

Learning Class and Domain Augmentations for Single-Source Open-Domain Generalization

Prathmesh Bele¹ Valay Bundele^{1*} Avigyan Bhattacharya^{1*} Ankit Jha¹ Gemma Roig² Biplab Banerjee¹

¹Indian Institute of Technology Bombay, India ²Goethe University Frankfurt, Germany

prathmeshbele.iitb@gmail.com, valaybundele@gmail.com, avigyanbhattacharya123@gmail.com
ankitjha16@gmail.com, roig@cs.uni-frankfurt.de, getbiplab@gmail.com

Abstract

Single-source open-domain generalization (SS-ODG) addresses the challenge of labeled source domains with supervision during training and unlabeled novel target domains during testing. The target domain includes both known classes from the source domain and samples from previously unseen classes. Existing techniques for SS-ODG primarily focus on calibrating source-domain classifiers to identify open samples in the target domain. However, these methods struggle with visually fine-grained open-closed data, often misclassifying open samples as closed-set classes. Moreover, relying solely on a single source domain restricts the model’s ability to generalize. To overcome these limitations, we propose a novel framework called SODG-NET that simultaneously synthesizes novel domains and generates pseudo-open samples using a learning-based objective, in contrast to the ad-hoc mixing strategies commonly found in the literature. Our approach enhances generalization by diversifying the styles of known class samples using a novel metric criterion and generates diverse pseudo-open samples to train a unified and confident multi-class classifier capable of handling both open and closed-set data. Extensive experimental evaluations conducted on multiple benchmarks consistently demonstrate the superior performance of SODG-NET compared to the literature.

1. Introduction

Deep learning models often face significant performance degradation when confronted with domain shift [3, 46], which occurs when the training and test data originate from different distributions. To address this issue, domain generalization (DG) has been proposed as a means to enable models to generalize to unknown target domains [45, 53]. In the classical DG setup, the assumption is made that multiple source domains are accessible during training, known as multi-source DG (Multi-DG). Notable advancements have



Figure 1. Differences between existing domain and sample interpolation methods (Mixstyle [55], Cumix [32]) and our proposed SODG-NET for style/sample synthesis. In both [55] and [32], α is sampled from the dist. $\mathcal{B}(\Gamma)$, where we propose a learning based scheme. \mathcal{F}_{ss} and \mathcal{F}_{fa} denote the proposed style synthesis and feature aggregation blocks, respectively.

been achieved in Multi-DG over the past decades, utilizing techniques such as domain alignment, meta-learning, domain augmentation, and self-supervision [2, 5–7, 9, 37].

However, collecting and labeling data from multiple domains can be expensive, making Multi-DG less practical in real-world scenarios. In contrast, Single-Source Domain Generalization (Single-DG) is a more challenging and realistic setting that has been less explored. In Single-DG, only one source domain is available for training, making it difficult for the model to learn domain-invariant information due to the absence of domain comparisons. As a result, the model is susceptible to overfitting the domain-specific signals present in the single source domain. In recent years, some efforts [8, 10, 47] have focused on domain augmentation for addressing Single-DG. However, it should be noted that accountable novel domain augmentation from a single source domain is a non-trivial problem.

While both Multi-DG and Single-DG are considered closed-set models, meaning that the source and target domains share the same set of categories, this assumption is often unrealistic in real-world applications. In practice, there is often a lack of prior information available for the target domain, leading to the possibility of encountering samples from previously unknown classes alongside the known class samples. This scenario has given rise to the problem of Open-Set Domain Generalization (ODG), which is applicable in both single and multi-DG settings. Despite the significant importance of ODG in various ap-

*equal contribution

plications such as self-driving cars, remote sensing, and medical imaging, there is a scarcity of ODG models in the existing literature [38, 56]. To this end, SS-ODG poses even greater challenges due to the scarcity of source domain styles and the presence of novel class samples in the target domain. It is worth noting that CROSSMATCH [56], the only existing single-source ODG model, tackles this issue by employing an adversarial augmentation strategy to synthesize out-of-distribution samples as pseudo-unknown data from the source domain. Furthermore, a majority voting-based classification strategy is utilized to identify open samples during testing.

Although CROSSMATCH demonstrates superior performance when combined with existing Single-DG methods, it does have a few potential bottlenecks: **i)** The generation of pseudo-open samples aims for a significant separation from the known class data, resulting in a coarse relationship between the known and pseudo-open samples. This approach may hinder the classification of fine-grained target open samples that have a closer relationship with the source domain data. **ii)** CROSSMATCH primarily focuses on the open-set classification task and lacks emphasis on improving the performance of closed-set classes. It does not support augmentation for the known-class samples and relies on existing Single-DG models for this purpose. **iii)** The multi-binary classifier used in CROSSMATCH neglects the class-level correlations, leading to performance degradations for fine-grained known and unknown samples.

The discussions highlight three significant research gaps in addressing SS-ODG: **i)** *How can we effectively synthesize diverse novel domains for the known classes when we only have a single source domain?* **ii)** *How can we generate pseudo-open samples that can be used to train a unified classifier capable of handling both open and closed classes in a confident fashion?* and **iii)** *Enhancing classification confidence within an ODG setting is a non-trivial task, especially considering the absence of prior knowledge about the target domain.* Existing style interpolation techniques [55] are not suitable for **(i)** due to the constraint of having only one source domain in SS-ODG, and style extrapolation methods [47, 54] may be challenging and costly to apply. For **(ii)**, traditional mix-up strategies [32, 50] based on stochastic BETA sampling may not always generate open samples reliably. *Given these challenges, there is a clear need for learning-based approaches to address (i)-(iii).*

Our proposed SODG-NET: In this paper, we present a comprehensive solution to address these gaps by introducing a novel end-to-end model called SODG-NET. Our approach involves synthesizing novel style primitives and generating representative open-set samples by separately leveraging domain and content knowledge extracted from the source domain images at the feature space, and further training a unified classifier for the closed and open classes while

optimizing the confidence of its predictive ability.

We recognize that the feature response statistics obtained from different layers of a vision encoder capture valuable domain-dependent information [28]. To capitalize on this, we propose a *style generator* module within SODG-NET. This module takes style information from two distinct images and produces a new style vector that maintains a significant margin from the input data’s style vectors (Fig. 1). This approach diverges from the conventional style interpolation methods [19, 31, 55]. While those methods simply blend the feature statistics of two input domain images through a basic convex combination, they lack the guarantee of producing distinct styles. Likewise, techniques involving style extrapolation [27] are restricted by the inherently uncertain nature of the problem, limiting their ability to generate a substantial variety of styles. Our solution to both these challenges lies in the innovative margin-based learning framework we introduce for style hallucination.

Furthermore, when generating the pseudo-open data, we depart from traditional ad-hoc mix-up techniques [51]. Instead, we introduce the concept of combining the feature embeddings of two distinct-class images using learnable feature weightings (Fig. 1). We then impose a constraint on the generated embedding, ensuring it is classified as belonging to the unknown class. By doing so, we establish a precise pseudo-open-closed boundary, facilitating improved identification of potential unknown-class samples. We identify our **major contributions** as follows,

- We address the SS-ODG problem through a data augmentation perspective, introducing SODG-NET, complementing the literature. SODG-NET can effectively learn to synthesize novel domain primitives and representative pseudo-open-class embeddings simultaneously.
- To ensure the distinctiveness of the generated domains both from the source domain and among themselves, we introduce a novel margin objective and a noisy-injected domain diversification strategy. Moreover, we put forth a weight learning approach that facilitates the fusion of feature embeddings from a pair of images, enabling the creation of a unified feature representation capturing the open space. Finally, we introduce intuitive objectives to increase the confidence of the predictions for both the closed and open classes, thus better handling visually alike samples.
- Thorough experiments were conducted on four benchmark datasets to evaluate SODG-NET’s performance for SS-ODG. On average, we observed a significant improvement of around 1-14% in the average h-score metric.

2. Related Works

Open-Set Domain Adaptation: OSDA is a related but different problem from ODG. OSDA considers a single labeled source domain and an unlabeled target domain, where the target domain contains additional samples from previ-

ously unknown classes. OSBP [36] adopts an adversarial technique to train representations that effectively distinguish unknown target samples. STA [30] employs a binary classifier to finely separate all target samples and applies weight adjustment to mitigate the negative influence of unfamiliar target samples. Additionally, there exist other approaches such as TIM [23], SHOT [29], JPOT [49], to name a few. In contrast, the SS-ODG problem assumes that source domain is only present during training, while unlabeled target domain appears only during inference. Hence, OSDA methods cannot be directly applied to SS-ODG.

Open Domain Generalization: On the other hand, ODG poses a more challenging and realistic problem, initially introduced by Shu et al. [38]. In this problem, the source domains, which are multiple in number, and the singular target domain have different label spaces. The objective is to develop a model that can learn from multiple sources of data and utilize that knowledge to classify new data points into known classes from the source data or into new, unknown classes. To tackle this problem, [38] proposed an approach that involves augmenting each source domain with missing class and domain knowledge through a novel DIRICHLET MIXUP and distilled soft-labeling technique. Subsequently, a meta-learning technique was employed over the augmented domains to acquire open-domain generalizable representations. While [38] is a multi-source ODG setup, CROSSMATCH [56] addresses a similar yet more complex problem, where there is only one source domain available for training but multiple unseen target domains for testing. [56] considered adversarial learning to hallucinate pseudo-open samples and deployed a multi-binary classifier, which is deemed to classify a given sample as belonging to a particular class or the open space.

While [56] serves as the closest existing framework to ours, there are significant differences between our approach and [56]. These differences primarily lie in two aspects: **i)** we recognize the challenges associated with extrapolating new domains using the unstable adversarial approach employed in [56]. Instead, we propose a margin-based objective that generates novel domain primitives, ensuring their distinguishability from the source domain, and **ii)** [56] relies on a voting-based classifier, which is susceptible to misclassification due to a lack of confidence. In contrast, we address this issue by directly learning an open-set classifier and introducing a learning-driven method for synthesizing representative pseudo-open data embeddings and explicitly enhancing the confidence of the predictions through novel losses. As a result, we are better equipped to tackle the SS-ODG problem compared to [56].

Augmentation Techniques: Previous works on domain or data augmentations, such as Goodfellow et al. [15], Kingma et al. [20], and Zhang et al. [51], have utilized variational autoencoders, GANs, or mixing strategies. Rahman et al.

[34] employed ComboGAN [1] to generate new samples and used various metrics to minimize the discrepancy between the generated and real data. In addition to generative models, mix-up techniques [32, 51] have been used to create new samples by interpolating between pairs of samples and their labels using BETA or DIRICHLET sampling. It has also been adopted in the areas of semi-supervised [39] and unsupervised learning [48] combined with pseudo-labeling and consistency regularization, respectively. [42] extends the mixup approach to the hidden states of deep neural networks while [44] applies it to the domain adaptation task in semantic segmentation.

Another approach by [54] involved training a GAN-based model using optimal transport to synthesize data from pseudo-novel domains. DLOW [14] bridged different domains by generating a continuous sequence of intermediate domains. MixStyle [55] randomly added batch-norm statistics of a pair of cross-domain samples.

Our approach to domain synthesis significantly departs from established domain interpolation techniques such as [31, 55]. Firstly, we introduce a novel approach to learning distinct styles by employing a margin-based objective tailored specifically for the SS-ODG problem. In contrast, the applicability of [31, 55] is more aligned with the Multi-DG problem and remains ad-hoc in nature. Furthermore, our incorporation of a diversification criterion ensures the creation of conspicuously distinctive generated domains. Along similar lines, our feature mixing strategy, facilitated by a neural network, outperforms traditional mix-up methods [32, 51]. This is due to our deliberate integration of learning-based criteria during the determination of mixing coefficients. The impact of these distinguishing features is evident in the results we present (Section 4).

3. Methodology

The SS-ODG problem revolves around a labeled source domain \mathcal{S} , which consists of training data $\mathcal{D}^s = \{x_i^s, y_i^s\}_{i=1}^{N_s}$. Here, $x^s \in \mathcal{X}^s$ represents input images sampled from the domain-specific distribution $\mathcal{P}(\mathcal{X}^s)$, and $y^s \in \mathcal{Y}^s$ represents the corresponding label set for \mathcal{S} . During testing, the model encounters unlabeled samples from a previously unknown target domain \mathcal{T} : $\mathcal{D}^t = \{x_j^t\}_{j=1}^{N_t}$. The label set \mathcal{Y}^t encompasses \mathcal{Y}^s , meaning $\mathcal{Y}^s \subset \mathcal{Y}^t$. Our objective is to learn a parameterized prediction model $f = g \circ h$, where g represents the generic feature extractor and h denotes the $\mathcal{C} + 1$ class classifier. Here, $|\mathcal{Y}^s| = \mathcal{C}$, indicating the number of classes in \mathcal{Y}^s .

3.1. Overview of SODG-NET

In this section, we present the architecture and working principles of SODG-NET, as depicted in Fig. 2. Our main objectives during the training of SODG-NET can be summarized as follows: **i)** Ensuring that the feature extractor

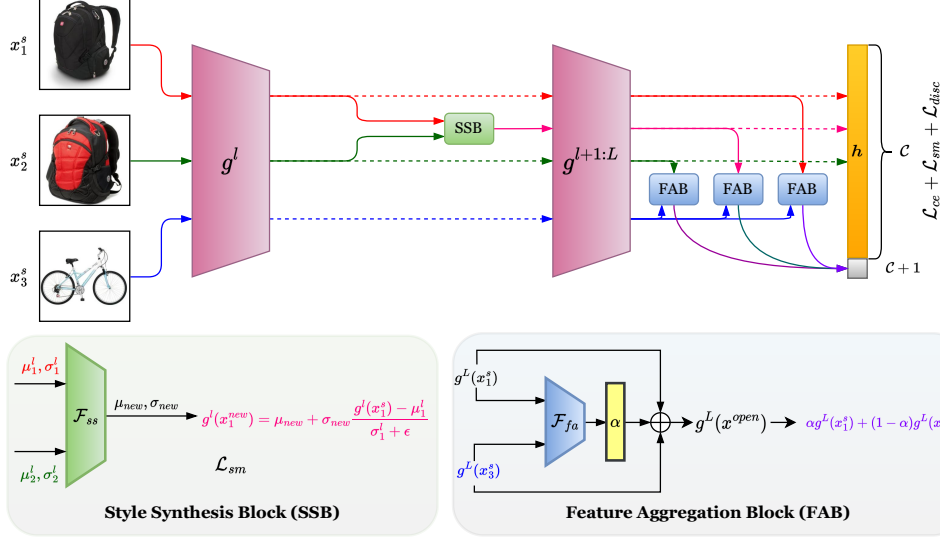


Figure 2. **The model architecture of SODG-NET.** The model consists of a feature extractor g and an open-closed classifier h . l denotes an encoder layer of l which consists of a total of L layers. \mathcal{F}_{ss} and \mathcal{F}_{fa} denote the proposed style synthesis and feature aggregation blocks. The figure also depicts the loss functions, which are explained in Section 3.

g is domain-generic, capable of extracting semantically coherent features from diverse visual domains. **ii)** Enabling the classifier h to simultaneously classify known-class samples from \mathcal{T} into one of the \mathcal{C} categories while consistently assigning a common label $\mathcal{C} + 1$ to all samples corresponding to the classes in $\mathcal{Y}^t - \mathcal{Y}^s$. **iii)** To improve the confidence of the predictions of h .

To ensure the domain independence of g , we propose augmenting \mathcal{D}^s with on-the-fly generated pseudo-domains and training g on the augmented domain set. However, synthesizing new domains from a single source domain is a challenging task. To address this, we introduce novel style synthesis blocks \mathcal{F}_{ss} within g . These blocks learn to generate new domain primitives by leveraging the domain properties of a pair of input images. It is important to note that the instance-wise feature statistics, specifically the mean (μ) and standard deviation (σ) from a CNN layer, capture domain artifacts, and different layers of g capture styles at different abstractions. Hence, it is possible to plug in \mathcal{F}_{ss} after the l^{th} encoder layer of g . We delve into the details of the style synthesis block and its functioning in Section 3.2.

Additionally, we aim to generate feature embeddings for the pseudo-open samples towards training h . To achieve this, we introduce a feature aggregation and novel feature generation network, denoted as \mathcal{F}_{fa} . This network combines the feature embeddings of two images with different class labels, producing a potential unknown class feature. The workings of \mathcal{F}_{fa} are elaborated further in Section 3.3.

3.2. Style synthesis block (SSB) to generate novel domains

The style synthesis block utilizes the style information from two input images to generate novel style character-

istics that are significantly different from the input image styles. For a given pair of images, x_1^s and x_2^s , belonging to the same class, we calculate the mean μ and standard deviation σ of their respective feature maps, $g^l(x_1^s)$ and $g^l(x_2^s)$, at the l^{th} encoder layer. Let us denote these as (μ_i^l, σ_i^l) , where $i \in 1, 2$. To obtain μ_{new}, σ_{new} , the mean and standard deviation vectors representing the synthesized style, we add noise sampled from a standard normal distribution $\mathbb{N}(0, 1)$ to each of the (μ_i^l, σ_i^l) and concatenate the mean and standard deviation vectors of the input images and pass them through \mathcal{F}_{ss}^l , denoting the style synthesis module coupled with g^l . *Our idea is to generate novel styles for each class separately, hence, we consider (x_1^s, x_2^s) from the same class.* The process can be expressed as follows:

$$\mu_{new}, \sigma_{new} \leftarrow \mathcal{F}_{ss}^l([\mu_1^l + \delta_1; \sigma_1^l + \delta_2; \mu_2^l + \delta_3; \sigma_2^l + \delta_4])$$

With the addition of the randomly sampled δ to the input vectors, we ensure generative modeling of the styles. Furthermore, to guarantee the distinctiveness of the generated $[\mu_{new}, \sigma_{new}]$ from the input $[\mu_1, \sigma_1, \mu_2, \sigma_2]$, we introduce a margin loss criterion, denoted as \mathcal{L}_{sm} (detailed in Eq. 4). In this regard, we begin by applying instance normalization to the feature encoding $g^l(x_1^s)$. This step is crucial as it helps remove the original style characteristics present in the feature encoding of x_1^s [11, 17, 40]. After the instance normalization step, we utilize the newly generated mean and standard deviation vectors, μ_{new} and σ_{new} , to modify the style of $g^l(x_1^s)$. This modification results in a sample with a novel style, distinct from the input images' styles but with identical semantic properties. ϵ is a small constant in Eq. 1 to ensure numerical stability.

$$g^l(x_1^{new}) = \mu_{new} + \sigma_{new} \left(\frac{g^l(x_1^s) - \mu_1}{\sigma_1 + \epsilon} \right) \quad (1)$$

3.3. Feature aggregation block (FAB) for open sample generation

The feature aggregation block, \mathcal{F}_{fa} , in contrast to \mathcal{F}_{ss} , aims to synthesize samples representing out-of-distribution arbitrary classes by mixing the content features of samples from two different classes in \mathcal{Y}^s . Unlike mix-up based strategies [32], our approach learns the mixing coefficient α using \mathcal{F}_{fa} , with the intention of classifying the synthesized feature as belonging to the $\mathcal{C} + 1$ category directly. To achieve this, we utilize the output feature maps from the last convolutional block of g , denoted as $g^L(x_1^s)$ and $g^L(x_3^s)$, corresponding to images x_1^s and x_3^s with different class labels. By concatenating these feature maps and passing them through \mathcal{F}_{fa} , we obtain a weighting vector α . This vector determines the contribution of each image’s feature maps to generate a novel class sample, computed as $g^L(x^{open}) = \alpha \odot g^L(x_1^s) + (1 - \alpha) \odot g^L(x_3^s)$. The mixing process, utilizing higher-level feature maps that capture abstract content information, allows the generation of samples representing novel classes by leveraging diverse input samples’ content information. Additionally, to enforce style variations in the generated open samples, we use Eq. 1 to update their style. This ensures that the synthesized samples exhibit distinct style characteristics, further enhancing the diversity and realism of the generated open samples.

3.4. Loss functions, training, and inference

A combination of multiple losses is employed to train both g and h . The primary objective, considering the features of the closed and generated open samples, is the cross-entropy loss, which can be expressed as follows:

$$\mathcal{L}_{ce} = \mathbb{E}_{\mathcal{P}(\mathcal{X}^s \cup \{x^{open}\}, \mathcal{Y}^s \cup \mathcal{C} + 1)} - \sum_{k=1}^{\mathcal{C}+1} y_{[k]} \log(h(g(x))_{[k]}) \quad (2)$$

$y \in \mathbb{R}^{\mathcal{C}+1}$ is the augmented label representation combining the closed and open classes simultaneously.

The cross-entropy loss \mathcal{L}_{ce} can struggle with nuanced open and closed samples, resulting in less confidence for open sample predictions. To overcome this, we introduce an extra loss term targeting the entropy of predictions $h(g(x^{open}))$ for generated open samples $x^{open} \in \mathcal{X}^{open}$. This term aims to increase posterior probabilities for the synthesized open samples in the $\mathcal{C} + 1^{th}$ class, while decreasing probabilities for closed class indices $1 : \mathcal{C}$, ultimately enhancing confidence in open sample predictions. It’s worth noting that \mathcal{L}_{ce} already assigns these samples to the $\mathcal{C} + 1^{th}$ index, and our approach further bolsters the confidence of this classification. Furthermore, we introduce

a margin loss for class-posterior probabilities in closed-set samples. This enhances prediction certainty and prevents finely distinguished closed samples from being misclassified as open-class data. Specifically, our aim is to widen the gap between the highest closed-set probability within indices $1 : \mathcal{C}$ and the open-class probability (indexed as $\mathcal{C} + 1$) for samples in \mathcal{D}_s . These loss functions collectively constitute the discriminability objective (\mathcal{L}_{disc}), defined as follows, where $h(g(x))_{\mathcal{C}+1}$ and $h(g(x))_{top}$ represent posterior probabilities of the open class and the highest closed-set class, respectively, for a given input x .

$$\mathcal{L}_{disc} = \mathbb{E}_{\mathcal{P}(\mathcal{X}^{open}, \{\mathcal{Y}^s \cup \mathcal{C} + 1\})} - h(g(x)) \log(h(g(x))) - \mathbb{E}_{\mathcal{P}(\mathcal{X}^s, \mathcal{Y}^s)} |h(g(x^s))_{\mathcal{C}+1} - h(g(x^s))_{top}|_1 \quad (3)$$

Finally, we introduce a novel margin loss (\mathcal{L}_{sm}) to ensure distinctiveness between the synthesized style properties $(\mu_{new}, \sigma_{new})$ and the input images’ styles (x_1^s, x_2^s) . This objective encompasses separate losses for μ and σ , aiming to contain the generated $(\mu_{new}, \sigma_{new})$ within predefined bounds (a, b) concerning the input styles. To achieve this, we consider two margins, striking a balance between dissimilarity from the inputs and preserving the image’s semantic integrity. In Eq. 4, $\|\mu_{new} - \mu_i\|_2$ is denoted as $d(\mu_{new}, \mu_i)$ for $i \in 1, 2$. This enforces the required distinction between the synthesized and input style properties.

$$\mathcal{L}_{sm}(\mu_{new}, \mu_i, a, b) = \begin{cases} 0 & \text{if } d(\mu_{new}, \mu_i) \in [a, b] \\ a - d(\mu_{new}, \mu_i) & \text{if } d(\mu_{new}, \mu_i) < a \\ d(\mu_{new}, \mu_i) - b & \text{if } d(\mu_{new}, \mu_i) > b \end{cases} \quad (4)$$

Total loss and training: To train the entire network in an end-to-end fashion, a weighted combination of losses is utilized, resulting in the total loss $\mathcal{L}_{\mathcal{T}}$:

$$\mathcal{L}_{total} = w_{ce} \mathcal{L}_{ce} + w_{disc} \mathcal{L}_{disc} + w_{sm} \mathcal{L}_{sm} \quad (5)$$

where w_{ce}, w_{disc}, w_{sm} are the weights corresponding to the loss components and all are set to 1 in our experiments.

Testing: At test time, the image is fed to the prediction model $g \circ h$ and the class label with the highest softmax probability score is predicted.

4. Experiments

Datasets: Following the footsteps of [56], we conducted our experiments on four datasets: (1) **Office-31** [35], (2) **Digits** [13, 18, 24, 33], (3) **Office-Home** [41], and (4) **PACS** [26], respectively. The dataset details are mentioned in the SUPPLEMENTARY.

Implementation details: While experimenting on *Office31* and *Digits* datasets, we use Amazon and MNIST as source

domains while the rest are used as target domains, respectively. In the case of the rest of the two datasets, every domain serves as a source domain once, and the rest are treated as target domains. For all the datasets except Digits, we employ a RESNET-18 pre-trained on ImageNet [16] as the backbone network, whereas the LENET [25] is considered for Digits. Intermediate feature maps for the domain synthesis task are extracted after the fifth convolutional block of the network. \mathcal{F}_{ss} is realized through a fully connected neural network with an input layer of size 256. The input layer takes the concatenated values of $\mu_1, \sigma_1, \mu_2, \sigma_2$, with added noise from a standard normal distribution $\mathbb{N}(0, 1)$ to ensure style diversification. The input layer is followed by a hidden layer with 192 nodes and an output layer of dimension 128. The first 64 dimensions represent μ_{new} while the remaining half denotes σ_{new} . The ReLU activation function follows each of these layers. On the other hand, \mathcal{F}_{fa} works as follows: First, it takes the input of concatenated features and projects it into a layer of dimension 512. This layer is followed by a ReLU activation and a Batch-Norm layer. The output of this layer is then passed to an output layer with a dimension of 512 and a Sigmoid activation.

The data is loaded in the form of triplets, where each consists of two images belonging to the same class and one randomly chosen image from a different class. The images belonging to the *Digits* dataset are resized to the dimension of 28×28 , while for all other datasets, the images are resized to 128×128 . Normalization is performed using the ImageNet mean values of [0.485, 0.456, 0.406] and standard deviation values of [0.229, 0.224, 0.225] for the R-G-B channels, respectively. The data is loaded with a batch size of 160, and the SGD optimizer is used with a learning rate of 0.001 and a momentum of 0.9. During the training procedure, the triplets are shuffled every five epochs. The margin between the mean and standard deviation is kept within the ranges of [1.5, 3.5] and [0.1, 2], respectively. We provide the detailed architecture of our model, which had a total of 12M learnable parameters.

Baselines and evaluation metrics: In our study, we conducted a comparative analysis by benchmarking our results against various approaches. Firstly, we considered the Empirical Risk Minimization (ERM) method [21] as a baseline without incorporating Single-DG. Additionally, we evaluated two state-of-the-art Single-DG methods: Adversarial Data Augmentation (ADA) [43] and Maximum-Entropy Adversarial Data Augmentation (MEADA) [52]. Furthermore, we assessed the performance of ERM, ADA, and MEADA after integrating with CROSSMATCH (represented as “+CM”) [56]. To establish the baseline performance, we employed Open-Set Domain Adaptation by Back-propagation (OSDAP) [36], a prominent technique in open-set domain adaptation, as well as OpenMax [4], a method for open-set recognition.

Table 1. Results (% Accuracy) on **Office31** and **Digits** Dataset.

Method	Office31		Digits	
	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>
OSDAP [36]	76.51	77.68	41.42	40.46
OpenMax [4]	18.19	16.71	42.38	40.67
ERM [21]	79.82	40.69	49.17	17.97
ERM+CM [56]	78.3	51.14	49.07	40.15
ADA [43]	80.13	38.65	50.22	20.14
ADA+CM [56]	78.61	48.5	49.71	39.93
MEADA [52]	80.26	38.55	52.98	30.37
MEADA+CM [56]	78.98	54.69	51.27	38.70
SODG-NET	79.02	78.62	55.66	53.87
SODG-NET $-\mathcal{L}_{disc}$	78.34	78.06	53.64	52.82
SODG-NET $-\mathcal{L}_{disc} - \mathcal{L}_{sm}$	75.88	75.44	53.26	49.24
SODG-NET $-\mathcal{L}_{sm}$	68.66	68.49	47.80	46.38

During our experiments, we utilized several evaluation metrics, including overall accuracy (*acc*) and the h-score (*hs*) [12]. The h-score represents the harmonic mean of the accuracy values for known and unknown classes, providing a comprehensive assessment by assigning a high score only when both accuracies are significantly high. We present the average performance, computed over three seeds, using the leave-one-domain-out approach. In this approach, we fix each domain as the source while evaluating the average performance over the remaining domains treated as targets.

5. Results and Discussions

Office31 and Digits: Our experimental results on the *Office31* and *Digits* datasets, as presented in Table 1, demonstrate the efficacy of our proposed architecture when compared to existing approaches in the field. In terms of the *Office31* dataset, our method achieves performance on par with the highest accuracy (*acc*) reported by other methods. Notably, our approach exhibits an improvement of 0.94% in the highest score (*hs*) over the OSDAP. Furthermore, our method surpasses MEADA by a considerable margin of 40.07% in terms of *hs*. Regarding the *Digits* dataset, the existing literature indicates that MEADA and OpenMax achieve the highest *acc* and *hs* scores, respectively. However, our SODG-NET outperforms these methods with a margin of 2.68% in *acc* and 13.2% in the *hs* score, respectively. In Fig. 3, it is evident that the source domain classes maintain a notable level of clustering even when faced with test data from previously unseen target domains. Additionally, the appearance of a distinct red cluster at the centre of the plot for the generated pseudo-classes confirms the effectiveness of our learning-driven open-sample synthesis scheme, which ensures that the generated open samples should lie outside the closed-space support. We can see that the network trained on closed samples generated using the MIXSTYLE for SS-ODG underperforms our SODG-NET. The clusters in right t-SNE in Fig. 3 are not as dense as the ones in the left t-SNE.

Office-Home: Table 2 showcases the results obtained on the challenging *Office-Home* dataset. This dataset contains a larger number of classes compared to *Office31* and *Dig-*

Table 2. Results (% Accuracy) on **Office-Home** Dataset for different source domains.

Method	Art		Clipart		Product		Real-World		Average	
	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>
OSDAP [36]	45.61	52.35	52.78	58.82	41.45	47.95	53.51	58.40	48.34	54.38
OpenMax [4]	22.42	30.64	22.67	29.51	15.10	16.65	25.54	33.07	21.43	27.47
ERM [21]	65.00	31.07	64.12	35.78	60.53	36.33	66.59	33.92	64.06	34.28
ERM+CM [56]	65.49	52.85	63.37	50.51	58.03	47.25	67.75	52.60	63.66	50.80
ADA [43]	68.29	32.94	65.10	42.09	60.52	34.72	67.04	34.86	65.24	36.15
ADA+CM [56]	66.30	46.68	62.64	49.31	58.72	47.47	66.82	50.47	63.62	48.48
MEADA [52]	68.31	33.29	65.25	42.05	60.43	35.68	67.04	34.65	65.01	36.42
MEADA+CM [56]	65.85	53.22	62.90	48.87	58.36	45.34	67.10	50.77	63.55	49.55
SODG-NET	65.31	57.65	69.20	63.85	67.04	61.24	68.18	62.89	67.43	61.41
SODG-NET $-\mathcal{L}_{disc}$	65.03	56.49	64.81	62.62	65.99	60.92	67.82	62.32	65.91	60.59
SODG-NET $-\mathcal{L}_{disc} - \mathcal{L}_{sm}$	58.67	55.77	60.66	60.82	64.08	59.21	62.65	61.19	61.51	59.25

Table 3. Results (% Accuracy) on **PACS** Dataset for different source domains.

Method	Art Painting		Cartoon		Sketch		Photo		Average	
	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>
OSDAP [36]	53.30	46.58	43.73	38.81	42.05	41.03	30.81	32.89	42.47	39.83
OpenMax [4]	52.59	53.60	31.71	25.23	29.85	19.87	27.60	19.47	35.44	29.54
ERM [21]	62.24	38.90	55.34	40.96	39.19	28.89	38.32	35.74	48.77	36.12
ERM+CM [56]	63.52	44.90	57.60	48.31	38.53	30.43	42.52	41.60	50.54	41.31
ADA [43]	62.48	39.02	56.43	41.55	39.03	26.93	40.28	38.13	49.56	36.41
ADA+CM [56]	64.26	42.40	60.41	51.81	42.48	35.18	43.97	42.76	52.78	43.04
MEADA [52]	62.43	38.85	56.10	41.34	38.89	26.43	39.88	38.24	49.33	36.22
MEADA+CM [56]	62.63	41.88	60.03	51.36	41.51	35.76	43.50	41.60	51.92	42.65
SODG-NET	57.02	57.44	56.01	56.62	58.36	58.40	46.27	43.60	54.41	54.02
SODG-NET $-\mathcal{L}_{disc}$	56.41	56.35	55.41	55.93	58.12	57.30	43.07	43.45	53.25	53.26
SODG-NET $-\mathcal{L}_{disc} - \mathcal{L}_{sm}$	55.33	55.96	50.66	51.20	57.56	55.70	40.52	40.06	51.02	50.73

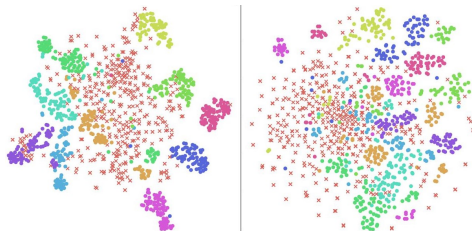


Figure 3. t-SNEs on **Office31** dataset. Left t-SNE corresponds to the results obtained from SODG-NET and the right one to the network where the SSB is replaced with MIXSTYLE. The red crosses ‘ \times ’ in the central region represent the unknown class, while the clusters belong to the ten known classes from the source and the target domain, showing the overlap in the clusters.

its and exhibits substantial distributional shifts between domains. Our proposed approach outperforms existing state-of-the-art methods in terms of *hs*, regardless of the chosen source domain. When considering the average performance across all domains, the ADA model achieves the highest *acc*, while OSDAP achieves the highest *hs* in the literature. However, we achieve a remarkable improvement of 2.19% over the best *acc* achieved by previous methods and surpass the best *hs* by an impressive margin of 7.03%.

PACS: The experimental results for the **PACS** dataset are summarized in Table 3. Similar to our performance on the **Office-Home** dataset, our approach consistently outperforms the existing literature in terms of *hs*, regardless of the chosen source domain. When considering the average performances, our method achieves the highest accuracy (*acc*) and *hs* scores, beating the next best technique combining ADA with CROSSMATCH by 1.63% and 10.98%, respectively, in the *acc* and *hs* metrics. The higher *hs* is a clear

indication of a better balance between the acc_k and acc_u .

Furthermore, we show the sensitivity to the loss terms \mathcal{L}_{sm} and \mathcal{L}_{disc} for all the datasets, highlighting the importance of both the loss terms. We note that removing \mathcal{L}_{sm} signifies an unbounded space for the generated styles. (see SUPPLEMENTARY for acc_k and acc_u)

5.1. Ablation Study

Diversity of the generated styles and open samples: To quantitatively assess the diversity of the generated styles and open samples compared to the source domains, we compute the average cosine distance between the concatenated (μ, σ) of the source samples and the generated $(\mu_{new}, \sigma_{new})$. Similarly, we calculate this metric between the sets x^s and x_{open} . In the case of the **Office31** dataset, we observe a mean cosine distance of 0.58 and 0.72 between the original and synthesized styles’ mean and standard deviation, indicating significant separation in the embedding space. Likewise, the mean cosine distance between the closed and open features is 0.55, further emphasizing their distinct placement.

Varying the number of known classes: We conducted experiments to evaluate the effectiveness of our method on the Real-World domain of the **Office-Home** dataset, considering different \mathcal{Y}^s . The $|\mathcal{Y}^s|$ was varied from 10 to 60, incrementing in steps of 10. Fig 4 illustrates the variation of *acc* and *hs* for the state-of-the-art MEADA method, the CROSSMATCH method applied over MEADA (MEADA+CM), and our proposed method. We observe that in our results, the *hs* and *acc* values are consistently close to each other, which again proves that our method has much more balance

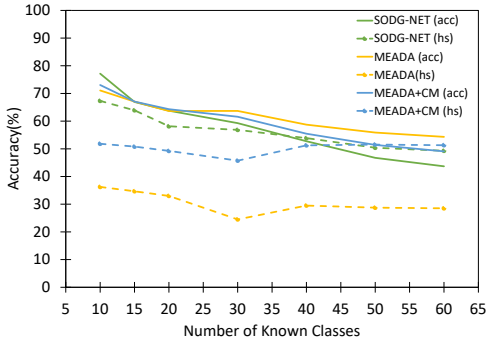


Figure 4. **Accuracy vs Number of Known Classes.** While performing similar to the MEADA and MEADA+CM in terms of *acc*, our SODG-NET outperforms the two methods for *hs*.

Table 4. Ablation on different style augmentation techniques.

Method	Office31		Digits	
	<i>acc</i>	<i>hs</i>	<i>acc</i>	<i>hs</i>
Jin <i>et al.</i> [19]	64.56	58.43	54.98	40.33
Kundu <i>et al.</i> [22]	58.98	54.36	51.27	46.70
Luo <i>et al.</i> [31]	64.58	64.56	54.05	51.37
SODG-NET	79.02	78.62	55.66	53.87

in acc_k and acc_u as compared to the other two. In particular, our *hs* metric outperforms the other two methods when the number of known classes is smaller. As the number of known classes surpasses 50, our method yields comparable yet better results to MEADA+CM in terms of *hs*.

Comparison of the style synthesis approaches: In Table 4, we conduct a comparative analysis of the style synthesis block within SODG-NET against three techniques from the DG literature [19, 22, 31]. These methods rely on feature statistics interpolation to create new domains. The results unequivocally demonstrate the superior performance of SODG-NET across both closed and open-set scenarios. This substantial advantage can be attributed to our metric-driven approach that fosters diversified style generation, profoundly enhancing the model’s overall generalizability.

\mathcal{F}_{f_a} and mixup based open sample generation methods:

In order to generate pseudo-samples for open-set recognition, we employed \mathcal{F}_{f_a} that learns a weight to linearly combine the content features obtained from the backbone network. As an alternative approach, we conducted experiments on *Office31* dataset using the following samples as unknown classes: (1) Taking two images from different classes, cropping them in half, and joining them together [32], (2) Calculating the mean of corresponding pixel values in images from two different classes, and (3) Randomly replacing a 30×30 patch from one image with a patch of the same size from an image of another class. When comparing these alternative methods with our proposed approach of generating open-set representations, we observed that the first method resulted in a performance lag of 4.22%

and 6.04% for accuracy and harmonic mean score, respectively. The other two methods were unable to generalize effectively on known classes, consequently failing to achieve consistent training accuracy.

Table 5. Effects of variations in the interval $[a, b]$ on *acc*.

		Standard Deviation				
		[0.1,1]	[1,2]	[2,3]	[3,4]	[4,5]
Mean	[0.1,1]	77.52	75.89	76.7	78.61	76.98
	[1,2]	76.7	76.7	76.16	76.84	77.38
	[2,3]	75.34	76.98	77.38	77.66	75.07
	[3,4]	75.48	77.25	76.84	74.48	75.07
	[4,5]	76.84	77.25	76.43	75.07	75.75

Table 6. Effects of variations in the interval $[a, b]$ on *hs*.

		Standard Deviation				
		[0.1,1]	[1,2]	[2,3]	[3,4]	[4,5]
Mean	[0.1,1]	76.71	75.45	75.99	77.17	76.9
	[1,2]	75.94	76.23	75.87	75.75	76.98
	[2,3]	75.24	76.62	76.6	77.46	74.78
	[3,4]	75.1	76.06	73.94	73.83	74.97
	[4,5]	76.47	75.25	76.37	74.49	74.5

Effects of a and b in \mathcal{L}_{sm} : In order to produce different styles, we induce a lower bound denoted as a to ensure a minimum distance between the predicted statistical features and the original style primitives. Simultaneously, an upper bound denoted as b is applied to prevent significant alterations in the distribution that could result in changes to the semantic information of the feature map. The experimentation involved evaluating five different ranges of $[a, b]$ for both the mean and standard deviation, resulting in a total of 25 combinations. The experimental results, presented in tables 5 and 6, provide insights into the *acc* and *hs* achieved on the *Office31* dataset.

6. Takeaways

This paper addresses the challenge of single-source ODG by proposing an architecture called SODG-NET. Our goal is to develop a model that can effectively generalize to diverse target domains using data from a single source domain, while also accommodating unknown classes in the target domain. To achieve this, we propose novel losses that enable learning to synthesize statistical style features different from the source domain, facilitating effective generalization for known classes. Additionally, SODG-NET generates representations for unknown classes by combining content features from images of two distinct classes using a learnable feature weighting and a classification constraint. Extensive experimentation on various datasets validates that our SODG-NET significantly improves generalization to different target domains using only a single source domain, while effectively detecting unknown classes. Future directions include applying this model to other application domains such as person re-identification.

References

- [1] Asha Anoopseh, Eirikur Agustsson, Radu Timofte, and Luc Van Gool. Combogan: Unrestrained scalability for image domain translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 783–790, 2018. [3](#)
- [2] Yogesh Balaji, Swami Sankaranarayanan, and Rama Chellappa. Metareg: Towards domain generalization using meta-regularization. *Advances in neural information processing systems*, 31, 2018. [1](#)
- [3] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79:151–175, 2010. [1](#)
- [4] Abhijit Bendale and Terrance E. Boult. Towards open set deep networks. *CoRR*, abs/1511.06233, 2015. [6](#), [7](#)
- [5] Silvia Bucci, Antonio D’Innocente, Yujun Liao, Fabio M Carlucci, Barbara Caputo, and Tatiana Tommasi. Self-supervised learning across domains. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5516–5528, 2021. [1](#)
- [6] Fabio M Carlucci, Antonio D’Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2229–2238, 2019. [1](#)
- [7] Fabio Maria Carlucci, Paolo Russo, Tatiana Tommasi, and Barbara Caputo. Hallucinating agnostic images to generalize across domains. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3227–3234. IEEE, 2019. [1](#)
- [8] Ilke Cugu, Massimiliano Mancini, Yanbei Chen, and Zeynep Akata. Attention consistency on visual corruptions for single-source domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4165–4174, 2022. [1](#)
- [9] Qi Dou, Daniel Coelho de Castro, Konstantinos Kamnitsas, and Ben Glocker. Domain generalization via model-agnostic learning of semantic features. *Advances in Neural Information Processing Systems*, 32, 2019. [1](#)
- [10] Thomas Duboudin, Emmanuel Dellandréa, Corentin Abgrall, Gilles Hénaff, and Liming Chen. Encouraging intra-class diversity through a reverse contrastive loss for single-source domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 51–60, 2021. [1](#)
- [11] Vincent Dumoulin, Jonathon Shlens, and Manjunath Kudlur. A learned representation for artistic style. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. [4](#)
- [12] Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. *Learning to Detect Open Classes for Universal Domain Adaptation*, pages 567–583. 11 2020. [6](#)
- [13] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1180–1189, Lille, France, 07–09 Jul 2015. PMLR. [5](#)
- [14] Rui Gong, Wen Li, Yuhua Chen, and Luc Van Gool. Dlow: Domain flow for adaptation and generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2477–2486, 2019. [3](#)
- [15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. [3](#)
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. [6](#)
- [17] Xun Huang and Serge J. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22–29, 2017*, pages 1510–1519. IEEE Computer Society, 2017. [4](#)
- [18] J.J. Hull. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):550–554, 1994. [5](#)
- [19] Xin Jin, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. Style normalization and restitution for domain generalization and adaptation. *IEEE Transactions on Multimedia*, 24:3636–3651, 2021. [2](#), [8](#)
- [20] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. [3](#)
- [21] Vladimir Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: École d’Été de Probabilités de Saint-Flour XXXVIII-2008*, volume 2033. 01 2011. [6](#), [7](#)
- [22] Jogendra Nath Kundu, Akshay Kulkarni, Amit Singh, Varun Jampani, and R Venkatesh Babu. Generalize then adapt: Source-free domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7046–7056, 2021. [8](#)
- [23] Jogendra Nath Kundu, Naveen Venkat, Ambareesh Revanur, R Venkatesh Babu, et al. Towards inheritable models for open-set domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12376–12385, 2020. [3](#)
- [24] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4):541–551, 12 1989. [5](#)
- [25] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. [6](#)
- [26] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. Deeper, broader and artier domain generalization. *CoRR*, abs/1710.03077, 2017. [5](#)
- [27] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018. [2](#)

- [28] Yanghao Li, Naiyan Wang, Jiaying Liu, and Xiaodi Hou. Demystifying neural style transfer. *arXiv preprint arXiv:1701.01036*, 2017. 2
- [29] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International conference on machine learning*, pages 6028–6039. PMLR, 2020. 3
- [30] Hong Liu, Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Qiang Yang. Separate to adapt: Open set domain adaptation via progressive separation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2927–2936, 2019. 3
- [31] Yawei Luo, Ping Liu, Tao Guan, Junqing Yu, and Yi Yang. Adversarial style mining for one-shot unsupervised domain adaptation. *Advances in neural information processing systems*, 33:20612–20623, 2020. 2, 3, 8
- [32] Massimiliano Mancini, Zeynep Akata, Elisa Ricci, and Barbara Caputo. Towards recognizing unseen categories in seen domains. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIII 16*, pages 466–483. Springer, 2020. 1, 2, 3, 5, 8
- [33] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bisacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*, 2011. 5
- [34] Mohammad Mahfujur Rahman, Clinton Fookes, Mahsa Baktashmotlagh, and Sridha Sridharan. Multi-component image translation for deep domain generalization. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 579–588. IEEE, 2019. 3
- [35] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision – ECCV 2010*, pages 213–226. Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. 5
- [36] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018. 3, 6, 7
- [37] 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. 1
- [38] Yang Shu, Zhangjie Cao, Chenyu Wang, Jianmin Wang, and Mingsheng Long. Open domain generalization with domain-augmented meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9624–9633, 2021. 2, 3
- [39] 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. 3
- [40] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Instance normalization: The missing ingredient for fast stylization. *CoRR*, abs/1607.08022, 2016. 4
- [41] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 5
- [42] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. Manifold mixup: Better representations by interpolating hidden states. In *International conference on machine learning*, pages 6438–6447. PMLR, 2019. 3
- [43] Riccardo Volpi, Hongseok Namkoong, Ozan Sener, John C Duchi, Vittorio Murino, and Silvio Savarese. Generalizing to unseen domains via adversarial data augmentation. *Advances in neural information processing systems*, 31, 2018. 6, 7
- [44] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Dada: Depth-aware domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7364–7373, 2019. 3
- [45] Jindong Wang, Cuiling Lan, Chang Liu, Yidong Ouyang, Tao Qin, Wang Lu, Yiqiang Chen, Wenjun Zeng, and Philip Yu. Generalizing to unseen domains: A survey on domain generalization. *IEEE Transactions on Knowledge and Data Engineering*, 2022. 1
- [46] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018. 1
- [47] Zijian Wang, Yadan Luo, Ruihong Qiu, Zi Huang, and Mahsa Baktashmotlagh. Learning to diversify for single domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 834–843, 2021. 1, 2
- [48] 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
- [49] Renjun Xu, Pelen Liu, Yin Zhang, Fang Cai, Jindong Wang, Shuoying Liang, Heting Ying, and Jianwei Yin. Joint partial optimal transport for open set domain adaptation. In *IJCAI*, pages 2540–2546, 2020. 3
- [50] Sangdoon Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6023–6032, 2019. 2
- [51] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017. 2, 3
- [52] Long Zhao, Ting Liu, Xi Peng, and Dimitris Metaxas. Maximum-entropy adversarial data augmentation for improved generalization and robustness. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 14435–14447. Curran Associates, Inc., 2020. 6, 7

- [53] Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization in vision: A survey. *arXiv preprint arXiv:2103.02503*, 2021. [1](#)
- [54] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16*, pages 561–578. Springer, 2020. [2](#), [3](#)
- [55] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. *arXiv preprint arXiv:2104.02008*, 2021. [1](#), [2](#), [3](#)
- [56] Ronghang Zhu and Sheng Li. Crossmatch: Cross-classifier consistency regularization for open-set single domain generalization. In *International Conference on Learning Representations*, 2022. [2](#), [3](#), [5](#), [6](#), [7](#)