Beyond Boundaries: A Novel Data-Augmentation Discourse for Open Domain Generalization

Shirsha Bose  
Technical University of Munich  
shirhabosecs@gmail.com

Ankit Jha  
Indian Institute of Technology Bombay  
ankitjha16@gmail.com

Hitesh Kandala  
Indian Institute of Technology Bombay  
khitesh2000@gmail.com

Biplab Banerjee  
Indian Institute of Technology Bombay  
getbiplab@gmail.com

Reviewed on OpenReview: https://openreview.net/forum?id=jpZmhiIys1

Abstract

The problem of Open Domain Generalization (ODG) is multifaceted, encompassing shifts in domains and labels across all source and target domains. Existing approaches have encountered challenges such as style bias towards training domains, insufficient feature-space disentanglement to highlight semantic features, and discriminativeness of the latent space. Additionally, they rely on a confidence-based target outlier detection approach, which can lead to misclassifications when target open samples visually align with the source domain data. In response to these challenges, we present a solution named ODG-NET. We aim to create a direct open-set classifier within a discriminative, unbiased, and disentangled semantic embedding space. To enrich data density and diversity, we introduce a generative augmentation framework that produces style-interpolated novel domains for closed-set images and novel pseudo-open images by interpolating the contents of paired training images. Our augmentation strategy skillfully utilizes disentangled style and content information to synthesize images effectively. Furthermore, we tackle the issue of style bias by representing all images in relation to all source domain properties, which effectively accentuates complementary visual features. Consequently, we train a multi-class semantic object classifier, incorporating both closed and open class classification capabilities, along with a style classifier to identify style primitives. The joint use of style and semantic classifiers facilitates the disentanglement of the latent space, thereby enhancing the generalization performance of the semantic classifier. To ensure discriminativeness in both closed and open spaces, we optimize the semantic feature space using novel metric losses. The experimental results on six benchmark datasets convincingly demonstrate that ODG-NET surpasses the state-of-the-art by an impressive margin of 1 – 4% in both open and closed-set DG scenarios.

1 Introduction

Domain Generalization (DG), as explored in the work of [Zhou et al. (2022)], aims to establish a shared embedding space derived from labeled source domains, which can then be applied to an unseen target domain. However, current DG methods are primarily tailored to closed-set scenarios, such as those seen in [Zhou et al. (2020b) and Zhou et al. (2020a)], where both source and target domains possess identical label sets. Nonetheless, this approach might not always be viable in dynamic real-world contexts, as exemplified by Robotics, where a navigating robot might encounter categories that are either common or unique to its surroundings [Zhao & Shen (2022)]. This realization underscores the necessity of tackling the more pragmatic and intricate realm of Open Domain Generalization (ODG) [Shu et al. (2021)], which revolves around training on labeled source domains housing both shared and domain-specific categories.
In the context of ODG, the target domain consists of samples belonging to either familiar classes or novel classes exclusive to that particular domain, thereby introducing several noteworthy challenges. Firstly, a substantial data imbalance arises due to the uneven representation of known classes within the source domains. Secondly, the task of establishing an embedding space that is both domain-agnostic and discriminatory becomes formidable in light of unregulated shifts in both domains and labels. Lastly, the absence of prior knowledge concerning the open space in the target domain adds to the complexity.

While one potential approach to addressing ODG could involve the fusion of a pre-existing closed-set DG technique with a readily available open-set recognition (OSR) method, such as Openmax [Bendale & Boult (2016b)], this strategy may not yield optimal results. The DG technique could potentially be influenced by the domain-specific classes found in ODG, which are subject to significant under-representation.

Surprisingly, Open Domain Generalization (ODG) has garnered little attention in the Domain Generalization (DG) literature, with DAML [Shu et al. (2021)] being the sole model explicitly designed for ODG. However, our extensive investigation revealed three significant limitations in DAML’s approach. Initially, DAML augments source domains via multi-domain mix-up features, using a Dirichlet-distribution based weight scheme, where identical weight configurations correspond to labels for generated features. However, merging content and style details in the raw latent features might result in semantically inconsistent feature-label pairs. This was evidenced in a PACS dataset [Li et al. (2017)] experiment, where classifying the synthesized features yielded notably low accuracy. Additionally, DAML overlooks disentangling content features from domain attributes, potentially introducing biases towards specific styles, hindering adaptability across diverse visual domains. Furthermore, DAML’s outlier rejection relies on thresholding source domain classifier responses, that can produce erroneous results, especially when target open samples visually align with certain source domain classes. These limitations underscore the imperative need for further exploration and refinement of ODG techniques, thereby unlocking their full potential and meaningfully addressing the inherent complexities.

To overcome these challenges, we introduce an innovative approach for training a robust generalized semantic open-set classifier. This classifier can distinguish known and potential new samples in an unbiased and disentangled embedding space. However, we face a hurdle as we lack representative open-space samples during training. To address this, we propose a novel technique called "sample hallucination," which generates these essential samples, facilitating comprehensive classifier training. Additionally, we tackle the issue of class imbalance in ODG by diversifying the appearances of training classes. Our method emphasizes the importance of representing images in a feature space unaffected by training domains. To achieve this, we advocate separating semantic features from style elements, revealing each image’s true content. Ultimately, we ensure that the semantic space effectively discriminates between classes, enabling our classifier to make precise distinctions across various categories. These innovations pave the way for more effective and accurate open domain generalization.

**Our proposed ODG-Net:** In this paper, we introduce our novel model, ODG-Net (depicted in Fig. 1), which addresses the aforementioned concerns comprehensively. ODG-Net consists of three pivotal modules, each devoted to addressing model bias, feature disentanglement, and the creation of a discriminative semantic feature space.

At the core of our research is the aim to enrich the existing source domains using two types of synthesized images generated by a novel conditional GAN. These image types serve specific purposes: expanding the diversity of closed-set classes and creating representative pseudo-open images. The first image type, called domain or style mix-up images, involves a sophisticated interpolation of style properties from source domains using a Dirichlet distribution. This
Our approach introduces new domains, diversifying the style of source images while preserving their inherent semantic object characteristics. The second type, known as pseudo open-space images, results from skillful interpolation of both domain and class identifiers from source domain images, achieved through Dirichlet sampling. To ensure the creation of diverse and meaningful samples, we’ve introduced diversification regularization to prevent potential mode collapse during the generation of domain or label-interpolated samples. Additionally, a structural cycle consistency mechanism has been implemented to maintain the structural integrity of the generated images.

Our approach tackles a range of formidable challenges, spanning from class imbalance and limited style diversity to the absence of an open-space prior. A unique aspect of our methodology lies in its capacity to unify label and domain mix-up concepts while offering customization in conditioning. This surpasses existing augmentation methods, which are limited to style or random image mix-ups [Mancini et al. (2020a); Zhou et al. (2021)]. ODG-Net strives for an unbiased latent embedding space, devoid of source domain bias. We achieve this through an innovative approach involving the training of domain-specific classifiers, which adeptly capture domain-specific features from each source domain. Consequently, each image is represented as a concatenation of features from all domain-specific models, creating a comprehensive and impartial embedding.

To disentangle domain-specific attributes from semantic object features in latent representations, we train two attention-driven classifiers: a domain classifier for domain label identification and an object classifier for recognizing class labels from the augmented domain set. This enriches our model’s grasp of object semantics while significantly mitigating the influence of domain-specific artifacts. For a highly discriminative semantic feature space, we introduce a contrastive loss among known classes, accentuating differences between various categories. Additionally, our entropy minimization objective strategically pushes pseudo outliers away from the known-class boundary, bolstering the model’s robustness.

We summarize our major contributions as:

1) In this paper, we introduce ODG-Net, an end-to-end network that tackles the challenging ODG problem by jointly considering closed and open space domain augmentation, feature disentanglement, and semantic feature-space optimization.

2) To synthesize augmented images that are diverse from the source domains, we propose a novel conditional GAN with a cycle consistency constraint and an anti-mode-collapse regularizer that interpolates domain and category labels. We also adopt a classification-based approach for feature disentanglement. Finally, we ensure the separability of the semantic feature space for closed and open classes through novel metric objectives.

3) We evaluate ODG-Net on six benchmark datasets in both open and closed DG settings. Our experiments demonstrate that ODG-Net consistently outperforms the literature. For instance, on ODG for Multi-dataset Shu et al. (2021) and on closed DG for DomainNet [Peng et al. (2019)], ODG-Net outperforms the previous state-of-the-art by approximately 3%.

2 Related Works

Open-set learning and open-set domain adaptation: The Open-set Recognition (OSR) [Bendale & Boult (2016c); Kong & Ramanan (2021); Pal et al. (2023b); Vaze et al. (2021)] challenge involves effectively identifying novel, unknown-class samples during testing, leveraging training samples from known closed-set classes. However, OSR doesn’t account for any differences in distributions between the training and test sets. Another relevant problem is Open-set Domain Adaptation (OSDA) [Panareda Busto & Gall (2017); Saito et al. (2018); Kundu et al. (2020); Bucci et al. (2020)], which addresses the scenario of a labeled source domain and an unlabeled target domain. The target domain contains unlabeled samples from the same semantic classes as the source domain, along with novel-class samples unique to the target domain. OSDA operates in a transductive manner, where both source and target domains are simultaneously employed during training. In contrast, Open Domain Generalization (ODG) sets itself apart from OSR and OSDA. In ODG, the target domain remains unseen during training, making it distinct. Additionally, the multiple source domains consist of a combination of shared and private categories. This diversity of categories, including shared and domain-specific ones, renders ODG even more intricate than the other tasks.

(Open) DG: DG refers to the problem of learning a supervised learning model that is generalizable across any target distribution without any prior knowledge. The initial studies in closed-set DG focused on domain adaptation (DA) [Li et al. (2010, 2012, 2014)].
In the context of ODG, we have access to multiple source domains denoted as \( D_1, D_2, \ldots, D_S \). Each of these domains has different distributions and contains a combination of domain-specific and shared categories. During training, we use labeled samples from each domain \( D_s = (x_s^i, y_s^i)_{i=1}^{n_s} \), where \( y_s \in Y_s \) is the label for \( x_s \in X_s \). The total number of classes in \( D \) is denoted by \( C \). The target domain \( D_T = \{ x_T^j \}_{j=1}^{n_T} \) has a distribution that is different from that of \( D \). It consists of unlabeled samples that belong to one of the source classes present in \( D \) or novel classes.

### Augmentation in DG

Data augmentation is a crucial technique in DG, and it can be implemented using various methods such as variational autoencoders, GANs, and mixing strategies. For instance, Rehman et al. (2020), Kingma & Welling (2013), and Zhang et al. (2017) used ComboGAN to generate new data and optimized ad hoc domain divergence measures to learn a domain-generic space. Zhou et al. (2020b) used mix-up techniques, which interpolates between a pair of samples and their labels. Similarly, Li et al. (2021b) used adversarial training to generate domains instead of samples. Mix-up, on the other hand, generates new data by interpolating between a pair of samples and their labels.

Recently, mix-up techniques [Yun et al. (2019); Mancini et al. (2020a); Zhou et al. (2021)] have become popular in the DG literature, applied to either the image or feature space.

The augmentation approach used by ODG-NET stands out from the existing literature by going beyond simple style or image mix-up. Our approach ensures that the object properties of images remain intact when using style mix-up, and we also have control over label mix-up to generate pseudo-open samples that can scatter the open space with varying levels of similarity to the source domains.

**Augmentation in DG**: Data augmentation is a crucial technique in DG, and it can be implemented using various methods such as variational autoencoders, GANs, and mixing strategies. For instance, Rehman et al. (2020), Kingma & Welling (2013), and Zhang et al. (2017) used ComboGAN to generate new data and optimized ad hoc domain divergence measures to learn a domain-generic space. Zhou et al. (2020b) combined GAN-based image generation with optimal transport to synthesize images different from the source data. Gong et al. (2019) treated generation as an image-to-image translation process and extracted intermediate images given an image pair. Similarly, Li et al. (2021b) used adversarial training to generate domains instead of samples. Mix-up, on the other hand, generates new data by interpolating between a pair of samples and their labels. Recently, mix-up techniques [Yun et al. (2019); Mancini et al. (2020a); Zhou et al. (2021)] have become popular in the DG literature, applied to either the image or feature space.

The augmentation approach used by ODG-NET stands out from the existing literature by going beyond simple style or image mix-up. Our approach ensures that the object properties of images remain intact when using style mix-up, and we also have control over label mix-up to generate pseudo-open samples that can scatter the open space with varying levels of similarity to the source domains.

Among the existing augmentation strategies, Zhou et al. (2020b) and Gong et al. (2019) are the closest to our approach as they both use conditional GANs. However, there are several key differences between our method and theirs: (a) Gong et al. (2019) requires paired training data to sample intermediate pseudo-stylized images, whereas we use conditional generation without the need for paired data; (b) Zhou et al. (2020b) uses extrapolation for domains, which is ill-posed, while we use Dirichlet distributions to interpolate domains and classes; and (c) while both Zhou et al. (2020b) and Gong et al. (2019) use style mix-up for closed-set data, we generate both closed and pseudo-open samples judiciously.

**Disentangled representation learning**: Disentangled feature learning refers to the process of modeling distinct and explanatory data variation factors. As per Dittadi et al. (2020), disentanglement can aid in out-of-distribution tasks. Previous efforts have focused on disentangling semantic and style latent variables in the original feature space using encoder-decoder models [Wang et al. (2022); Cai et al. (2019)], causality [Ouyang et al. (2021)], or in the Fourier space [Wang et al. (2022)]. These models are complex and require sophisticated knowledge to improvement the feature learning of the models. In contrast, ODG-NET proposes to use simple to implement yet effective, attention-based classifiers, to separate the style and semantic primitives from the latent visual representations.

### 3 Problem Definition and Proposed Methodology

In the context of ODG, we have access to multiple source domains denoted as \( D = \{ D_1, D_2, \ldots, D_S \} \). Each of these domains has different distributions and contains a combination of domain-specific and shared categories. During training, we use labeled samples from each domain \( D_s = (x_s^i, y_s^i)_{i=1}^{n_s} \), where \( y_s \in Y_s \) is the label for \( x_s \in X_s \). The total number of classes in \( D \) is denoted by \( C \). The target domain \( D_T = \{ x_T^j \}_{j=1}^{n_T} \) has a distribution that is different from that of \( D \). It consists of unlabeled samples that belong to one of the source classes present in \( D \) or novel classes.
that were not seen during training. The objective is to model a common classifier that can reject outliers while properly recognizing the known class samples given $D$ and then evaluate its performance on $D_T$.

In our formulation, each domain/style in $D$ is represented using an $S$-dimensional one-hot vector $v_d$. In contrast, a pseudo-domain (synthesized style) is represented by $\hat{v}_d$, which is sampled from a Dirichlet distribution with parameter $\alpha$ and has the same length as $v_d$. For instance, if $S = 3$, a source domain can be represented as a three-dimensional one-hot vector (e.g., $[0, 0, 1]$), while a $\hat{v}_d$ could be $[0.2, 0.3, 0.5]$.

Similarly, we denote the label $y$ as a $C$-dimensional one-hot vector. On the other hand, an interpolated label space is represented by $\hat{y}$, which is sampled in the same way as $\hat{v}_d$. A real image-label pair from $D$ is denoted as $(x_r, y)$. In contrast, a cGAN synthesized image is denoted by $(x_{cs}^a, y)$. We condition $F_G$ on four variables: the domain label $v_d/\hat{v}_d$, the class label $y/\hat{y}$, an input image $x_r/x_f^a/x_f^a$, and a noise tensor $\eta_1/\eta_2/\eta_3$.

Figure 2: A depiction of the ODG-NET architecture. It shows the model components: the embedding networks: $(F_{im}, F_v, F_y, F_\eta)$, cGAN consisting of $(F_G, F_{disc})$, the local domain-specific classifiers $\{F_l^s = (F_{bs}^s, F_{ls}^s)\}_{s=1}^S$, and the global domain and semantic classifiers $(F_d, F_o)$ with corresponding attention blocks $A_d$ and $A_o$, respectively. Colors indicate the flow of information for different data items.

3.1 Architecture and training overview for ODG-NET

Our objective is to craft a direct open-set classifier operating within a disentangled, discriminative, and unbiased semantic feature space. Additionally, we introduce a method for generating training samples for the open space by proposing the creation of pseudo-open samples through a generative module.

To accomplish our aims, we introduce ODG-NET, composed of four modules (illustrated in Fig. 2). First and foremost, ODG-NET employs a generative augmentation strategy, utilizing a conditional GAN equipped with a U-Net-based generator denoted as $F_G$ and a binary discriminator referred to as $F_{disc}$. We condition $F_G$ on four variables: the domain label $v_d/\hat{v}_d$, the class label $y/\hat{y}$, an input image $x_r/x_f^a/x_f^a$, and a noise tensor $\eta_1/\eta_2/\eta_3$.
In this context, we train \( F \) to generate two types of images: (i) \((x_{f}^{cs}, y)\) when conditioned on \((x_r, \hat{v}_d, y, \eta_1)\), where \(x_{f}^{cs}\) retains the class label \(y\) of \(x_r\) while the style changes according to \(\hat{v}_d\), and (ii) \((x_{f}^{os}, \hat{y})\) when conditioned on \((x_r, v_d/\hat{v}_d, \hat{y}, \eta_2)\), thereby modifying the semantic and stylistic characteristics of \(x_r\) according to \(\hat{y}\) and \(v_d/\hat{v}_d\) in \(x_{f}^{os}\). We employ a standard min-max formulation to train the conditional GAN and introduce a regularizer to ensure that the generated samples do not closely resemble the data from \(D\) (as expressed in Eq. [1]).

Furthermore, we introduce a cycle consistency loss to maintain the structural robustness of the images. This approach constructs a multi-view representation space that encompasses diverse and complementary perspectives of the images.

### 3.2 Loss functions, training, and inference

**Regularized cGAN objectives with structural cycle consistency for image synthesis:** As previously mentioned, we employ a cGAN model to generate potential open-space images. Within this framework, \( F_G \) and \( F_{disc} \) engage in a min-max adversarial game. The primary objective of \( F_{disc} \) is to accurately discern between real and synthesized images, while \( F_G \) endeavors to deceive \( F_{disc} \). To prevent the generated images from being too similar to those in \( D \), we introduce a regularization term denoted as \( W \). This regularization term penalizes scenarios in which the real and synthesized images, represented by their respective features \( F_{ei}(x_r) \) and \( F_{ei}(x_{f}^{cs/os}) \), become indistinguishable for slightly different values of \((v_d, \hat{v}_d)\) or \((\hat{y}_d, \hat{\hat{y}}_d)\). Here, \(\delta\) represents the cosine similarity, and \(\epsilon\) is a small constant. Essentially, if \(\delta(v_d, \hat{v}_d)\) or \(\delta(\hat{y}_d, \hat{\hat{y}}_d)\) tends toward 1, \(W\) enforces \(\delta(F_{ei}(x_r), F_{ei}(x_{f}^{cs/os}))\) to approach 0, minimizing the loss. The comprehensive loss formulation is presented below.

\[
L_{\text{GAN}} = \mathbb{E}_{P_D, P_{\text{cs/os}}^{\text{noise}}} \left[ \log(F_{\text{disc}}(F_{ei}(x_r))) + \log(1 - F_{\text{disc}}(F_{ei}(x_{f}^{cs/os}))) + \beta \frac{\delta(F_{ei}(x_r), F_{ei}(x_{f}^{os/cs})) + \epsilon}{\delta(v_d, \hat{v}_d) + \delta(\hat{y}_d, \hat{\hat{y}}_d) + \epsilon} \right] 
\]

In this context, \(P_D\), \(P_{\text{cs}}^{\text{noise}}\), and \(P_{\text{os}}^{\text{noise}}\) denote the data distribution of the source domains in \(D\) and the noise used to generate closed-set and open-set samples, respectively. We set \(P_{\text{cs}}^{\text{noise}} = \mathcal{N}(0, \|)\), and \(P_{\text{os}}^{\text{noise}} = \mathcal{N}(0, \sigma)\), where \(\sigma\) is a large value. Our goal is to limit the space of generated closed-set images so that they represent similar semantic concepts while allowing for more scattering in the pseudo-open space, aiding in learning a robust open-set classifier.

To maintain the structural robustness of \(F_G\) against variations in style or label, we propose a method for reconstructing \(x_r\). We take into account the embeddings of the synthesized \(x_{f}^{cs/os}\), the actual class label \(y\), the domain identifier \(v_d\), and a noise vector \(\eta_3 \in \mathcal{N}(0, \|)\) as inputs to \(F_G\): \(x_r^{\text{rec}} = F_G(x_{f}^{cs/os}, v_d, y, \eta_3)\). By following the path \(x_r \rightarrow x_{f}^{cs/os} \rightarrow x_r^{\text{rec}}\), we ensure that \(x_{f}^{cs/os}\) represents a meaningful image rather than noisy data.

To compute the loss, we use the standard \(\ell_1\) distance between the original real domain images and the remapped images \(x_r^{\text{rec}}\), given by:

\[
L_{\text{rec}} = \mathbb{E}_{P_D, \mathcal{N}(0, \|)} \left[ \|x_r - x_r^{\text{rec}}\|_1 \right].
\]

**Learning style agnostic latent representations for all the images:** Subsequently, our aim is to guarantee that the latent feature embeddings of images remain impartial and not skewed toward any particular training source domain. To mitigate the risk of overfitting to any specific source domain, we introduce a method wherein input images are represented based on the characteristics of all source domains, leveraging the feature representation \(F_{ei}(x)\). This approach constructs a multi-view representation space that encompasses diverse and complementary perspectives of the images.
To achieve this, we train \( \mathcal{F}_s^p \) using \( D_s \) where \( s \) belongs to \( \{1, 2, \cdots, S\} \). We consider \( S \) multiclass cross-entropy losses (\( L_{\text{CE}} \)) for this purpose (as shown in Eq. (3)), where \( P^s_{D} \) represents the data distribution for the \( s^{th} \) source domain.

\[
L_{\text{Local}} = \frac{1}{S} \sum_{s \in \{1, 2, \cdots, S\}} E_{P^s_{D}} [L_{\text{CE}}(\mathcal{F}_s^p(x_s), y_s)].
\]  

Disentangling latent features of the images to highlight the semantic contents through global classifiers: Simultaneously, our objective is to disentangle domain-specific attributes from the semantic object features within the previously derived latent representations. This disentanglement facilitates the object classifier in focusing exclusively on the semantic content, thereby enhancing its generalizability across novel target domains.

In this context, the global domain classifier, denoted as \( \mathcal{F}_d \), is tasked with identifying domain identifiers based on the attended features, which are defined as \( F_{\text{es}}(x) = F_{\text{el}}(x) \otimes A_d + F_{\text{el}}(x) \). This is achieved through the use of a multiclass cross-entropy loss. It is worth noting that \( \mathcal{F}_d \) implicitly ensures that \( \mathcal{F}_G \) generates images in accordance with the specified conditioning domain identifiers. The corresponding loss function is delineated below.

\[
L_{\text{dom}} = \mathbb{E}_{P_{D}, P_{\text{no/sem}}^s} [L_{\text{CE}}(\mathcal{F}_d(F_{\text{es}}(x_r^{\text{rec}})), v_d) + L_{\text{CE}}(\mathcal{F}_d(F_{\text{es}}(x_r^{\text{os/sem}})), \hat{v}_d)].
\]  

In contrast, the open-set classifier, \( \mathcal{F}_{os} \), is trained to accurately identify all samples belonging to known classes while disregarding the generated pseudo-outliers by labeling them as \( C + 1 \), from \( F_{\text{os}}(x) = F_{\text{el}}(x) \otimes A_0 + F_{\text{el}}(x) \). This is achieved using a multiclass cross-entropy loss. Similar to \( \mathcal{F}_d \), \( \mathcal{F}_{os} \) also aids \( \mathcal{F}_G \) in producing high-quality synthesized images and supplements \( L_{\text{rec}} \).

\[
L_{\text{class}} = \mathbb{E}_{P_{D}, P_{\text{no/sem}}^s} [L_{\text{CE}}(\mathcal{F}_{os}(F_{\text{es}}(x_r)), \hat{y}) + L_{\text{CE}}(\mathcal{F}_{os}(F_{\text{es}}(x_r^{\text{os/sem}})), \hat{y})].
\]  

\( \mathcal{F}_d \) and \( \mathcal{F}_{os} \) work together on \( F_{\text{el}}(x) \), and seek to learn the domain-specific and semantic features separately, suggesting that fact that both the networks are devoted to disentangling the latent features wisely.

**Algorithm 1 ODG-NET training algorithm**

**Require:** Initialized \( \mathcal{F}_G, \mathcal{F}_{im}, \mathcal{F}_y, \mathcal{F}_\eta, \mathcal{F}_d, \mathcal{F}_{os}, \mathcal{F}_{disc}, \{\mathcal{F}_s\}_{s=1}^S \)

1. while Not Converged do
2. Sample a batch of \((x_r, y, v_d)\) from \( D \) and \( \eta_1 \sim \mathcal{N}(0, I), \eta_2 \sim \mathcal{N}(0, \sigma), \eta_3 \sim \mathcal{N}(0, I) \). \( \sigma \) is the noise variance for generating the pseudo-open samples.
3. Generate \( \hat{v}_d, \hat{y} \) using Dirichlet(\( \alpha \)). \( \alpha \) is the parameter of the distribution.
4. Obtain a batch of \( x_r^{cs} = \mathcal{F}_G(x_r, \hat{v}_d, y, \eta_1) \).
5. Obtain a batch of \( x_r^{os} = \mathcal{F}_G(x_r, v_d, \hat{v}_d, \hat{y}, \eta_2) \).
6. Obtain \( x_r^{\text{rec}} = \mathcal{F}_G(x_r^{cs/\text{os}}, v_d, y, \eta_3) \).
7. Obtain \( F_{\text{el}}, \) the latent representation corresponding to \((x_r, x_r^{cs}, x_r^{os}, x_r^{\text{rec}})\).
8. Obtain \( F_{\text{es}}, \) the attended domain features, and \( F_{\text{os}}, \) the attended semantic features from \( F_{\text{el}} \).
9. Solve: \( \arg \min \mathcal{F}_{\text{im}}, \mathcal{F}_{\text{os}}, \mathcal{F}_y, \mathcal{F}_\eta, \mathcal{F}_d, \mathcal{F}_{\text{disc}}, \{\mathcal{F}_s\}_{s=1}^S \) \[ w_{\text{Gan}}L_{\text{Gan}} + w_{\text{rec}}L_{\text{rec}} + w_{\text{local}}L_{\text{local}} \].
10. Solve: \( \arg \min \mathcal{F}_{\text{os}}, \mathcal{F}_d \) \( L_{\text{dom}} + L_{\text{class}} + w_{\text{sem}}L_{\text{sem}} \).
end while

**Ensuring discriminativenss of the semantic feature space.** The inherent diversity within multi-domain data poses a challenge for \( \mathcal{F}_{os} \) in creating an optimized semantic feature space based on \( \mathcal{F}_{os}^p \). In this space, it’s expected that closed-set classes from the augmented domains should form distinct clusters, while pseudo-open-set samples should be effectively pushed away from the support region of the closed set. To enhance the discriminative qualities, we propose the utilization of a contrastive loss for closed-set samples across the augmented domain set. Simultaneously, we aim to minimize the entropy (\( \mathcal{E} \)) of \( \mathcal{F}_{os} \) predictions for pseudo-open samples. Minimizing the entropy effectively acts as a weighting mechanism for \( \mathcal{F}_{os} \) when handling pseudo-open samples. This weighting strategy increases the posterior probability \( p(y = C + 1|x_r^{cs/\text{os}}) \) for pseudo-open samples while reducing the posteriors associated with the known-class indices \((1 - C)\). This approach contributes to the overall objective of creating a more discriminative semantic feature space, which is crucial for the successful separation of classes and pseudo-open samples.
To implement the contrastive loss, \( L_{\text{con}} \), we select an anchor sample, \( x_a \), and randomly obtain a positive sample, \( x_+ \), which shares the same class label as \( x_a \), as well as a set of negative samples, \( \{x^m\}_{m=1}^M \), where no restrictions are imposed on the styles of the samples. The goal is to maximize the cosine similarity, \( \delta \), for \( (x_a, x_+) \), while minimizing it for all possible pairs of \( (x_a, x^m) \). The semantic feature space optimization loss can be expressed as follows:

\[
L_{\text{sem}} = \mathbb{E}_{p_{D}, p_{\text{dis}}/\text{con}} [\mathcal{E}(\mathcal{F}_o(F_{\text{enc}}(x^a_i))) + L_{\text{con}}].
\]

where \( L_{\text{con}} \) is defined as follows,

\[
L_{\text{con}} = \left[ -\log \frac{\exp(\delta(F^b(x_a), F^b(x_+))}{\sum_{m=1}^M \exp(\delta(F^b(x_a), F^b(x^m)))} \right].
\]

**Total loss and training.** We follow an alternate optimization strategy in each training episode for ODG-NET, mentioned in Algorithm 1. In the vanilla training stage, we train the embedding networks \( (F_i)_s \), \( G \), and the local domain-specific networks \( (F^s_i)_{s=1}^S \), given the fixed \( (F_d, F_o) \) to produce meaningful images. \( \mathcal{L} \) represents the loss contributions and we set them to the value 1 in all our experiments.

\[
\text{argmin}_{F_i, F_d} \text{argmax}_{F_o} [w_{\text{GAN}} L_{\text{GAN}} + w_{\text{rec}} L_{\text{rec}} + w_{\text{local}} L_{\text{local}}].
\]

Subsequently, we train \( F_o \) and \( F_d \) to obtain the optimized semantic classifier keeping other parameters fixed.

\[
\text{arg min}_{F_o, F_d} [L_{\text{dom}} + L_{\text{class}} + w_f L_{\text{sem}}].
\]

**Testing.** During inference, images from \( D_T \) are provided as input to \( (F^s_i)_{s=1}^S \). The class labels with the highest softmax probability scores are predicted according to \( F_o \).

### 4 Experimental Evaluations

**Datasets.** We present our results on six widely used benchmark datasets for DG. Specifically, we follow the approach of [Shu et al.] (2021) and use the following datasets: (1) Office-Home [Venkateswara et al.] (2017), (2) PACS [Li et al.] (2017), (3) Multi-Dataset [Shu et al.] (2021). In addition, we introduce the experimental setup of ODG for two additional DG datasets, namely VLCS [Fang et al.] (2013) and Digits-DG [Zhou et al.] (2020b) in this paper. For our closed-set DG experiment, we also utilize the large-scale DomainNet [Peng et al.] (2019).

**Implementation details:** To ensure clarity, we use a ResNet-18 based backbone [He et al.] (2016) for \( F_o \) consistently, while we adopt standard architectures per benchmark for closed-DG tasks, following established literature [Zhou et al.] (2020b). Our attention modules \( (A_d, A_o) \) are composed of a pair of spatial and spectral attention modules, implemented using the query-key-value processing-based approach [Han et al.] (2022). In total, ODG-NET comprises 48 million parameters for \( S = 3 \), and the training stage requires 65 GFLOPS.

**Training protocol and model selection.** We employ a standardized training protocol across all datasets. During each training iteration, we first optimize Eq. 3 using the Adam optimizer [Kingma & Ba] (2014), with a learning rate of \( 2\epsilon - 4 \) and betas of \((0.5, 0.99)\). We then minimize Eq. 9 using Adam with a learning rate of \( 2\epsilon - 2 \) and betas of \((0.9, 0.99)\). Our batch size is typically set to 64, and we train for 30 epochs, except for DomainNet, where we use a batch size of 128 and train for 40 epochs. We follow a cross-validation approach to estimate the loss weights, holding out 10% of samples per domain and using held-out pseudo-open-set validation samples obtained through cumix [Mancini et al.] (2020), that the model has not seen to select the best-performing model. In this regard, the mixup samples do not have a clear semantic meaning as they are generated by randomly combining two images. Hence, they can be considered representative open samples. We further consider \( \beta = 0.5 \) to put \( \mathcal{W} \) as a soft constraint in \( L_{\text{GAN}} \). A large \( \beta \) instigates the generation of ambiguous images in order to make them different from \( D \). Besides, we set \( \alpha = 0.5 \) following [Shu et al.] (2021).

**Evaluation protocol.** For ODG experiments, we report the top-1 accuracy for closed-set samples (Acc) and the H-score for closed and open samples. For closed-set DG experiments, we consider the top-1 performance. We report the mean \( \pm \) std. over three runs.
4.1 Results on open DG tasks

**Baselines.** Our baseline method, AGG, involves merging the source domains with different label sets and training a unified classifier on all the classes. In comparison, we evaluate the performance of ODG-NET against traditional DG methods that are less sensitive to label changes between different source domains, as outlined in Shu et al. (2021). These include state-of-the-art meta-learning-based and augmentation-based DG methods Li et al. (2018a; 2019a); Mancini et al. (2020a); Zhou et al. (2021); Shi et al. (2021); Rame et al. (2022), heterogeneous DG Li et al. (2019b), and methods that produce discriminative and generic embedding spaces Wang et al. (2020b); Huang et al. (2020); Zhang et al. (2022). As per Shu et al. (2021), we employ a confidence-based classifier for our competitors. Here, a sample is classified as unknown if the class probabilities are below a predefined threshold. Alternatively, we also compare against the only existing ODG technique, DAML Shu et al. (2021) and consider a variant where we combine DAML with Openmax Bendale & Boult (2016b) based OSR. Finally, we report the results of two open-set recognition baselines, Openmax Bendale & Boult (2016b) and MORGAN Pal et al. (2023b).

**Table 1:** Comparative analysis for PACS on ODG. (In %)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Art Acc</th>
<th>Sketch H-score</th>
<th>Photo Acc</th>
<th>Cartoon H-score</th>
<th>Avg Acc</th>
<th>Avg H-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGG</td>
<td>51.35</td>
<td>49.75</td>
<td>53.15</td>
<td>46.43</td>
<td>55.17</td>
<td>44.78</td>
</tr>
<tr>
<td>OpenMax Bendale &amp; Boult (2016b)</td>
<td>53.33</td>
<td>54.18</td>
<td>73.39</td>
<td>53.87</td>
<td>63.98</td>
<td>50.54</td>
</tr>
<tr>
<td>MORGAN Pal et al. (2023b)</td>
<td>44.56</td>
<td>52.11</td>
<td>60.31</td>
<td>48.69</td>
<td>58.37</td>
<td>46.37</td>
</tr>
<tr>
<td>MLDG Li et al. (2018a)</td>
<td>45.59</td>
<td>51.29</td>
<td>62.20</td>
<td>57.43</td>
<td>54.00</td>
<td></td>
</tr>
<tr>
<td>FC Li et al. (2019b)</td>
<td>51.12</td>
<td>51.15</td>
<td>69.32</td>
<td>58.13</td>
<td>46.69</td>
<td></td>
</tr>
<tr>
<td>EpisFCR Li et al. (2019a)</td>
<td>54.16</td>
<td>46.35</td>
<td>72.00</td>
<td>60.64</td>
<td>48.47</td>
<td></td>
</tr>
<tr>
<td>PAR Wang et al. (2020b)</td>
<td>52.97</td>
<td>52.00</td>
<td>67.77</td>
<td>56.56</td>
<td>44.95</td>
<td></td>
</tr>
<tr>
<td>RSC Huang et al. (2009)</td>
<td>50.47</td>
<td>45.49</td>
<td>67.51</td>
<td>58.92</td>
<td>45.05</td>
<td></td>
</tr>
<tr>
<td>CuMixM Mancini et al. (2020a)</td>
<td>53.85</td>
<td>37.70</td>
<td>74.16</td>
<td>57.85</td>
<td>41.05</td>
<td></td>
</tr>
<tr>
<td>Fish Shi et al. (2021)</td>
<td>52.22</td>
<td>55.54</td>
<td>69.85</td>
<td>61.75</td>
<td>48.61</td>
<td></td>
</tr>
<tr>
<td>Disentanglement Zhang et al. (2022)</td>
<td>53.18</td>
<td>53.36</td>
<td>70.54</td>
<td>63.02</td>
<td>47.42</td>
<td></td>
</tr>
<tr>
<td>Mixstyle Zhou et al. (2022)</td>
<td>53.41</td>
<td>54.44</td>
<td>71.54</td>
<td>63.55</td>
<td>48.30</td>
<td></td>
</tr>
<tr>
<td>DAML Shu et al. (2021)</td>
<td>54.10</td>
<td>55.33</td>
<td>76.85</td>
<td>65.49</td>
<td>51.88</td>
<td></td>
</tr>
<tr>
<td>DAML + OpenMax Bendale &amp; Boult (2016a)</td>
<td>52.73</td>
<td>56.82</td>
<td>75.84</td>
<td>65.23</td>
<td>51.45</td>
<td></td>
</tr>
<tr>
<td><strong>ODG-NET</strong></td>
<td>57.21</td>
<td>47.67</td>
<td>68.80</td>
<td>55.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2:** Comparative analysis for Office-Home on ODG. (In %)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Clipart Acc</th>
<th>Real-World H-score</th>
<th>Product Acc</th>
<th>Art H-score</th>
<th>Avg Acc</th>
<th>Avg H-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGG</td>
<td>42.83</td>
<td>62.40</td>
<td>54.27</td>
<td>42.22</td>
<td>50.43</td>
<td>47.41</td>
</tr>
<tr>
<td>OpenMax Bendale &amp; Boult (2016b)</td>
<td>43.29</td>
<td>59.86</td>
<td>56.71</td>
<td>48.76</td>
<td>53.81</td>
<td>48.30</td>
</tr>
<tr>
<td>MORGAN Pal et al. (2023b)</td>
<td>39.68</td>
<td>59.76</td>
<td>52.29</td>
<td>48.55</td>
<td>49.00</td>
<td></td>
</tr>
<tr>
<td>MLDG Li et al. (2018a)</td>
<td>41.82</td>
<td>55.84</td>
<td>52.52</td>
<td>42.58</td>
<td>51.07</td>
<td>47.58</td>
</tr>
<tr>
<td>FC Li et al. (2019b)</td>
<td>41.80</td>
<td>55.16</td>
<td>54.01</td>
<td>44.13</td>
<td>51.03</td>
<td>48.02</td>
</tr>
<tr>
<td>EpisFCR Li et al. (2019a)</td>
<td>37.13</td>
<td>54.73</td>
<td>54.63</td>
<td>44.46</td>
<td>50.25</td>
<td>48.48</td>
</tr>
<tr>
<td>PAR Wang et al. (2020b)</td>
<td>41.27</td>
<td>57.60</td>
<td>54.13</td>
<td>42.40</td>
<td>51.26</td>
<td>49.03</td>
</tr>
<tr>
<td>RSC Huang et al. (2009)</td>
<td>38.60</td>
<td>53.73</td>
<td>54.61</td>
<td>44.99</td>
<td>49.56</td>
<td>47.89</td>
</tr>
<tr>
<td>CuMixM Mancini et al. (2020a)</td>
<td>41.54</td>
<td>50.58</td>
<td>42.76</td>
<td>41.67</td>
<td>51.67</td>
<td>49.40</td>
</tr>
<tr>
<td>Fish Shi et al. (2021)</td>
<td>43.76</td>
<td>58.74</td>
<td>49.78</td>
<td>54.16</td>
<td>51.75</td>
<td></td>
</tr>
<tr>
<td>Disentanglement Zhang et al. (2022)</td>
<td>44.89</td>
<td>59.51</td>
<td>54.44</td>
<td>45.49</td>
<td>53.16</td>
<td>50.31</td>
</tr>
<tr>
<td>Mixstyle Zhou et al. (2022)</td>
<td>42.88</td>
<td>60.38</td>
<td>54.34</td>
<td>45.43</td>
<td>51.16</td>
<td>49.53</td>
</tr>
<tr>
<td>DAML Shu et al. (2021)</td>
<td>45.13</td>
<td>60.93</td>
<td>53.12</td>
<td>52.34</td>
<td>54.47</td>
<td>54.43</td>
</tr>
<tr>
<td>DAML + OpenMax Bendale &amp; Boult (2016a)</td>
<td>45.51</td>
<td>61.46</td>
<td>51.34</td>
<td>52.34</td>
<td>54.47</td>
<td>54.43</td>
</tr>
<tr>
<td><strong>ODG-NET</strong></td>
<td>49.81</td>
<td>68.45</td>
<td>65.05</td>
<td>59.40</td>
<td>56.69</td>
<td></td>
</tr>
</tbody>
</table>

Quantitative and qualitative analysis. Tables 1-5 present a performance comparison of ODG-NET with the literature on five datasets. ODG-NET consistently outperforms others in terms of Acc and H-score for all domain combinations and the average leave-one-out case where all the domains except one are used during training and the model is validated on the held-out target domain. For example, on PACS, ODG-NET achieves an Acc of 68.80% and an H-score of 55.81%, beats the previous best of DAML+OpenMax which obtained 65.23% and 52.15%, respectively. Our method outperforms Shu et al. (2021) by ≈ 3% for Office-Home and ≈ 5% for VLCS and Digits-DG in H-score. For the Multi-
### Table 3: Comparative analysis for VLCS on ODG. (In %)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Caltech</th>
<th>LabelMe</th>
<th>Pascal VOC</th>
<th>Sun</th>
<th>AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc</td>
<td>H-score</td>
<td>Acc</td>
<td>H-score</td>
<td>Acc</td>
</tr>
<tr>
<td>AGG</td>
<td>65.49</td>
<td>62.59</td>
<td>46.15</td>
<td>42.78</td>
<td>48.29</td>
</tr>
<tr>
<td>OpenMax Bendale &amp; Boult (2016b)</td>
<td>64.19</td>
<td>62.54</td>
<td>47.77</td>
<td>45.41</td>
<td>50.22</td>
</tr>
<tr>
<td>MORGAN Pal et al. (2023a)</td>
<td>61.59</td>
<td>59.87</td>
<td>43.33</td>
<td>40.98</td>
<td>47.83</td>
</tr>
<tr>
<td>MLDG Li et al. (2018a)</td>
<td>66.91</td>
<td>63.11</td>
<td>45.65</td>
<td>41.76</td>
<td>49.37</td>
</tr>
<tr>
<td>FC Li et al. (2019a)</td>
<td>65.59</td>
<td>60.48</td>
<td>45.23</td>
<td>44.22</td>
<td>49.23</td>
</tr>
<tr>
<td>EP-FCR Li et al. (2019a)</td>
<td>66.81</td>
<td>62.98</td>
<td>47.83</td>
<td>45.33</td>
<td>50.22</td>
</tr>
<tr>
<td>PAR Wang et al. (2018b)</td>
<td>65.78</td>
<td>61.25</td>
<td>46.21</td>
<td>42.54</td>
<td>50.11</td>
</tr>
<tr>
<td>RSC Bian et al. (2020)</td>
<td>64.43</td>
<td>61.39</td>
<td>45.61</td>
<td>43.71</td>
<td>49.60</td>
</tr>
<tr>
<td>CuMix Mancini et al. (2020a)</td>
<td>66.21</td>
<td>63.76</td>
<td>46.72</td>
<td>45.59</td>
<td>50.54</td>
</tr>
<tr>
<td>Fish Shi et al. (2021)</td>
<td>65.82</td>
<td>62.29</td>
<td>47.66</td>
<td>46.52</td>
<td>50.11</td>
</tr>
<tr>
<td>Disentanglement Zhang et al. (2022)</td>
<td>63.27</td>
<td>61.86</td>
<td>48.65</td>
<td>45.39</td>
<td>50.53</td>
</tr>
<tr>
<td>Mixstyle Zhou et al. (2021)</td>
<td>66.11</td>
<td>63.19</td>
<td>46.72</td>
<td>46.22</td>
<td>49.75</td>
</tr>
<tr>
<td>DAML Shu et al. (2021)</td>
<td>69.18</td>
<td>64.65</td>
<td>48.22</td>
<td>47.71</td>
<td>49.87</td>
</tr>
<tr>
<td>DAML + OpenMax Bendale &amp; Boult (2016a)</td>
<td>67.24</td>
<td>66.51</td>
<td>46.43</td>
<td>46.18</td>
<td>52.49</td>
</tr>
<tr>
<td>ODG-N</td>
<td>73.42</td>
<td>69.93</td>
<td>51.89</td>
<td>51.56</td>
<td>56.27</td>
</tr>
</tbody>
</table>

### Table 4: Comparative analysis for Digit-DG on ODG. (In %)

<table>
<thead>
<tr>
<th>Methods</th>
<th>MNIST</th>
<th>MNIST_M</th>
<th>SVHN</th>
<th>SYN</th>
<th>AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc</td>
<td>H-score</td>
<td>Acc</td>
<td>H-score</td>
<td>Acc</td>
</tr>
<tr>
<td>AGG</td>
<td>69.45</td>
<td>63.59</td>
<td>43.51</td>
<td>42.15</td>
<td>50.26</td>
</tr>
<tr>
<td>OpenMax Bendale &amp; Boult (2016b)</td>
<td>73.87</td>
<td>65.39</td>
<td>46.71</td>
<td>44.63</td>
<td>53.87</td>
</tr>
<tr>
<td>MORGAN Pal et al. (2023a)</td>
<td>72.45</td>
<td>63.59</td>
<td>41.78</td>
<td>43.32</td>
<td>50.67</td>
</tr>
<tr>
<td>MLDG Li et al. (2018a)</td>
<td>71.33</td>
<td>69.22</td>
<td>43.19</td>
<td>41.78</td>
<td>48.73</td>
</tr>
<tr>
<td>FC Li et al. (2019a)</td>
<td>71.29</td>
<td>66.29</td>
<td>41.22</td>
<td>40.67</td>
<td>47.72</td>
</tr>
<tr>
<td>EP-FCR Li et al. (2019a)</td>
<td>72.39</td>
<td>68.33</td>
<td>45.83</td>
<td>43.34</td>
<td>51.27</td>
</tr>
<tr>
<td>PAR Wang et al. (2018b)</td>
<td>70.88</td>
<td>67.47</td>
<td>44.62</td>
<td>42.65</td>
<td>49.34</td>
</tr>
<tr>
<td>RSC Bian et al. (2020)</td>
<td>72.77</td>
<td>66.34</td>
<td>42.27</td>
<td>41.43</td>
<td>48.32</td>
</tr>
<tr>
<td>CuMix Mancini et al. (2020a)</td>
<td>72.10</td>
<td>67.52</td>
<td>45.88</td>
<td>43.74</td>
<td>52.22</td>
</tr>
<tr>
<td>Fish Shi et al. (2021)</td>
<td>74.43</td>
<td>66.89</td>
<td>42.65</td>
<td>44.45</td>
<td>53.31</td>
</tr>
<tr>
<td>Disentanglement Zhang et al. (2022)</td>
<td>71.29</td>
<td>68.83</td>
<td>45.38</td>
<td>41.59</td>
<td>50.16</td>
</tr>
<tr>
<td>Mixstyle Zhou et al. (2021)</td>
<td>76.56</td>
<td>70.56</td>
<td>47.81</td>
<td>45.66</td>
<td>54.97</td>
</tr>
<tr>
<td>DAML Shu et al. (2021)</td>
<td>73.98</td>
<td>69.88</td>
<td>46.49</td>
<td>45.62</td>
<td>53.34</td>
</tr>
<tr>
<td>DAML + OpenMax Bendale &amp; Boult (2016a)</td>
<td>75.77</td>
<td>71.38</td>
<td>48.51</td>
<td>47.49</td>
<td>55.61</td>
</tr>
<tr>
<td>ODG-N</td>
<td>78.56</td>
<td>75.75</td>
<td>50.52</td>
<td>50.22</td>
<td>57.81</td>
</tr>
</tbody>
</table>

Figure 3: (a) Depiction of the generated pseudo-open-set samples by ODG-N. (b) Depiction of the pseudo-stylized images (columns 2-4) generated by ODG-N w.r.t. the input images mentioned in column 1. (c) Results of the intermediate images (two images for both cases.) showing the transition between a pair of domain/label.
A drop of approximately latent conditioning information in a meaningful way. The model without embedding layers resulted in a performance loss functions for PACS and Office-Home datasets. We confirmed that embedding networks are crucial for learning Model and loss ablation

outperforms DAML Shu et al. (2021) by a significant margin of at least 5 to the more diversified training set on which ODG-N ET 50 observe an improvement of at least 50% 3
observe an improvement of at least 50% 3.

Moreover, we present a collection of synthetically created images produced by our novel ODG-N ET. As shown in Figure 3, (a) displays the generated pseudo-open-set examples, while (b) exhibits the pseudo-stylized pictures (columns 2-4) derived from their corresponding input images (column 1). This comparison highlights the evident transformation from the original input images to the synthesized pseudo images. Additionally, in (c), we demonstrate the domain-independent nature of the semantic space given the augmented source data.

4.2 Results on closed DG tasks

In the context of closed-set DG tasks, we compare the performance of ODG-NET against the existing literature, focusing on supervised pre-training methods that use meta-learning, regularization, and domain augmentation techniques Zhou et al. (2020b), Chen et al. (2021), Kang et al. (2022), Chattopadhyay et al. (2020), Xu et al. (2021), Shu et al. (2021), Shi et al. (2021), Du et al. (2020), Zhao et al. (2020), among others.

As shown in Table 6 for the five benchmark DG datasets, ODG-NET outperforms all comparative techniques in the average leave-one-out DG evaluations, despite these techniques being designed explicitly for closed-set DG. We observe an improvement of at least 3 – 4% across all datasets. For DomainNet, ODG-NET achieves an average accuracy of 50.16%, which is 4% better than the previous state-of-the-art method SWAD Cha et al. (2021), likely due to the more diversified training set on which ODG-NET is trained. Finally, in closed-set DG experiments, ODG-NET outperforms DAML Shu et al. (2021) by a significant margin of at least 5 – 7%.

4.3 Ablation analysis

Model and loss ablation. In Table 7 we present the effects of different components of the ODG-NET model and the loss functions for PACS and Office-Home datasets. We confirmed that embedding networks are crucial for learning latent conditioning information in a meaningful way. The model without embedding layers resulted in a performance drop of approximately 5 – 6% in the H-score. Similarly, attention modules helped to highlight the style and semantic features better, and using \((A_d, A_s)\) resulted in a 3 – 4% improvement in the H-score for both datasets. Additionally, we experimented with using a common backbone for the source domains instead of \(\{F_i\}_{i=1}^S\). This approach significantly reduced the H-score by 4% and 6% for both datasets, indicating the importance of multi-view features learned by the domain-specific backbones.

Furthermore, \(F_d\) helped to discriminate the domain and semantic properties of \(F_{el}(x)\), and the omission of \(F_d\) reduced performance by almost 5%. We also removed both \(F_d\) and the local classifiers \(\{F_i\}_{i=1}^S\) simultaneously, resulting in a performance drop of over 6%. When we trained \(F_e\) from scratch instead of using a pre-trained ResNet-18 backbone, we observed a performance drop of around 3%. The pre-trained ResNet-18 backbone is already rich in discriminative information, which helps our DG tasks. Concerning the loss function of feature optimization, it is evident that both

<table>
<thead>
<tr>
<th>Methods</th>
<th>Acc</th>
<th>H-score</th>
<th>Acc</th>
<th>H-score</th>
<th>Acc</th>
<th>H-score</th>
<th>Acc</th>
<th>H-score</th>
<th>Avg</th>
<th>H-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGG</td>
<td>29.78</td>
<td>34.06</td>
<td>65.33</td>
<td>64.72</td>
<td>44.30</td>
<td>51.04</td>
<td>27.59</td>
<td>35.41</td>
<td>41.75</td>
<td>46.31</td>
</tr>
<tr>
<td>OpenMax Bendale &amp; Boult 2016b</td>
<td>36.72</td>
<td>34.41</td>
<td>63.29</td>
<td>62.88</td>
<td>44.19</td>
<td>50.75</td>
<td>34.51</td>
<td>32.29</td>
<td>44.67</td>
<td>45.08</td>
</tr>
<tr>
<td>MORGAN Park et al. 2020</td>
<td>29.45</td>
<td>28.59</td>
<td>59.97</td>
<td>62.22</td>
<td>43.75</td>
<td>47.53</td>
<td>31.19</td>
<td>30.04</td>
<td>41.09</td>
<td>42.09</td>
</tr>
<tr>
<td>ML-DG Li et al. 2018</td>
<td>29.66</td>
<td>35.11</td>
<td>65.37</td>
<td>54.40</td>
<td>44.04</td>
<td>50.53</td>
<td>26.83</td>
<td>34.57</td>
<td>41.48</td>
<td>43.65</td>
</tr>
<tr>
<td>FC Li et al. 2019b</td>
<td>29.91</td>
<td>35.42</td>
<td>64.77</td>
<td>63.65</td>
<td>44.13</td>
<td>50.07</td>
<td>28.56</td>
<td>34.10</td>
<td>41.84</td>
<td>45.81</td>
</tr>
<tr>
<td>Epi-FCR Li et al. 2019a</td>
<td>27.70</td>
<td>37.62</td>
<td>60.31</td>
<td>64.95</td>
<td>39.57</td>
<td>50.24</td>
<td>26.76</td>
<td>33.74</td>
<td>38.59</td>
<td>46.64</td>
</tr>
<tr>
<td>PAR Wang et al. 2020</td>
<td>29.29</td>
<td>39.99</td>
<td>64.09</td>
<td>62.59</td>
<td>42.36</td>
<td>46.37</td>
<td>30.21</td>
<td>39.96</td>
<td>41.49</td>
<td>47.23</td>
</tr>
<tr>
<td>ES Hu et al. 2020h</td>
<td>27.57</td>
<td>34.98</td>
<td>60.36</td>
<td>60.02</td>
<td>37.76</td>
<td>42.21</td>
<td>26.21</td>
<td>30.44</td>
<td>37.98</td>
<td>41.91</td>
</tr>
<tr>
<td>CuMix Mancini et al. 2020a</td>
<td>30.03</td>
<td>40.18</td>
<td>64.61</td>
<td>65.07</td>
<td>44.37</td>
<td>48.70</td>
<td>29.72</td>
<td>33.70</td>
<td>42.18</td>
<td>46.91</td>
</tr>
<tr>
<td>Fish Shi et al. 2021</td>
<td>32.78</td>
<td>35.42</td>
<td>65.43</td>
<td>67.77</td>
<td>45.37</td>
<td>48.81</td>
<td>32.35</td>
<td>32.45</td>
<td>43.98</td>
<td>46.11</td>
</tr>
<tr>
<td>Disentanglement Zhang et al. 2022</td>
<td>28.76</td>
<td>33.33</td>
<td>64.48</td>
<td>64.44</td>
<td>42.29</td>
<td>50.05</td>
<td>30.65</td>
<td>35.87</td>
<td>41.54</td>
<td>45.92</td>
</tr>
<tr>
<td>Mixstyle Zhou et al. 2021</td>
<td>30.03</td>
<td>40.18</td>
<td>64.61</td>
<td>65.07</td>
<td>44.37</td>
<td>48.70</td>
<td>29.72</td>
<td>33.70</td>
<td>42.18</td>
<td>46.91</td>
</tr>
<tr>
<td>DAML Shi et al. 2021</td>
<td>37.62</td>
<td>44.27</td>
<td>66.54</td>
<td>67.80</td>
<td>47.80</td>
<td>52.93</td>
<td>34.48</td>
<td>41.82</td>
<td>46.61</td>
<td>51.71</td>
</tr>
<tr>
<td>DAML + OpenMax Bendale &amp; Boult 2016a</td>
<td>38.55</td>
<td>45.51</td>
<td>66.87</td>
<td>68.89</td>
<td>48.51</td>
<td>53.12</td>
<td>35.61</td>
<td>42.56</td>
<td>47.38</td>
<td>52.52</td>
</tr>
<tr>
<td>ODG-NET</td>
<td>40.75</td>
<td>47.54</td>
<td>69.49</td>
<td>71.22</td>
<td>50.11</td>
<td>55.39</td>
<td>37.58</td>
<td>44.10</td>
<td>49.48</td>
<td>54.56</td>
</tr>
</tbody>
</table>

Published in Transactions on Machine Learning Research (11/2023)
Figure 4: (a) T-SNE of real and cGAN synthesized images in the semantic feature space $F^b_b$ for PACS dataset. (b) Accuracy comparison between ODG-NET and Cumix [Mancini et al., 2020a] based pseudo-open sample generation. (c) Accuracy comparison between ODG-NET and Mixstyle [Zhou et al., 2021] and L2A-OT [Zhou et al., 2020b] based closed-set sample generation. (d) The Frechet distance [Dowson & Landau, 1982] between real data and the closed and pseudo-open images generated, with and without the consideration of $W$ in $I_{Gan}$.

Figure 5: (a) Frechet distance between the source and target domains for the closed-set classes. (b) Openness analysis.
Table 6: Results of PACS, VLCS, Office-Home, Digits-DG and DomainNet datasets under close-set DG. (In %)

<table>
<thead>
<tr>
<th>Methods</th>
<th>PACS</th>
<th>VLCS</th>
<th>Office-Home</th>
<th>Digits-DG</th>
<th>DomainNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSA Motiian et al. (2017)</td>
<td>79.40</td>
<td>70.20</td>
<td>64.90</td>
<td>74.50</td>
<td>-</td>
</tr>
<tr>
<td>SFA-A Li et al. (2023)</td>
<td>81.70</td>
<td>74.00</td>
<td>-</td>
<td>79.60</td>
<td>-</td>
</tr>
<tr>
<td>MetaReg Balan et al. (2018)</td>
<td>81.70</td>
<td>-</td>
<td>-</td>
<td>43.62</td>
<td>-</td>
</tr>
<tr>
<td>MixStyle Zhou et al. (2021)</td>
<td>83.70</td>
<td>-</td>
<td>65.50</td>
<td>-</td>
<td>34.0</td>
</tr>
<tr>
<td>JiGen Carlucci et al. (2019)</td>
<td>80.51</td>
<td>73.19</td>
<td>61.20</td>
<td>-</td>
<td>40.30</td>
</tr>
<tr>
<td>SagNet Wu et al. (2019)</td>
<td>83.25</td>
<td>-</td>
<td>62.34</td>
<td>-</td>
<td>38.90</td>
</tr>
<tr>
<td>RSC Huang et al. (2020)</td>
<td>85.15</td>
<td>75.43</td>
<td>63.12</td>
<td>-</td>
<td>38.90</td>
</tr>
<tr>
<td>DDAIG Zhou et al. (2020a)</td>
<td>83.10</td>
<td>-</td>
<td>65.50</td>
<td>77.58</td>
<td>-</td>
</tr>
<tr>
<td>L2A-OT Zhou et al. (2020b)</td>
<td>82.80</td>
<td>-</td>
<td>65.60</td>
<td>78.10</td>
<td>-</td>
</tr>
<tr>
<td>FACT Xu et al. (2021)</td>
<td>83.25</td>
<td>-</td>
<td>62.34</td>
<td>-</td>
<td>34.0</td>
</tr>
<tr>
<td>Style Neo Kang et al. (2022)</td>
<td>85.47</td>
<td>-</td>
<td>65.89</td>
<td>-</td>
<td>44.60</td>
</tr>
<tr>
<td>Liu et al. Liu et al. (2021)</td>
<td>-</td>
<td>76.48</td>
<td>67.85</td>
<td>80.02</td>
<td>-</td>
</tr>
<tr>
<td>MMD-AE Li et al. (2018b)</td>
<td>77.00</td>
<td>72.30</td>
<td>62.70</td>
<td>-</td>
<td>74.60</td>
</tr>
<tr>
<td>Cross-Grad Shankar et al. (2018)</td>
<td>80.70</td>
<td>-</td>
<td>64.40</td>
<td>75.83</td>
<td>-</td>
</tr>
<tr>
<td>MASF Dou et al. (2019)</td>
<td>81.03</td>
<td>74.11</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EISNet Wang et al. (2020a)</td>
<td>82.15</td>
<td>74.65</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MetaVIB Du et al. (2020)</td>
<td>-</td>
<td>74.54</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DGER Zhao et al. (2020)</td>
<td>-</td>
<td>74.38</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>MixUp Zhang et al. (2020)</td>
<td>83.10</td>
<td>-</td>
<td>65.50</td>
<td>-</td>
<td>38.90</td>
</tr>
<tr>
<td>ODG-N</td>
<td>90.66</td>
<td>79.85</td>
<td>72.92</td>
<td>86.75</td>
<td>50.16</td>
</tr>
</tbody>
</table>

Table 7: Model and loss ablation analysis on PACS & Office-Home datasets. (In %)

<table>
<thead>
<tr>
<th>Model variants of ODG-N</th>
<th>PACS Acc</th>
<th>H-score</th>
<th>Office-Home Acc</th>
<th>H-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>w/o ($\mathcal{F}_{in}$, $\mathcal{F}_v$, $\mathcal{F}<em>y$, $\mathcal{F}</em>\eta)$</td>
<td>63.43</td>
<td>50.82</td>
<td>53.07</td>
<td>50.55</td>
</tr>
<tr>
<td>w/o $\mathcal{A}_d$ and $\mathcal{A}_o$</td>
<td>66.01</td>
<td>52.93</td>
<td>56.58</td>
<td>53.53</td>
</tr>
<tr>
<td>w/o ($\mathcal{F}<em>{x}$)</em>{x=1}, but a common backbone for the source domains</td>
<td>63.73</td>
<td>51.53</td>
<td>53.14</td>
<td>50.50</td>
</tr>
<tr>
<td>w/o $\mathcal{F}_d$</td>
<td>64.98</td>
<td>52.09</td>
<td>55.55</td>
<td>52.69</td>
</tr>
<tr>
<td>w/o $\mathcal{F}<em>d$ and ($\mathcal{F}</em>{x}$)_{x=1}</td>
<td>62.38</td>
<td>49.67</td>
<td>51.95</td>
<td>49.02</td>
</tr>
<tr>
<td>w/o Entropy loss</td>
<td>66.81</td>
<td>53.15</td>
<td>57.35</td>
<td>54.36</td>
</tr>
<tr>
<td>w/o L_{sem}</td>
<td>65.36</td>
<td>51.73</td>
<td>55.40</td>
<td>51.60</td>
</tr>
<tr>
<td>w/o L_{sem}</td>
<td>64.74</td>
<td>51.12</td>
<td>54.57</td>
<td>51.11</td>
</tr>
<tr>
<td>w/o $\mathcal{F}<em>d$ and ($\mathcal{F}</em>{x}$)<em>{x=1} and L</em>{sem}</td>
<td>59.07</td>
<td>47.07</td>
<td>48.91</td>
<td>45.56</td>
</tr>
<tr>
<td>w/ with training $\mathcal{F}_o$ from scratch</td>
<td>65.96</td>
<td>52.86</td>
<td>56.38</td>
<td>53.33</td>
</tr>
<tr>
<td>w/ w/ L in L_{GAN}</td>
<td>66.26</td>
<td>53.96</td>
<td>58.38</td>
<td>54.93</td>
</tr>
</tbody>
</table>

Sensitivity to noise variance of GAN for synthesizing closed and pseudo-open samples

| $P_{noise}^{cs}$ $\sim$ $\mathcal{N}(0, 1)$; $P_{noise}^{pos}$ $\sim$ $\mathcal{N}(0, 1)$ | 65.30 | 52.51 | 55.38 | 52.53 |
| $P_{noise}^{cs}$ $\sim$ $\mathcal{N}(0, 2)$ | 66.35 | 53.32 | 56.44 | 53.40 |
| $P_{noise}^{cs}$ $\sim$ $\mathcal{N}(0, 3)$ | 67.29 | 54.21 | 57.18 | 54.49 |
| $P_{noise}^{cs}$ $\sim$ $\mathcal{N}(0, 4)$ | 67.98 | 55.14 | 58.29 | 55.52 |
| $P_{noise}^{cs}$ $\sim$ $\mathcal{N}(0, 5)$ | 68.22 | 55.37 | 58.72 | 56.11 |
| $P_{noise}^{cs}$ $\sim$ $\mathcal{N}(0, 0.5)$ | 67.13 | 54.22 | 55.65 | 52.24 |

ODG-N($P_{noise}^{cs}$ $\sim$ $\mathcal{N}(0, 1)$; $P_{noise}^{pos}$ $\sim$ $\mathcal{N}(0, 5)$) | 68.80 | 55.81 | 59.40 | 56.69 |

the closed-set contrastive and open-space entropy regularizer assists in generating a more discriminative feature space. The model without any of these losses or the full L_{sem} led to performance degradation by approximately 4%. Finally, we removed the diversity regularization W in L_{GAN}. Although, as per Fig. 4(d), W induces diversity in the generated images, empirically, we observed a nominal change in the accuracy (1–1.5%) in the presence of W.

Comparison of our augmentation with methods from the literature. Our augmentation technique enables more
controlled style and label mix-up, and we compared it against two types of augmentation techniques from the literature: existing style diversification approaches [Zhou et al. (2020b; 2021) for closed-set classes, and Cumix Mancini et al. (2020a), which performs random image mix-up so that the generated images can be a proxy for open-space. Our results in Fig. 4(b)-4(c) demonstrate that ODG-NET performs better with our proposed augmentation. Our method is interpolation-based, which allows us to generate more style primitives than Zhou et al. (2020b). Since our method is image-based as opposed to the feature-based method of [Zhou et al. (2021)], we can handle the semantics better. Similarly, for pseudo-open samples, we can generate more meaningful images with varied similarities to the closed classes than the random mix-up of Mancini et al. (2020a).

Sensitivity to variances of $P_{\text{noise}}^{\text{pos}}$. In this experiment, we tune the $\sigma$ parameter of $P_{\text{noise}}^{\text{pos}}$ while fixing $P_{\text{noise}}^{\text{cs}}$ for $F_d$. As shown in Table 7, we observe that as we increase $\sigma$ from 1 to 5, the model performance continuously improves from 52.51% to 55.81% for PACS and from 52.53% to 56.69% for Office-Home. With high variance, the open samples are sparsely distributed, better covering the open space. However, the performance improvements are found to saturate beyond $\sigma = 5$. On the other hand, increasing the variance of $P_{\text{noise}}^{\text{cs}}$ significantly affects the performance, leading to a drop of at least 3%. This occurs because the generated images may deviate from the original semantic concepts, degrading the quality of the generated images.

Frechet distance for domain alignment. To assess the domain independence of $F_o$, we calculate the Frechet distance Dowson & Landau (1982) between the closed-set classes of the source and target domains for Office-Home, with the target domain being Real-world. In Fig. 5(a), we show the Frechet distance of the baseline AGG, DAML, and two variants of ODG-NET, with and without the domain classifier $F_d$. The full ODG-NET produces the minimum Frechet distance, indicating that it performs the best domain alignment among the compared models. The model without $F_d$ performs poorly compared to the full ODG-NET, suggesting that the use of $F_d$ helps disentangle features better, making $F_o$ less affected by domain properties and focus on shared components.

Sensitivity to number of target open classes. Since there is no restriction on the number of open classes in the target domain, we are interested in assessing whether ODG-NET can handle different numbers of open classes during inference. Here, we considered the average leave-one-out H-score for Office-Home and simulated three scenarios with different numbers of open classes in the target: 10, 20, and 30. In Figure 5(b), ODG-NET consistently outperforms DAML Shu et al. (2021) by at least 3% for different openness factors. The entropy minimization component of $I_{\text{sen}}$ widens the gap between open and closed spaces, which is helpful in this regard. We validate this by removing the entropy component of $I_{\text{sen}}$ and re-running the experiments, in which case we find that performance drops by 2 – 3%.

Performance comparison when the source domains have a disjoint set of classes. Here, we present a novel experimental scenario in ODG, where the source domains have completely different sets of classes, and the target domain consists of all the classes from the sources, plus previously unknown class samples. We compare the performance of ODG-NET with that of DAML Shu et al. (2021) for this setup in Table 8. We find that ODG-NET outperforms DAML by around 3%, demonstrating its robustness to extreme domain and label shifts within the source domains.

Table 8: Comparison between DAML Shu et al. (2021) and ODG-NET when the source domains have mutually disjoint classes on Office-Home dataset in terms of H-score. (In %)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Clipart</th>
<th>RealWorld</th>
<th>Product</th>
<th>Art</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAML Shu et al. (2021)</td>
<td>40.12</td>
<td>58.72</td>
<td>54.85</td>
<td>47.25</td>
<td>50.23</td>
</tr>
<tr>
<td>ODG-NET</td>
<td>45.69</td>
<td>62.77</td>
<td>56.95</td>
<td>49.86</td>
<td>53.81</td>
</tr>
</tbody>
</table>

5 Takeaways

In this paper, we present ODG-NET, a solution to the challenging problem of open domain generalization. This task combines domain generalization, open-set learning, and class imbalance in a common setting. One of the key features of ODG-NET is the novel generative augmentation, which enables continuous domain and label conditional image synthesis through interpolation of conditioning variables. This augmented training set is utilized to learn a discriminative and unbiased semantic space for an open-set classifier while minimizing the effects of domain-dependent artifacts. In our experiments, ODG-NET achieves state-of-the-art performance for both open-set and closed-set domain generalization on six benchmark datasets. We plan to extend our evaluation to more safety-critical applications in the future.
References


