed a joint domain-aware label and dual classifier framework for learning a domain-generalizable model when only one source domain is fully labelled while the others are completely unlabelled. Recently, Stylematch [57] extends [32] for SSDG with stochastic modeling to reduce overfitting and multi-view consistency learning for generalizing across domains. Multi-view consistency operates within the confines of the input pixel space and assumes some style variance between the data distributions. Such an approach struggles or is unable to improve performance when presented with different types of distributions such as background shifts or corruption shifts as shown in our experiments (sec. 4). In this work, we propose a new SSDG approach, for partially-labelled source domains, that explores the feature space to develop a feature-based conformity mechanism and a semantics alignment constraint. Feature space provides more flexibility compared to input pixel space for imposing various consistencies. This is beneficial towards effectively tackling SSDG challenges as demonstrated in our results.

3. Methodology

3.1. Preliminaries

Problem Settings: We adapt some notation from [17]. Similar to multi-source DG settings, let us characterize each domain d by $d = \{(\mathbf{x}_i^d, y_i^d)\}_{i=1}^n$, where $\mathbf{x}_i^d \in \mathbb{R}^{C \times H \times W}$ is an input image, y_i^d is the corresponding label, and comprised of n^* independent and identically distributed (i.i.d) examples drawn from a joint probability distribution $\mathcal{P}(\mathcal{X}^d, \mathcal{Y}^d)$ for all possible training (source) domains $d \in \{d_1, ..., d_{tr}\}$. $\mathcal{X}^d$ is an input space over which the domain d is defined and $\mathcal{Y}^d$ is the corresponding label space. Here, we consider the distribution shift in $\mathcal{P}(\mathcal{X}^d)$ while $\mathcal{P}(\mathcal{Y}^d)$

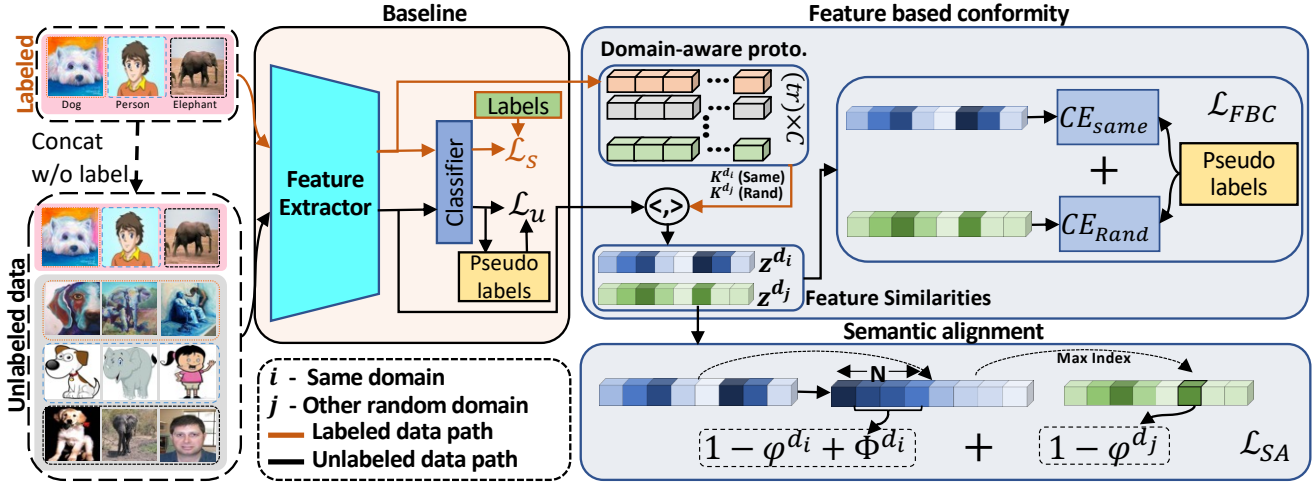*The value of n can be different for each domain.

Figure 4. Overall architecture of our method. Fundamentally, it is a semi-supervised baseline (e.g., FixMatch [32]) with a feature extractor and a classifier that involves pseudo-labelling and prediction consistency mechanisms. To tackle semi-supervised domain generalization, we first propose a *feature-based conformity* module (sec. 3.2) that aligns the posterior from feature space with the pseudo-label from output space. We then develop a *semantics alignment loss* (sec. 3.3) to regularize the semantic layout of feature space and further improve the effectiveness of feature-based conformity.

shares the same label space $\mathcal{Y}$ where $y \in \mathcal{Y} = \{1, 2, .., C\}$ is an associated class label.

In the SSDG setting, the number of labeled examples is constrained i.e., each source domain d has a labeled part $d^L = \{(\mathbf{x}_i^d, y_i^d)\}$ and an unlabeled part $d^U = \{(\mathbf{u}_i^d)\}$. Further, the number of examples in the unlabeled part is much higher than in the labeled part, i.e., $|d^U| \gg |d^L|$. Our goal is to learn a domain-generalizable model $\mathcal{F}$ using the training (source) domains $d \in \{d_1, ..., d_{tr}\}$ to accurately predict on an out-of-distribution data, whose examples are drawn from $\mathcal{P}(\mathcal{X}^{d_{te}}, \mathcal{Y}^{d_{te}})$, where d_{te} represents the target domain. For our methodology, we decompose the model as $\mathcal{F} = w \circ f$, $f : \mathbf{x} \to h$ is a feature encoder and $w : h \to y$ is a classifier. $\mathcal{F}$ maps input images to the target label space.

SSDG pipeline: We overview FixMatch [32] which emerged as a top-performing SSL-based SSDG baseline in our empirical investigations (Fig. 2). So we chose it as an example baseline to explain our method. However, as shown in the experimental results (sec. 4), our method is model-agnostic and can be applied to several SSL-based SSDG baselines. FixMatch, which was originally proposed for SSL, combines two prior SSL techniques: pseudo-labeling, and consistency regularization. Pseudo-labeling uses a model to generate artificial labels for unlabeled data which are obtained from the arg max of the model's prediction probability. It retains only those artificial labels whose largest class probability falls above a predefined threshold. On the other hand, consistency regularization [3] leverages unlabeled data by enforcing that the predictions for perturbed views of the same image should be similar.

FixMatch[32] applies a weak augmentation and a strong augmentation to all images in a minibatch. The overall loss function consists of a supervised loss $\mathcal{L}_s$ and an unsupervised loss $\mathcal{L}_u$. $\mathcal{L}_s$ is a standard cross entropy (CE) loss applied on the weakly augmented labeled images. For unlabeled images, FixMatch first computes a pseudo-label corresponding to a weakly augmented version of the image and then uses this to enforce the cross-entropy loss against the model's output for a strongly augmented version of the same image, denoted as $\mathcal{L}_u$. This introduces a form of consistency regularization for the model. Overall the FixMatch loss is formulated as: $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u$.

Discussion: Although the SSL-based SSDG baselines show relatively better performance in SSDG settings, there is considerable room for further improvement across several DG benchmarks (Fig 2). An important challenge, faced by them, is the selection of accurate pseudo-labels in the presence of multiple domain shifts (see Fig. 3). This is further aggravated by the scarcity of limited labels, which increases the chances of model overfitting. To tackle the SSDG problem, we leverage the feature space and propose to enforce the prediction consistency between the two quantities: feature space posteriors which are derived from the same and different domains, and the pseudo-label from the model's output space (sec. 3.2). It implicitly facilitates the learning of more accurate pseudo-labels (see Fig. 3) by penalizing those examples whose prediction from the feature space does not align with the corresponding pseudo-label. Towards further improving the model's discriminative ability under the SSDG setting, we develop a semantics alignment loss (sec. 3.3) that attempts to regularize the semantic layout in the feature space by domain-aware similarity-guided

cohesion and repulsion of training examples. Fig. 4 visualizes the overall architecture of our method. In our method, described next, we ignore the labels of labeled set d^L and merge it with the unlabeled set d^U and so collectively we treat all the images as unlabeled afterward following [57] unless stated otherwise.

3.2. Feature-based Conformity

We believe that, in the presence of multiple source domains manifesting various domain shifts, for an unlabeled example belonging to an arbitrary domain, the posterior from the same and different source domains in feature space, should align with its pseudo-label produced from the model's output space. In this work, we coin this as feature-based conformity (FBC) that implicitly facilitates the model towards generating more accurate pseudo-labels. To achieve feature-based conformity, for an unlabeled image feature, the following key steps are designed (Fig. 4). First, we build domain-aware class prototypes in the feature space with the features of labeled images. Second, to compute similarities for a given unlabeled image feature, we choose prototypes having the same domain label (referred as same-domain class prototypes) and also select another set of prototypes having a different domain label chosen randomly (referred as different-domain class prototypes). This yields same-domain and different-domain similarities. Third, we convert the same-domain and different-domain similarities to posterior probabilities, and finally, these two probabilities are aligned with the pseudo-labels.

Pseudo label generation: For a given image $\mathbf{u}$ a pseudo label $\tilde{y}$ is generated by $\tilde{y} = \arg\max(\sigma(\mathcal{F}(\alpha(\mathbf{u}))))$ if $\max(\sigma(\mathcal{F}(\alpha(\mathbf{u})))) \geq \tau$, where σ is a softmax function, $\alpha(.)$ represents the weak augmentation operation [32] and τ is a threshold to retain the most confident predictions as described in [32].

Domain-aware class prototypes: We want to build domain-aware class prototypes which take into account the domain label d and the class c. Specifically, we obtain the class prototypes by averaging the image features of labeled raw examples without any augmentation, from the penultimate layer, corresponding to the class c and the domain d in consideration. Thus, the domain-aware class prototype $\mathbf{K}_c^d$ for class c in domain d is defined as:

$$\mathbf{K}_c^d = \frac{1}{|\mathcal{S}_c^d|} \sum_{i=1}^{|\mathcal{S}_c^d|} \mathcal{S}_c^d[i] \quad \text{where } \mathcal{S}_c^d = \{f(\mathbf{x}_i^d)|(y_i = c)\}_{i=1}^{|d^L|}. \tag{1}$$

Here $\mathbf{x}_i^d$ is a labeled image from domain d with class label c. Our domain-aware class prototypes are dynamically updated at the end of each epoch.

Feature similarity and alignment: Once we obtain the domain-aware class prototypes, we leverage them to obtain the same-domain and different-domain posterior prob-

abilities in the feature space for the unlabeled images. To be more specific, for an image $\mathbf{u}^{d_i}$, we obtain similarities with the same-domain class prototypes and randomly chosen different-domain class prototypes. We compute the similarity with same-domain class prototype as $\langle f(\alpha(\mathbf{u}^{d_i})), \mathbf{K}_c^{d_i} \rangle$, where $\langle, \rangle$ symbolizes the cosine similarity and α is the weak-augmentation function [32]. After computing the similarities with C same-domain class prototypes, we get a vector of same-domain similarities, denoted as $\mathbf{z}^{d_i} \in \mathbb{R}^C$. We then take the softmax of $\mathbf{z}^{d_i}$ to obtain the same-domain posterior probability $p^{d_i}(\mathbf{K}_c^{d_i}|f(\alpha(\mathbf{u}^{d_i})))$. Like-wise for obtaining similarity with the randomly chosen different-domain class prototype, we choose a random domain $d_j(\neq d_i)$ and compute $\langle f(\alpha(\mathbf{u}^{d_i})), \mathbf{K}_c^{d_j} \rangle$. After computing the similarities with C different-domain class prototypes, we get a vector of different-domain similarities, denoted as $\mathbf{z}^{d_j} \in \mathbb{R}^C$. We take the softmax of $\mathbf{z}^{d_j}$ to obtain the different-domain posterior probability $p^{d_j}(\mathbf{K}_c^{d_j}|f(\alpha(\mathbf{u}^{d_i})))$. Now, we leverage these same-domain probabilities $\mathbf{p}^{d_i}$ and the different-domain probabilities $\mathbf{p}^{d_j}$ from feature space and propose to align with the pseudo-label $\tilde{y}$ using two cross-entropy losses:

$$\mathcal{L}_{FBC} = -\underbrace{\sum_{i=1}^{C} \tilde{y}_i \log(p_i^{d_i})}_{\text{Same domain}} - \underbrace{\sum_{i=1}^{C} \tilde{y}_i \log(p_i^{d_j})}_{\text{Random different domain}} \tag{2}$$

The feature-based conformity serves as a regularizer by aligning the *the same-domain and different-domain probabilities* from the feature space with the corresponding pseudo-label.

3.3. Semantics Alignment Constraint

We intend to regularize the semantic structure in the feature space to capture semantically harmonious features and therefore improve the feature discriminativeness under domain shifts (Fig. 5 left). To this end, we formulate a semantics alignment (SA) loss that achieves domain-aware similarity-guided cohesion and repulsion of training examples (Fig. 4). For an input feature, it attempts to maximize the similarity to the assigned prototype class while minimising the similarity to the hard non-assigned prototypes in the same domain. Hard non-assigned prototypes refer to the class prototypes which are nearest neighbors to the assigned prototype. Also, we maximize the similarity with a prototype of randomly chosen different domain, whereby this prototype is having the same class label as the assigned prototype in the same domain. Formally, let $\phi^{d_i} = \max(\mathbf{z}^{d_i})$ and $\phi^{d_j} = \mathbf{z}^{d_j}[\arg\max(\mathbf{z}^{d_i})]$ be the similarity corresponding to the assigned prototype in the same domain and the similarity indexed (with the assigned prototype) from different domain for an input feature corresponding to $\mathbf{u}^{d_i}$. We also include the similarity to hard non-assigned prototypes

Model	PACS	OH	VLCS	DigitsDG	TerraInc.
ERM	59.8 ± 2.5	56.7 ± 0.8	68.0 ± 0.5	29.1 ± 2.9	23.5 ± 1.4
EntMin	64.2 ± 2.2	57.0 ± 0.8	66.2 ± 0.3	39.3 ± 2.8	26.6 ± 2.6
MeanTeacher	61.5 ± 1.4	55.9 ± 0.5	66.2 ± 0.4	38.8 ± 2.9	25.0 ± 2.8
FlexMatch	$72,7 \pm 1.2$	53.7 ± 0.7	56.2 ± 2.1	68.9 ± 1.2	26.4 ± 1.8
FreeMatch	74.0 ± 2.7	56.2 ± 0.2	61.6 ± 1.3	67.5 ± 2.4	30.1 ± 1.2
FixMatch	76.6 ± 1.2	57.8 ± 0.3	70.0 ± 2.1	66.4 ± 1.4	30.5 ± 2.2
StyleMatch	79.4 ± 0.9	59.7 ± 0.2	73.5 ± 0.6	65.9 ± 1.9	29.9 ± 2.8
FlexMatch + Ours	75.3 ± 1.2	55.8 ± 0.4	58.7 ± 1.0	$\mathbf{73.1 \pm 1.1}$	30.9 ± 1.0
FreeMatch + Ours	77.3 ± 1.7	58.0 ± 0.4	62.6 ± 1.3	72.2 ± 0.4	32.4 ± 2.9
FixMatch + Ours	78.2 ± 1.2	59.0 ± 0.4	72.2 ± 1.0	70.4 ± 1.4	$\mathbf{34.7 \pm 1.9}$
StyleMatch + Ours	$\mathbf{80.5 \pm 1.1}$	$\mathbf{60.3 \pm 0.6}$	$\mathbf{74.2 \pm 0.5}$	67.7 ± 1.7	32.5 ± 1.8

Table 1. SSDG accuracy (%) with 10 labels per class. (Average over 5 independent seeds is reported.)

in the same domain by first sorting same domain similarities $\mathbf{z}^{d_i}$ (in descending order) to get $\mathbf{v}^{d_i} = \text{Sort}(\mathbf{z}^{d_i})$ and then selecting and averaging the top-N after excluding the highest similarity as: $\Phi^{d_i} = \frac{1}{N-1} \sum_{n=2}^{N} \mathbf{v}^{d_i}[n]$. With these quantities, our semantics alignment loss is formulated as:

$$\mathcal{L}_{SA} = \underbrace{(1 - \phi^{d_i} + \Phi^{d_i})}_{\text{Same domain}} + \underbrace{(1 - \phi^{d_j})}_{\text{Random different domain}} \quad (3)$$

To reduce the loss, the ϕ^{d_i} and ϕ^{d_j} terms will be maximized while the Φ^{d_i} term will be minimized. The same-domain component of the loss tries to align the feature to the assigned prototype and also tries to repel the hard non-assigned prototypes in the same domain. The different-domain component of the loss attempts to align the feature to the assigned prototype in different-domain.

Our overall loss has four different loss terms: the supervised and unsupervised losses from the baseline, and the feature-based conformity loss and semantic alignment loss from our method, $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u + \mathcal{L}_{FBC} + \mathcal{L}_{SA}$. We provide the pseudo-code of our methodology in Algorithm. 1.

4. Experiments

Datasets, training and implementation details: We utilize PACS [23], OfficeHome [36], Digits [55], TerraIncognita [5] and VLCS [34] datasets which are widely used to report DG performance. For a detailed description of the datasets, refer to the supplementary materials. We conduct experiments under two settings; 10 labels and 5 labels per class while labeled images are selected randomly. The latter setting is more challenging due to the extreme scarcity of labeled data. Following [57], we randomly sample 16 labeled and 16 unlabeled images from each source domain to construct a minibatch. The labeled subset of minibatch is used to calculate the supervised loss while the (complete) minibatch, including both the labeled images (with ground truth labels dropped) and unlabeled images, are utilized to calculate all the unsupervised losses [57]. We use ImageNet[12] pre-trained ResNet-18[18] as the backbone architecture and a single-layer MLP head as the classifier.

Algorithm 1 Pseudo-code

1: **Input:** Labeled batch $d^L = \{(\mathbf{x}_b^d, y_b^d) : b \in (1, \cdots, B)\}$ and Unlabeled batch $d^U = \{(\mathbf{u}_b^d) : b \in (1, \cdots, B)\}\ \forall d \in d_1, d_2, \cdots d_{tr}$, Confidence threshold τ, E is total epochs, B is the Number of batches in an epoch, Model: $\mathcal{F} = w \circ f$
2: **for** epoch=1 to E **do**
3:      # Create domain-aware prototypes from the whole labeled images d^L, for each domain and each class.
4:      $\mathcal{S}_c^d = \{f(\mathbf{x}_i^d) | (y_i = c)\}_{i=1}^{|d^L|}$ ▷ Set of image features from the domain d and class c
5:      $\mathbf{K}_c^d = \frac{1}{|\mathcal{S}_c^d|} \sum_{i=1}^{|\mathcal{S}_c^d|} \mathcal{S}_c^d[i]$
6:      # Compute the supervised loss
7:      $\mathcal{L}_s = CE(\mathcal{F}(\alpha(\mathbf{x_b})), y_b))$
8:      # Concat the labeled images to unlabeled (without labels)
9:      $u^d = [\mathbf{u}^d, \mathbf{x}^d]\ \forall\ d_i \in d$
10:      **for** $u^{d_i} \in u^d$ **do**
11:          # Generate pseudo label
12:          **if** $\max(\sigma(\mathcal{F}(u^{d_i}))) \geq \tau$ **then**
13:              $\tilde{y} = \text{argmax}(\sigma(\mathcal{F}(u^{d_i})))$
14:              # Compute unsupervised loss
15:              $\mathcal{L}_u = CE(\sigma(\mathcal{F}(u^{d_i})), \tilde{y})$
16:              # Compute FBC loss
17:              $\mathbf{z}^{d_i} = \langle f(u^{d_i}), \mathbf{K}_c^{d_i} \rangle\ ; p^{d_i} = \sigma(\mathbf{z}_i^{d_i})$
18:              $\mathbf{z}^{d_j} = \langle f(u^{d_i}), \mathbf{K}_c^{d_j} \rangle\ ; p^{d_j} = \sigma(\mathbf{z}^{d_j})$
19:              $\mathcal{L}_{FBC} = CE(p^{d_i}, \tilde{y}) + CE(p^{d_j}, \tilde{y})$ ▷ (Eq: 2)
20:              # Compute SA loss
21:              $\phi^{d_i} = max(\mathbf{z}^{d_i})$
22:              $\Phi^{d_i} = \frac{1}{N-1} \sum_{n=2}^{N} \text{sort}(\mathbf{z}^{d_i})[n]$
23:              $\phi^{d_j} = \mathbf{z}^{d_j}[\text{argmax}(\mathbf{z}^{d_i})]$
24:              $\mathcal{L}_{SA} = (1 - \phi^{d_i} + \Phi^{d_i}) + 1 - \phi^{d_j}$ ▷ (Eq: 3)
25:          **end if**
26:          return $\mathcal{L} = \mathcal{L}_s + \mathcal{L}_u + \mathcal{L}_{FBC} + \mathcal{L}_{SA}$
27:      **end for**
28: **end for**

We use SGD as the optimizer with an initial learning rate of 0.003 for the backbone and 0.01 for the classifier, respectively. Both learning rates are decayed using the cosine annealing rule and we train all the models for 20 epochs on all datasets except TerraIncognita (trained for 10 epochs).

Model	PACS	OH	VLCS	DigitsDG	TerraInc.
ERM	51.2 ± 3.0	51.7 ± 0.6	67.2 ± 1.8	22.7 ± 1.0	22.9 ± 3.0
EntMin	55.9 ± 4.1	52.7 ± 0.5	66.5 ± 1.0	28.7 ± 1.3	21.4 ± 3.5
MeanTeacher	53.3 ± 4.0	50.9 ± 0.7	66.4 ± 1.0	28.5 ± 1.4	20.9 ± 2.9
FlexMatch	65.1 ± 2.5	48.8 ± 0.3	56.0 ± 2.8	59.0 ± 2.0	24.9 ± 4.3
FreeMatch	72.8 ± 1.2	53.8 ± 0.7	60.3 ± 1.7	58.9 ± 1.4	23.5 ± 2.7
FixMatch	73.4 ± 1.3	55.1 ± 0.5	69.9 ± 0.6	56.0 ± 2.2	28.9 ± 2.3
StyleMatch	78.4 ± 1.1	56.3 ± 0.3	72.5 ± 1.5	55.7 ± 1.6	28.7 ± 2.7
FlexMatch + Ours	71.0 ± 1.4	51.3 ± 0.1	58.0 ± 2.1	$\mathbf{66.2 \pm 0.6}$	28.8 ± 2.6
FreeMatch + Ours	73.7 ± 3.6	55.0 ± 0.2	62.1 ± 1.4	65.0 ± 1.5	26.5 ± 3.2
FixMatch + Ours	77.3 ± 1.1	55.8 ± 0.2	71.3 ± 0.7	62.0 ± 1.5	$\mathbf{33.2 \pm 2.0}$
StyleMatch + Ours	$\mathbf{79.3 \pm 0.9}$	$\mathbf{56.5 \pm 0.2}$	$\mathbf{72.9 \pm 0.7}$	58.7 ± 1.7	30.4 ± 3.7

Table 2. SSDG accuracy (%) with 5 labels per class. (Average over 5 independent seeds is reported.)

We set $n = \lceil \frac{|C|}{2} \rceil$, from sec. 3.3, in all experiments.

Evaluation protocol: We use the leave-one-domain out protocol for evaluation which has been used widely in domain generalization [17]. In this protocol, one domain is used as the target while the remaining domains are used as the source data to train the model. The target domain is unseen during the training phase and the model is evaluated on this unseen target domain. We report top-1 accuracy averaged over 5 independent trials.

Baselines: Since our method combines SSL with DG, we select state-of-the-art methods in both paradigms. We choose EntMin [16], MeanTeacher [33], FixMatch [32], FlexMatch [50] and FreeMatch [41] methods in SSL. StyleMatch[57] is selected as an SSDG baseline as it shows promising performance under 5 and 10 labels settings. Further, we chose ERM[35] in DG as it shows competitive performance against many existing DG methods [17]. For a detailed comparison with existing DG methods, see supplementary materials.

4.1. Results

We report the performance of our method when integrated into four different baselines: StyleMatch, FixMatch, FlexMatch, and FreeMatch on five challenging DG datasets (Table 1,2). **PACS:** Our method consistently improves the performance of all four baselines in both the 10 labels and 5 labels settings. We obtain 1.6%, and 3.9% average gains with our approach on top of the FixMatch baseline for 10 labels, and 5 labels setting respectively. Further, we attain 80.5% average accuracy on 10 class settings, when applied with StyleMatch baseline. **VLCS:** Our method is capable of delivering gain over all four baselines in both the 10 labels and 5 labels per class settings. Specifically, when combined with FixMatch baseline our method achieves more than 1.8% average gain on both settings. **OfficeHome:** Our approach improves the FixMatch baseline by 1.2% on average in 10 labels settings. In the same setting, our method, in tandem with StyleMatch, achieves the best figures across all domains with an average of 60.3%. **Digits-DG:** When combined with FixMatch baseline, our approach improves

Method	Average
Baseline [32]	73.4
Baseline + $\mathcal{L}_{\text{FBC(same-domain)}}$	76.0
Baseline + $\mathcal{L}_{\text{FBC(different-domain)}}$	74.9
Baseline + $\mathcal{L}_{\text{FBC}}$	76.7
Baseline + $\mathcal{L}_{\text{SA}}$	74.8
Baseline + $\mathcal{L}_{\text{FBC}}$ + $\mathcal{L}_{\text{SA(same-domain)}}$	77.0
Baseline + $\mathcal{L}_{\text{FBC}}$ + $\mathcal{L}_{\text{SA}}$ (Ours)	**77.3**

Table 3. Ablation study on PACS (5 labels per class).

FixMatch by a significant gain of 4.0% in 10 label setting and 6.0% in the 5-label setting. On average, we consistently improve all the baselines by approximately 5% with our approach. **Terra Incognita:** Our method provides gain over all four baselines on the Terra Incognita dataset which has a number of real-world distribution shifts such as illumination shifts, blur, shifts in the size of the region of interest (ROI), occlusions, camouflage and shifts in perspective. In both 5 and 10 labels settings, we obtain over 4.0% gain with our approach over the FixMatch baseline.

Note that, StyleMatch is unable to improve on its baseline FixMatch when there are no style distributions present in the source domains (on datasets such as Digits and TerraIncognita) where our proposed method demonstrates significant gains with over 4.0% improvement. As our approach is model agnostic, it can be seamlessly integrated with different SSL and SSDG baselines. Further, the improvements over the baselines are consistent across all the DG datasets under both 10 and 5 labels per class settings.

4.2. Ablation Study and Analysis

Impact of different components: In Table 3 we report the performance contribution of individual components in our approach. We note the following important trends: (1) each component is capable of improving the performance over the baseline, (2) the feature-based conformity component, which is the first component of our proposed method, provides a high gain of 3.3% while showing each same domain and different domain feature alignment is capable of improving the baseline's performances individually. (3)
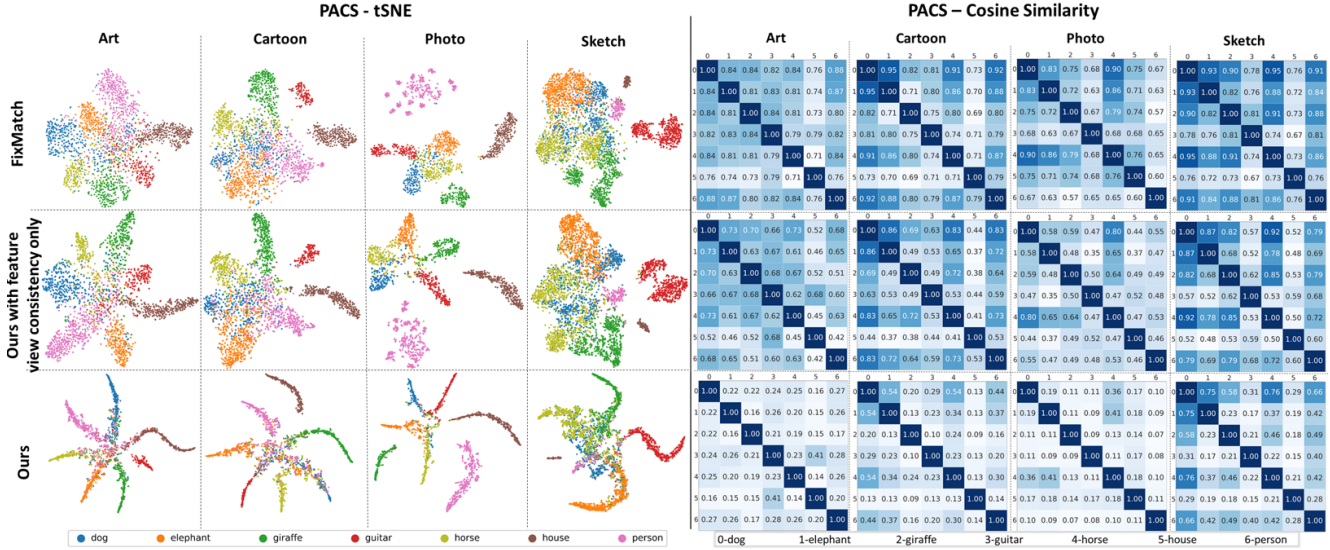
Figure 5. We visualize the feature space using tSNE (left), and the cosine similarity between the means of class-wise features (right) for PACS test domains of baseline (first row), ours with feature-based conformity only (second row), and ours with both feature-based conformity and semantic alignment loss (third row).

Domain Shift	Dataset	Method							
		FlexMatch	FreeMatch	FixMatch	StyleMatch	FlexMatch+Ours	FreeMatch+Ours	FixMatch+Ours	StyleMatch+Ours
Style Shifts	OH, PACS	56.9	63.3	64.1	64.25	61.2	64.4	66.6	**67.9**
Background Shifts	VLCS, Digits	57.6	59.6	62.9	64.1	62.1	63.6	**66.7**	65.8
Corruption Shift	Terra	24.9	23.5	28.9	28.7	28.8	26.5	**33.2**	30.4

Table 4. Accuracy(%) for different types of domain shifts in 5 labels per class setting. Corruption shifts include changes in illumination, changes in perspective, changes in ROI size, camouflage, occlusions, and blur which are often present in real-world distributions.

our proposed method provides the best gain of 3.9% when the feature-based conformity is coupled with the semantic alignment constraint.

Performance under various domain shifts: We evaluate the performance under various domain shifts (Table 4) including background shifts, style shifts, and corruption shifts which occur in real-world scenarios. Our approach provides visible gains over baselines under various domain shifts. Existing SSDG methods such as StyleMatch operate within the confines of the input pixel space and under the assumption that some style variations are present between source domains. Unlike our method which harnesses the information at feature space, such methods fail to generalize when they encounter real-world corruption shifts such as shifts in illumination, perspective, ROI size, blur, camouflage, and occlusions in the source domains.

Feature visualization: We visualize the feature space for PACS test domains with FixMatch as baseline (see Fig. 5 (left)). When FixMatch is coupled with our first component, feature-based conformity, we observe better class-wise discrimination over the baseline. When the feature-based conformity is supported by the semantic alignment constraint, we can observe well-separated and well-compact class-wise

clusters in feature space that help improve the classification performance. Also, our proposed method encourages classes to be orthogonal in the feature space (Fig. 5(right)).

5. Conclusion and Limitations

We approach the relatively unexplored problem of semi-supervised domain generalization (SSDG) and propose a new method, built with feature-based conformity and semantics alignment constraint modules, towards addressing the key challenges in SSDG. The feature-based conformity mechanism aligns the posterior distributions from two views, while the semantics alignment constraint further boosts the effectiveness of feature-based conformity by regularizing the semantic layout of feature space. Our approach is plug-and-play, parameter-free and model-agnostic, so it can be seamlessly integrated into different baselines as validated in our results. Extensive experiments on different challenging DG benchmarks show that our method delivers a consistent and notable gain over four recent baselines. An aspect to consider is that our approach may necessitate significant adjustments to accommodate semi-supervised single-source DG.

References

[1] Abulikemu Abuduweili, Xingjian Li, Humphrey Shi, Cheng-Zhong Xu, and Dejing Dou. Adaptive consistency regularization for semi-supervised transfer learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6923–6932, 2021. 3

[2] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*, 2019. 2

[3] Philip Bachman, Ouais Alsharif, and Doina Precup. Learning with pseudo-ensembles. *Advances in neural information processing systems*, 27, 2014. 4

[4] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. In *NeurIPS*, pages 998–1008, 2018. 1

[5] Sara Beery, Grant Van Horn, and Pietro Perona. Recognition in terra incognita. In *European Conference on Computer Vision*, pages 456–473, 2018. 2, 6

[6] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019. 1, 3

[7] Gilles Blanchard, Gyemin Lee, and Clayton Scott. Generalizing from several related classification tasks to a new unlabeled sample. *NeurIPS*, 24:2178–2186, 2011. 1

[8] Fabio M Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *CVPR*, pages 2229–2238, 2019. 1, 2

[9] Junbum Cha, Sanghyuk Chun, Kyungjae Lee, Han-Cheol Cho, Seunghyun Park, Yunsung Lee, and Sungrae Park. Swad: Domain generalization by seeking flat minima. *NeurIPS*, 34, 2021. 2

[10] Prithvijit Chattopadhyay, Yogesh Balaji, and Judy Hoffman. Learning to balance specificity and invariance for in and out of domain generalization. 2020. 2

[11] Hao Chen, Ran Tao, Yue Fan, Yidong Wang, Jindong Wang, Bernt Schiele, Xing Xie, Bhiksha Raj, and Marios Savvides. Softmatch: Addressing the quantity-quality tradeoff in semi-supervised learning. In *The Eleventh International Conference on Learning Representations*, 2023. 3

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 6

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 1

[14] Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. Domain generalization via model-agnostic learning of semantic features. In *NeurIPS*, pages 6450–6461, 2019. 2

[15] Muhammad Ghifary, W Bastiaan Kleijn, Mengjie Zhang, and David Balduzzi. Domain generalization for object recognition with multi-task autoencoders. In *ICCV*, 2015. 2

[16] Yves Grandvalet and Yoshua Bengio. Semi-supervised learning by entropy minimization. *Advances in neural information processing systems*, 17, 2004. 3, 7

[17] Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. *ArXiv*, abs/2007.01434, 2021. 3, 7

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1, 6

[19] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. Pmlr, 2018. 3

[20] Muhammad Haris Khan, Talha Zaidi, Salman Khan, and Fahad Shehbaz Khan. Mode-guided feature augmentation for domain generalization. In *Proc. Brit. Mach. Vis. Conf.*, 2021. 1, 2

[21] Daehee Kim, Youngjun Yoo, Seunghyun Park, Jinkyu Kim, and Jaekoo Lee. Selfreg: Self-supervised contrastive regularization for domain generalization. In *ICCV*, pages 9619–9628, 2021. 2

[22] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, page 896, 2013. 1, 3

[23] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *ICCV*, pages 5542–5550, 2017. 1, 2, 6

[24] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, and Timothy M Hospedales. Episodic training for domain generalization. *In ICCV*, 2019. 1, 2

[25] Haoliang Li, Sinno Jialin Pan, Shiqi Wang, and Alex C Kot. Domain generalization with adversarial feature learning. In *CVPR*, pages 5400–5409, 2018. 2

[26] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*, 41(8):1979–1993, 2018. 3

[27] Saeid Motiian, Marco Piccirilli, Donald A Adjeroh, and Gianfranco Doretto. Unified deep supervised domain adaptation and generalization. In *ICCV*, pages 5715–5725, 2017. 2

[28] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *ICML*, 2013. 1, 2

[29] Seonguk Seo, Yumin Suh, Dongwan Kim, Geeho Kim, Jongwoo Han, and Bohyung Han. Learning to optimize domain specific normalization for domain generalization. 2020. 2

[30] Shiv Shankar, Vihari Piratla, Soumen Chakrabarti, Siddhartha Chaudhuri, Preethi Jyothi, and Sunita Sarawagi. Generalizing across domains via cross-gradient training. *arXiv preprint arXiv:1804.10745*, 2018. 2

[31] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[32] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020. 1, 2, 3, 4, 5, 7

[33] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017. 1, 3, 7

[34] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR 2011*, pages 1521–1528. IEEE, 2011. 2, 6

[35] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 1999. 2, 7

[36] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, pages 5018–5027, 2017. 2, 6

[37] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. In *NeurIPS*, 2018. 2

[38] Haohan Wang, Xindi Wu, Pengcheng Yin, and Eric P Xing. High frequency component helps explain the generalization of convolutional neural networks. *arXiv preprint arXiv:1905.13545*, 2019. 3

[39] Ruiqi Wang, Lei Qi, Yinghuan Shi, and Yang Gao. Better pseudo-label: Joint domain-aware label and dual-classifier for semi-supervised domain generalization. *Pattern Recognition*, 133:108987, 2023. 3

[40] Shujun Wang, Lequan Yu, Caizi Li, Chi-Wing Fu, and Pheng-Ann Heng. Learning from extrinsic and intrinsic supervisions for domain generalization. 2020. 1, 2

[41] Yidong Wang, Hao Chen, Qiang Heng, Wenxin Hou, Yue Fan, , Zhen Wu, Jindong Wang, Marios Savvides, Takahiro Shinozaki, Bhiksha Raj, Bernt Schiele, and Xing Xie. Freematch: Self-adaptive thresholding for semi-supervised learning. 2023. 2, 3, 7

[42] Qizhe Xie, Zihang Dai, Eduard Hovy, Thang Luong, and Quoc Le. Unsupervised data augmentation for consistency training. *Advances in neural information processing systems*, 33:6256–6268, 2020. 3

[43] Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V Le. Self-training with noisy student improves imagenet classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10687–10698, 2020. 3

[44] Zheng Xu, Wen Li, Li Niu, and Dong Xu. Exploiting low-rank structure from latent domains for domain generalization. In *European Conference on Computer Vision*, pages 628–643. Springer, 2014. 2

[45] Shen Yan, Huan Song, Nanxiang Li, Lincan Zou, and Liu Ren. Improve unsupervised domain adaptation with mixup training. *arXiv preprint arXiv:2001.00677*, 2020. 2

[46] Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 2022. 1

[47] Minxiang Ye, Yifei Zhang, Shiqiang Zhu, Anhuan Xie, and Senwei Xiang. Semi-supervised domain generalization with graph-based classifier. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023. 3

[48] Junkun Yuan, Xu Ma, Defang Chen, Kun Kuang, Fei Wu, and Lanfen Lin. Label-efficient domain generalization via collaborative exploration and generalization. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2361–2370, 2022. 1, 3

[49] Anran Zhang, Yandan Yang, Jun Xu, Xianbin Cao, Xiantong Zhen, and Ling Shao. Latent domain generation for unsupervised domain adaptation object counting. *IEEE Transactions on Multimedia*, 2022. 1

[50] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34:18408–18419, 2021. 2, 3, 7

[51] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 3

[52] Kun Zhang, Bernhard Schölkopf, Krikamol Muandet, and Zhikun Wang. Domain adaptation under target and conditional shift. In *International conference on machine learning*, pages 819–827. PMLR, 2013. 1

[53] Youshan Zhang. A survey of unsupervised domain adaptation for visual recognition. *arXiv preprint arXiv:2112.06745*, 2021. 1

[54] Dengyong Zhou, Olivier Bousquet, Thomas Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. *Advances in neural information processing systems*, 16, 2003. 3

[55] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Deep domain-adversarial image generation for domain generalisation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 13025–13032, 2020. 2, 6

[56] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *European Conference on Computer Vision*, pages 561–578. Springer, 2020. 1, 2

[57] Kaiyang Zhou, Chen Change Loy, and Ziwei Liu. Semi-supervised domain generalization with stochastic stylematch. *International Journal of Computer Vision*, pages 1–11, 2023. 1, 2, 3, 5, 6, 7