OSLoPrompt: Bridging Low-Supervision Challenges and Open-Set Domain Generalization in CLIP

Mohamad Hassan N C[1] Divyam Gupta[1] Mainak Singha[1] Sai Bhargav Rongali[1] Ankit Jha[2]
Muhammad Haris Khan[3] Biplab Banerjee[1]
[1]Indian Institute of Technology Bombay [2]The LNM Institute of Information Technology (LNMIIT)
[3]Mohamed Bin Zayed University of Artificial Intelligence

1. Contents of the supplementary

In the supplementary materials, we list the following information:

1. Detailed descriptions of the datasets, including known and novel class splits in Section 2.
2. List of variable names used in the paper and their purpose is detailed in Section 3.
3. Pseudo-code representation of the proposed method in Section 4.
4. Ablation study examining the impact of context length in Section 5.
5. Comprehensive literature survey on *classification with attributes in CLIP* in Section 6.
6. List of attributes and generated pseudo-open class names in Table 6 and 7 respectively.
7. Implementation details for both comparative and ablation methods in Section 8.
8. Additional visualizations of the generated pseudo-open images in Figure 1 (Section 9).
9. Expanded tables detailing domain combinations across the five datasets in Table 9, 10, 11, 12 and 13.
10. Few limitations of OSLoPrompt in Section 11.

Table 1. Summary of the datasets used.

Dataset	Images	Classes	Domains
Office-Home [32]	15,500	65	4 (Art, Clipart, Product, Real)
PACS [15]	9,991	7	4 (Artpaint, Cartoon, Sketch, Photo)
Multi-Dataset[29]	Combined	20 open	Various (Office-31 [27], STL-10[6], VisDA2017[19], DomainNet[20])
Mini-DomainNet [20]	362,470	125	4 (Clipart, Painting, Real, Sketch)
VLCS [11]	10,729	5	4 (PASCAL VOC 2007[8], Caltech[12], LabelMe[26], Sun[35])

Table 2. Known-novel class splits for the ODG settings: PACS, VLCS, Office-Home, Multi-dataset, and Mini-DomainNet datasets. Indices follow the alphabetical order of the class names.

Domain	PACS	VLCS	Office-Home	Multi-Datasets	Mini-DomainNet
Source 1	3, 0, 1	0, 1	0–14, 21–31	0–30	0–19, 40–59
Source 2	4, 0, 2	1, 2	0–8, 15–20, 32–42	1, 31–41	0–9, 20–39, 80–89
Source 3	5, 1, 2	2, 3	0–2, 9–20, 43–53	31, 33–34, 41–47	10–19, 40–49, 60–79
Target	0–6	0–4	0, 3–4, 9–10, 15–16, 21–23, 32–34, 43–45, 54–64	0, 1, 5–6, 10–11, 14, 17, 20, 26, 31–36, 39–43, 45–46, 48–67	0–4, 8–17, 25–34, 43–47, 75–79, 83–87, 90–125

2. Datasets descriptions

Office-Home Dataset [32]: The Office-Home dataset comprises **15,500 images**, carefully organized into **65 distinct classes** that span across four visually diverse domains: *Art*, *Clipart*, *Product*, and *Real*. Each domain represents a unique visual style, ranging from artistic renderings to photographic images, making the dataset highly valuable for evaluating domain adaptation and transfer learning models. This dataset is particularly suited for domain generalization, multi-domain learning, and visual recognition tasks.

PACS Dataset [10]: The PACS (Photo, Artpaint, Cartoon, Sketch) dataset consists of **9,991 images**, categorized into **seven broad classes**: *Dog*, *Elephant*, *Giraffe*, *Guitar*, *House*, *Horse*, and *Person*. These images are drawn from four distinct domains: *Artpaint*, *Cartoon*, *Sketch*, and *Photo*, representing varying styles and abstraction levels. The dataset is widely recognized for its benchmark utility in domain generalization research, especially for testing models' robustness to domain shifts.

VLCS Dataset[11]: This dataset combines images from four different datasets namely (PASCAL VOC 2007 [8], Caltech [12], LabelMe [26] and Sun [35]) consisting of images spread across five categories namely Bird, Car, Chair, Dog, and Person. We consider four categories as closed-set and the remaining category as open-set. Each of the datasets is considered as a separate domain.

Multi-Dataset [29]: The Multi-Dataset combines data from several prominent public datasets, including *Office-31* [27], *STL-10* [6], and *VisDA2017* [19]. Additionally, it incorporates four domains from *DomainNet* [20], resulting in a richly diverse dataset. This composite dataset includes **20 open classes**, intentionally absent from the joint label set of the source domains. This design facilitates tasks such as *open-set domain adaptation*, where models are challenged to handle unseen categories and cross-domain learning, providing a comprehensive benchmark for domain adaptation techniques.

Mini-DomainNet [20]: The Mini-DomainNet is a compact yet diverse subset of the DomainNet dataset, featuring images from **125 categories** across **four domains**: *Clipart*, *Painting*, *Real*, and *Sketch*. Each domain reflects a distinct visual characteristic, offering a balanced data distribution to evaluate models in multi-domain and transfer learning scenarios. This smaller-scale dataset is optimized for quick experimentation while maintaining the challenge and diversity of the full DomainNet.

Table 1 shows the dataset details, while Table 2 details the known-novel class splits, following [3, 28, 30]. In Table 2, the class names are indexed to integers alphabetically.

3. Variable names and their purpose

Table 3 details the same.

4. Pseudo-code of our training process

We detail the training process of OSLoPrompt in Algorithm 1, using the variable names from the main paper.

5. Ablation on the context lengths

In this particular section, we analyze the effect of the number of directly learnable tokens q that are introduced in addition to tokens derived from the visual prompts as given in Eq 9. As observed in Table 4, we can see that when $q = 2$, it leads to the best harmonic score. This highlights the need for the balance of the directly learnable context tokens and tokens derived from visual prompts. In Table 5, we can see that there is better H-score for the context length $\mathcal{M} = 8$ for PACS, but we follow context length 4 since it is followed majorly in the literature including [30], giving better results on majority of the datasets.

6. Literature survey on prompting with descriptions in CLIP

The classification accuracy of CLIP on downstream tasks and open-vocabulary datasets is highly influenced by the quality of text prompts [23]. Prior works have explored this sensitivity through simple handcrafted templates (*e.g.*, "a photo of a `[CLS]`") [23] or by augmenting these templates with semantically richer attributes generated by large language models (LLMs) [21].

Expanding beyond prompt engineering, methods such as LaCLIP [9], LaBo [36], and VFC [17] refine CLIP's visual-textual alignment by leveraging LLM-enriched captions to improve performance across diverse tasks and domains. Similarly, ARGUE [31] employs LLM-generated attributes, followed by attribute sampling, to enhance visual-semantic mapping. Another perspective is introduced by Kim *et al.* [14], which integrates visual attribute learning into prompts using contrastive learning.

Despite these advancements, the nuanced interplay between LLM-generated attributes and visual data remains underexplored. This is where one of the novelties of OSLoPrompt lies, bridging this gap by effectively integrating LLM-driven semantic attributes with visual cues to enhance open-set recognition and domain generalization in the low-supervision setting.

Table 3. Variables used in the OSLOPROMPT framework.

Variable	Description
Dataset and domains	
$\mathcal{D}$	Source domains.
$\mathcal{D}_s$	s^{th} source domain.
$\mathcal{X}_s, \mathcal{Y}_s$	Input images and labels in the s^{th} source domain.
$\mathcal{C}$	Combined set of classes across all the source domains.
$\mathcal{D}_t, \mathcal{X}_t, \mathcal{Y}_t$	Target domain dataset, inputs, and labels.
Target domain class definitions	
$\mathcal{Y}_t^{\text{known}}$	Known target domain classes.
$\mathcal{Y}_t^{\text{novel}}$	Novel or outlier classes.
Data augmentation	
$\mathcal{C}^{\text{open}}$	Synthesized pseudo-open class names by GPT-4o.
$\mathcal{D}^{\text{open}}$	Synthesized pseudo-open images by Stable Diffusion.
$\mathcal{D}^{\text{aug}}$	Augmented dataset: $\mathcal{D} \cup \mathcal{D}^{open}$.
Prompts and attributes	
$\mathbf{Prompt_v}$	Learnable visual prompts at the first ViT layer of $\mathcal{F}_v$.
$\mathbf{Prompt_s^{y^s}}$	Domain-specific static prompts.
$\mathcal{A}_{y^s}, \mathcal{A}'_{y^s}(x^s)$	GPT-4o generated class-wise attributes and attribute-enhanced image embeddings through cross-attention.
$\mathcal{A}''(x^s)$	Class-agnostic semantic encodings for the images.
$\overline{\mathbf{Prompt_s^{y^s}}}(x^s)$	Final dynamic domain-specific prompt conditioned on the image.
$\mathbf{Prompt}_{\text{gen}}^{\mathbf{y}}$	Domain-agnostic prompts.
Training objectives	
$\mathcal{L}_{\text{ce}}^{\text{dom-spec}}$	Supervised contrastive loss for domain-specific prompts.
$\mathcal{L}_{\text{align}}$	Context alignment loss.
$\mathcal{L}_{\text{ce}}^{\text{dom-gen}}$	Supervised contrastive loss for domain-agnostic prompts.
$\mathcal{L}_{\text{total}}$	Total loss combining all objectives.
Model components	
$\mathcal{F}_v, \mathcal{F}_t$	CLIP visual and textual encoders.
Proj_{vt}	Projector to transform the visual prompts onto the subset of context tokens of the domain-agnostic prompts.
$\mathbf{w_k}, \mathbf{w_v}, \mathbf{w_q}$	Projections for the query, key, and value for cross-attention: $\mathcal{F}_v^q = \mathcal{F}_v \mathbf{w_q}^T$, $\mathcal{F}_t^k = \mathcal{F}_t \mathbf{w_k}^T$, $\mathcal{F}_t^v = \mathcal{F}_t \mathbf{w_v}^T$.

Table 4. Ablation on PACS dataset for the number (q) of directly learnable context tokens in the domain-agnostic prompts when the total context length $\mathcal{M}$ is 4.

Number of Tokens (q)	H-score
0	94.32
1	94.74
2	**94.86**
3	93.47
4	92.15

Algorithm 1 OSLOPROMPT: Training algorithm for obtaining domain-agnostic prompts capable of solving LSOSDG

Require: Training data $\mathcal{D}$, domains $\{s\}_{s=1}^{\mathcal{N}}$, the synthesized pseudo-open dataset $\mathcal{D}^{\text{open}}$: $\mathcal{D}^{\text{aug}} = \mathcal{D} \cup \mathcal{D}^{\text{open}}$, $\mathcal{F}_v$, $\mathcal{F}_t$
Ensure: Optimize parameters $\{\nu_{1:q}\}, \mathbf{w^q}, \mathbf{w^k}, \mathbf{w^v}, \text{Proj}_{vt}, \mathbf{Prompt_v}$

Initialize $\mathbf{Prompt_v}$ for the visual prompt learning in $\mathcal{F}_v$
Construct and initialize the domain-agnostic prompt $\mathbf{Prompt}_{\text{gen}}$ as given in Eq 9
while training not converged **do**
    Sample a batch $\{(x, y)\}$ from $\mathcal{D}^{\text{aug}}$
    for $s = 1$ to $\mathcal{N}$ **do**
        Extract samples $\{(x^s, y^s)\}$ belonging to $\mathcal{D}_s$
        Initialize the domain-specific static prompt $\mathbf{Prompt_s^{y^s}}$ of $\mathcal{D}_s$ Eq 3
        Compute the attribute-enhanced embedding $\mathcal{A}'_{y^s}(x^s)$ using Eq 4 through the notion of cross attention using query-key-value formulation
        Class agnostic encoding $\mathcal{A}''(x^s)$ is computed by averaging across all the classes Eq 5
        Updated image-driven semantic attributes conditioned domain specific prompt $\overline{\mathbf{Prompt_s^{y^s}}}(x^s)$ is obtained from $\mathbf{Prompt_s^{y^s}}$ and $\mathcal{A}''(x^s)$ Eq 6
        Compute the class-posterior probability $p(y^s|x^s)$ using Eq 8 and $\mathcal{L}_{\text{ce}}^{\text{dom-spec}}$ using Eq 7
    end for
    $\mathcal{L}_{\text{align}}$ is obtained by computing cosine similarity of $\mathbf{Prompt}_{\text{gen}}$ and $\{\overline{\mathbf{Prompt_s}}\}_{s=1}^{\mathcal{N}}$ Eq 10 for the known classes in $\mathcal{C}$
    $\mathcal{L}_{\text{ce}}^{\text{dom-gen}}$ is calculated given $(x, y) \in \mathcal{D}^{\text{aug}}$ for the classes $\mathcal{C} \cup$ Unknown
    Training objectives: $\mathcal{L}_{\text{total}} \leftarrow \min_{\substack{\{\nu_{1:q}\}, \mathbf{w^q}, \mathbf{w^k}, \mathbf{w^v}, \\ \text{Proj}_{vt}, \mathbf{Prompt_v}}} \left[\mathcal{L}_{\text{align}} + \mathcal{L}_{\text{ce}}^{\text{dom-spec}} + \mathcal{L}_{\text{ce}}^{\text{dom-gen}} \right]$ (Eq. 11)
end while

Table 5. Ablation on context length for the PACS dataset[10] 1-shot setting.

Context Length ($\mathcal{M}$)	H-score
4	94.86
8	**95.44**
16	94.21

7. Class attributes, and the pseudo-open class names generated by GPT-4o

In Table 6 and Table 7 highlight the class-wise four attributes in $\mathcal{A}$ and the pseudo-open class names generated in $\mathcal{C}^{\text{open}}$, both using GPT-4o.

Table 6. Closed-set classes and their attributes generated by GPT-4o for the PACS dataset[10].

Class	Attributes
Dog	Fur, snout, tail, paw pads
Elephant	Large ears, trunk, tusks, wrinkled skin
Giraffe	Long neck, spotted pattern, horns, slender legs
Guitar	Curved body, strings, fretboard, soundhole
Horse	Mane, hooves, muscular build, tail
House	Roof, windows, doors, chimney

8. Implementation details for both comparative and ablation methods

We evaluate our proposed methods against several state-of-the-art approaches using their official implementations, incorporating necessary modifications to ensure compatibility with 1-shot and 5-shot settings. For **PromptSRC** [13] and **STYLIP** [4], we extend their frameworks by integrating synthetic open samples as outlined in our method and employing an "unknown" prompt. These enhancements refine the original designs to more effectively address open-set scenarios. Similarly,

Table 7. Fine-grained pseudo-open-set class names vs. closed-set class names for the PACS dataset[10] generated by GPT-4o.

Closed-set Classes	Related Fine-Grained Pseudo Open-Set Classes Outputted by GPT-4o
Dog	Wolf, Fox, Coyote, Jackal, Dhole, Fennec Fox, Hyena, Maned Wolf
Elephant	Mastodon, Woolly Mammoth, Rhinoceros, Hippopotamus
Giraffe	Okapi, Pronghorn, Impala, Sable Antelope, Kudu, Eland, Gazelle, Springbok, Nyala, Gerenuk
Guitar	Mandolin, Banjo, Lute, Bouzouki, Sitar, Balalaika, Charango, Oud, Lyre, Zither
Horse	Zebra, Donkey, Onager, Kiang, Tarpan, Wild Ass, Quagga
House	Castle, Hut, Palace
Additional Fine-Grained Pseudo Open-Set Classes: Alpaca, Emu, Lynx, Peacock, Ferret, Armadillo, Pangolin, Tamarin, Mongoose, Marten, Caracal, Serval, Ocelot, Civet, Quokka, Wallaby, Pademelon, Koala, Pika, Aye-aye, Tarsier, Wombat, Kinkajou, Agouti, Coati, Cuscus, Galago, Jerboa, Marmoset	

for **CLIP+OpenMax** [2], we adapt the approach by computing Mean Activation Vectors (MAVs) for each class using a modified data loader and optimizing thresholds for improved open-set recognition accuracy.

For **MORGAN** [18] and **2LM** [22], we implement meta-learning strategies on the dataset $\mathcal{D}$, modifying the backbone for **MORGAN** and incorporating OpenMax for open-set recognition in **2LM**, analogous to **CLIP+OpenMax**. Meta-training is conducted over 30 episodes for both methods, during which convergence was observed. Methods such as **StyLIP** [4] and **ODG-Net** [3] are evaluated using the official implementations provided by their authors. All methods are trained for the default number of epochs specified in their respective implementations.

For the ImageNet experiments, we extend **OSLoPrompt** by introducing additional domain-specific prompts selected from ImageNet templates provided in the official CLIP implementation [24]. To ensure consistency, we use a ViT-B/32 backbone across all models. Optimization is performed using the SGD optimizer with a learning rate of 0.0035 over six epochs. These adjustments ensure robust and fair comparisons across methods, emphasizing adaptability in joint open-set and low-shot domain generalization (DG) scenarios. All comparative methods are evaluated in the LSOSDG setting for consistency.

In the ablation experiments, we generate pseudo-open samples using a Mix-up-based [16] approach, with λ uniformly sampled from $[0.3, 0.7]$. For two samples, $x_i^s \in \mathcal{X}_s$ and $x_j^{s'} \in \mathcal{X}_{s'}$, from source domains s and s', respectively, the generated pseudo-open sample $x^{\text{open}} \in \mathcal{D}^{\text{open}}$ is defined as:

$$x^{\text{open}} = \lambda x_i^s + (1 - \lambda) x_j^{s'}$$

where $\lambda \in [0.3, 0.7]$. For manual domain-specific prompting, ad hoc attributes are created by concatenating the four attributes of each class along with the class name. For instance, the attributes of the dog class, as shown in Table 6, are "Fur, snout, tail, paw pad." The resulting manual prompt for the dog class becomes: "A {domain} of dog with Fur, snout, tail, paw pad."

For the image-conditioning experiments, projected image features from the Meta-Net are incorporated into the prompts before being passed to the CLIP [24] text encoder. The Meta-Net consists of two linear layers with an intermediate representation size of 32, with a ReLU activation [1] applied between the layers. Additionally, we evaluate two types of domain-generic prompting: (1) textual prompting and (2) textual prompting combined with image conditioning, inspired by **CoOp** [38] and **CoCoOp** [37], respectively. In both cases, the context length is fixed at 4.

9. Generated pseudo-open samples by ODG-CLIP [30] and OSLoPrompt

Fig. 1 compares the pseudo-open images generated by the method of ODG-CLIP, which are mostly coarse-grained, and OSLoPrompt, which are fine-grained in nature, given the closed-set images. In Table 8, we compare the FID distance [7] between the closed set images and the generated pseudo-open samples. The proposed fine-grained pseudo-open samples are highly similar to closed-set samples when compared to ODG-CLIP's synthesized pseudo-open samples. The FID score in our case is low by 4.9 points than that of ODG-CLIP.

10. Detailed results on all the datasets

We report the detailed results for all the domain combinations for the five datasets in Table 9 - 13. In the leave-one-domain-out protocol, all but one domains are considered as sources, while the rest acts as the target domain.

Figure 1. Pseudo open images generated by [30] and OSLoPrompt, given the known-class images, for PACS.

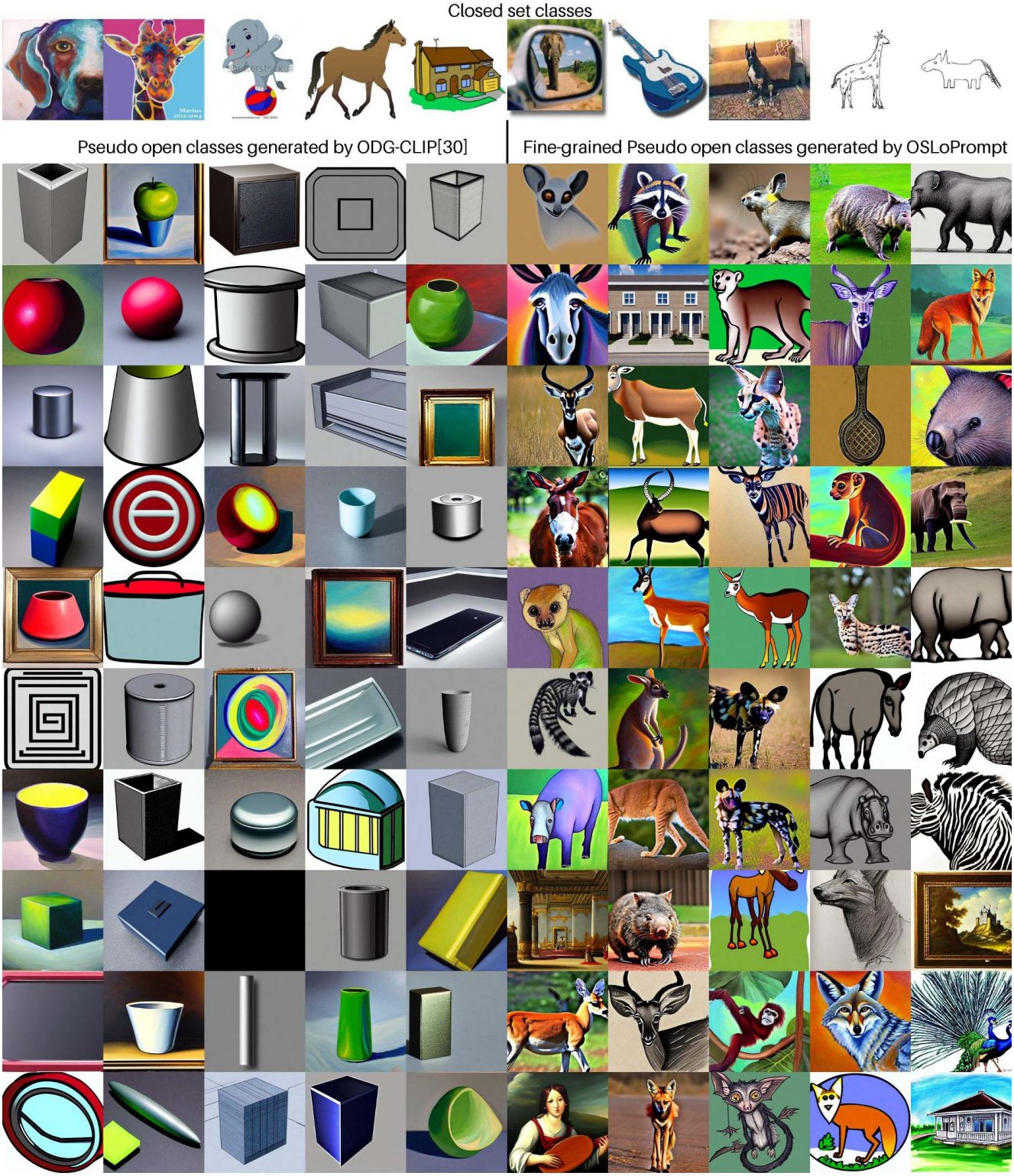

Table 8. FID between the closed-set images and generated pseudo-open images for the PACS dataset.

Pseudo-open image synthesis method	FID score
ODG-CLIP[30] pseudo open samples	13.04
Fine-grained pseudo open samples (ours)	**8.14**

Table 9. Accuracy over different target domains in the Mini-DomainNet dataset. The 1-shot results are at the top, and the 5-shot results are at the bottom. For each case, the other domains are considered as the source domains.

Method	Clipart		Painting		Sketch		Real		Avg Acc	Avg H-score
	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score		
CLIP + OpenMax [2]	20.50	14.28	19.50	32.42	7.69	32.42	20.00	33.08	16.92	28.05
CLIPN [33]	47.43	40.68	50.81	39.56	42.60	39.91	49.67	43.50	47.63	40.91
MORGAN [18]	26.59	11.28	14.16	16.95	28.41	23.34	20.44	11.25	22.40	15.70
STYLIP [4]	62.48	59.45	60.78	56.15	47.24	43.72	67.25	70.51	59.44	57.46
PROMPTSRC [13]	27.38	19.71	26.91	25.54	20.24	21.72	26.25	14.79	25.20	20.44
2LM [22]	28.06	18.71	23.94	19.62	21.36	20.88	24.65	11.78	24.50	17.75
ODG-Net [3]	21.76	17.50	24.42	27.99	21.98	11.44	20.04	19.40	22.05	19.08
MEDIC [34]	26.53	21.99	24.63	24.08	19.84	20.67	23.93	9.48	23.73	19.05
SCI-PD [5]	17.50	22.95	15.50	23.20	16.00	23.12	16.00	24.04	16.25	23.33
ODG-CLIP [30]	66.84	**76.10**	54.74	66.25	59.47	**60.55**	63.16	59.12	61.05	65.50
OSLOPROMPT	**76.00**	66.00	**61.50**	**66.43**	**60.00**	59.15	**78.50**	**78.69**	**69.00**	**67.57**
CLIP + OpenMax [2]	31.50	47.17	30.00	45.54	32.82	48.93	35.50	51.17	32.46	48.20
CLIPN [33]	53.39	44.65	58.53	53.04	55.50	50.91	55.71	46.31	55.78	48.53
MORGAN [18]	32.10	21.15	36.26	29.31	35.26	32.32	47.62	25.45	37.81	27.06
STYLIP [4]	64.12	60.03	62.52	60.65	53.87	44.12	75.61	77.93	64.03	60.68
PROMPTSRC [13]	37.68	32.38	33.21	32.00	39.80	31.04	34.80	29.86	36.37	31.32
2LM [22]	38.55	28.27	34.34	25.47	35.19	35.10	45.65	25.94	38.43	28.70
ODG-Net [3]	45.69	30.53	38.53	20.86	39.71	23.41	35.88	20.08	39.95	23.72
MEDIC [34]	39.87	33.08	30.42	28.23	35.83	32.16	41.68	28.94	36.95	30.60
SCI-PD [5]	19.50	28.23	23.50	34.60	15.50	24.18	26.50	35.28	21.25	30.57
ODG-CLIP [30]	79.00	52.12	60.53	58.54	**78.97**	**88.03**	79.00	63.28	74.38	65.49
OSLOPROMPT	**86.00**	**63.68**	**63.00**	**66.06**	63.58	62.03	**85.50**	**74.53**	**74.52**	**66.58**

11. Potential limitations

We find two potential areas of improvements for OSLOPROMPT, as discussed in the following,

1. **Challenges with highly fine-grained datasets:** In fine-grained datasets, where differences between classes are subtle, generating meaningful pseudo-open samples is tricky, and may require more insights in our prompting scheme.
2. **Impact of pseudo-open image quality:** There is dependence on the Stable diffusion [25] model to generate pseudo-open samples. When the prompts are fine-grained and specific, there is a chance that the model can introduce artifacts unrelated to the object in the image.

Table 10. Accuracy over different target domains in the Multi-dataset benchmark. The 1-shot results are shown at the top and the 5-shot results at the bottom. For each case, the other domains are considered as the source domains.

Method	Clipart		Painting		Sketch		Real		Avg Acc	Avg H-score
	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score		
CLIP + OpenMax [2]	10.00	18.17	16.15	27.80	12.79	22.67	8.02	14.85	11.74	20.87
CLIPN [33]	41.51	37.31	34.53	34.48	36.38	34.49	46.92	38.82	39.84	36.28
MORGAN [18]	31.66	46.72	26.76	36.43	23.48	31.91	38.08	33.96	30.00	37.26
STYLIP [4]	56.32	50.22	49.10	45.16	39.92	35.28	60.65	59.89	51.50	47.64
PROMPTSRC [13]	27.38	25.06	30.51	31.82	29.10	31.97	33.65	35.85	30.16	31.18
2LM [22]	28.24	38.49	27.47	32.13	31.19	32.19	32.00	36.37	29.73	34.80
ODG-Net [3]	25.88	25.48	24.05	35.85	25.80	27.46	40.92	28.79	29.16	29.40
MEDIC [34]	31.62	34.15	26.31	30.99	27.71	31.29	35.75	36.00	30.35	33.11
SCI-PD [5]	15.37	11.78	17.62	21.48	20.52	24.88	14.30	18.56	16.95	19.18
ODG-CLIP [30]	66.42	71.71	55.25	65.09	58.21	**65.71**	75.08	75.60	63.74	69.53
OSLOPROMPT	**79.04**	**78.18**	**66.06**	**69.49**	**72.55**	62.74	**87.55**	**87.55**	**76.30**	**74.49**
CLIP + OpenMax [2]	48.04	63.09	62.63	73.06	51.72	65.22	63.97	73.97	56.59	68.84
CLIPN [33]	42.28	38.18	44.76	39.56	47.25	36.32	51.69	42.86	46.50	39.23
MORGAN [18]	35.65	45.48	38.28	44.86	30.48	34.62	37.48	46.26	35.47	42.80
STYLIP [4]	59.70	55.58	52.05	46.25	45.13	37.00	61.43	60.21	54.58	49.76
PROMPTSRC [13]	34.44	36.36	31.50	35.52	38.40	39.75	38.39	40.86	35.68	38.12
2LM [22]	35.31	37.22	34.74	34.43	31.68	31.15	38.44	38.72	35.04	35.38
ODG-Net [3]	29.62	32.93	28.59	33.63	32.41	39.75	46.16	41.41	34.20	36.93
MEDIC [34]	32.24	34.92	34.05	36.31	36.00	34.05	39.38	39.74	35.42	36.26
SCI-PD [5]	27.58	28.05	31.26	22.56	39.41	38.49	30.54	26.46	32.20	28.89
ODG-CLIP [30]	67.65	73.33	71.63	**76.26**	**75.60**	71.52	82.71	83.46	74.40	76.14
OSLOPROMPT	**83.21**	**78.69**	**73.12**	75.36	71.86	**76.13**	**90.81**	**90.00**	**79.75**	**80.05**

Table 11. Accuracy and H-score over different target domains in the Office-Home dataset. The 1-shot results are shown at the top and the 5-shot results at the bottom. For each case, the other domains are considered as the source domains.

Method	Clipart		Product		Real World		Art		Avg Acc	Avg H-score
	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score		
CLIP + OpenMax [2]	15.91	31.45	18.70	31.45	25.62	40.58	19.74	27.09	19.99	32.64
CLIPN [33]	48.52	36.07	40.41	31.50	44.50	31.70	43.27	32.06	44.18	32.83
MORGAN [18]	24.88	16.73	4.59	14.95	30.28	24.17	17.07	18.19	19.21	18.51
STYLIP [4]	42.35	20.37	60.00	15.79	62.51	34.13	44.49	17.52	52.34	21.95
PROMPTSRC [13]	22.37	17.44	14.30	12.93	30.10	14.10	21.30	14.94	22.02	14.85
2LM [22]	27.88	13.47	10.12	14.95	30.28	13.61	16.93	12.35	21.30	13.60
ODG-Net [3]	31.30	11.44	1.89	10.89	26.21	13.27	22.48	10.18	20.47	11.45
MEDIC [34]	26.70	11.25	10.83	11.29	28.68	11.64	19.03	12.80	21.31	11.75
SCI-PD [5]	25.00	34.51	29.20	38.44	48.76	56.48	38.10	47.79	35.27	44.31
ODG-CLIP [30]	46.21	54.72	50.41	55.30	58.68	46.87	39.47	54.83	48.69	52.93
OSLOPROMPT	**59.09**	**63.38**	**81.30**	**67.35**	**79.33**	**70.48**	**59.21**	**54.96**	**69.73**	**64.04**
CLIP + OpenMax [2]	29.55	44.32	29.61	44.39	30.58	42.76	52.63	65.64	35.59	49.28
CLIPN [33]	44.22	40.51	46.81	40.46	49.38	40.47	51.34	39.87	47.94	40.33
MORGAN [18]	40.07	17.62	30.93	11.12	37.63	29.23	36.17	16.55	36.20	18.63
STYLIP [4]	50.62	42.91	65.32	40.28	78.42	51.98	53.11	34.67	61.87	42.46
PROMPTSRC [13]	32.30	20.82	30.20	22.08	28.21	16.98	33.70	21.51	31.10	20.35
2LM [22]	35.97	11.23	28.33	14.91	28.01	25.35	25.22	24.32	29.38	18.95
ODG-Net [3]	43.65	22.11	36.01	12.59	28.74	16.05	29.68	13.08	34.52	15.96
MEDIC [34]	35.27	16.11	27.42	16.31	31.66	20.59	27.24	20.77	30.40	18.45
SCI-PD [5]	25.76	34.97	35.77	41.26	45.40	**57.40**	25.00	36.36	32.98	42.50
ODG-CLIP [30]	45.45	47.90	73.17	**81.22**	71.07	30.61	31.57	37.50	55.32	49.31
OSLOPROMPT	**61.36**	**57.68**	**91.86**	66.81	**80.99**	57.03	**67.10**	**66.80**	**75.33**	**62.08**

Table 12. Accuracy and H-score across target domains in the VLCS dataset. The 1-shot results are shown at the top and the 5-shot results at the bottom. For each case, the other domains are considered as the source domains.

Method	CALTECH		SUN09		VOC2007		LABELME		Avg Acc	Avg H-score
	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score		
CLIP + OpenMax [2]	22.94	37.32	2.08	4.08	21.42	34.85	35.92	51.08	20.59	31.83
CLIPN [33]	25.70	19.14	21.43	19.16	27.40	18.98	26.83	20.66	25.34	19.49
MORGAN [18]	32.40	25.58	22.81	31.12	39.77	28.26	30.44	23.91	31.35	27.22
STYLIP [4]	11.04	19.23	21.45	31.17	25.75	35.54	53.52	52.48	27.94	34.61
PROMPTSRC [13]	26.28	22.25	29.58	27.59	19.34	12.07	24.71	18.26	24.98	20.04
2LM [22]	33.09	29.65	22.70	32.38	37.37	27.01	33.28	26.01	31.61	28.76
ODG-Net [3]	34.93	27.62	21.76	33.37	41.08	30.25	31.57	25.42	32.33	29.17
MEDIC [34]	33.41	27.37	27.86	27.08	35.62	23.80	34.88	26.86	32.94	26.28
SCI-PD [5]	22.63	23.03	26.42	25.77	11.90	12.13	18.56	17.47	19.88	19.60
ODG-CLIP [30]	80.62	87.25	54.10	50.11	52.95	50.61	22.05	30.84	52.43	54.70
OSLOPROMPT	**99.47**	**99.73**	**62.75**	**66.60**	**78.02**	**79.55**	**75.30**	**61.69**	**78.89**	**76.89**
CLIP + OpenMax [2]	77.98	87.63	55.35	67.57	64.41	75.17	67.25	68.60	66.25	74.74
CLIPN [33]	27.33	27.07	35.46	29.75	36.52	26.99	32.38	28.00	32.92	27.95
MORGAN [18]	34.59	40.39	48.41	39.91	39.39	39.50	46.27	35.01	42.16	38.70
STYLIP [4]	46.01	55.80	41.25	45.47	42.09	49.02	53.75	45.44	45.78	48.93
PROMPTSRC [13]	36.95	29.57	35.18	32.19	34.27	33.91	38.24	33.77	36.16	32.36
2LM [22]	37.22	38.42	44.73	37.52	38.03	36.79	46.71	36.69	41.67	37.36
ODG-Net [3]	36.33	40.18	52.73	38.70	38.95	35.51	44.20	36.46	43.05	37.71
MEDIC [34]	35.53	36.20	44.33	36.31	40.17	35.95	42.09	34.76	40.53	35.56
SCI-PD [5]	30.24	30.77	32.71	34.01	29.16	28.29	28.34	28.84	30.11	30.48
ODG-CLIP [30]	76.96	86.86	**70.82**	42.43	47.56	48.22	56.39	50.06	62.93	56.89
OSLOPROMPT	**98.95**	**99.39**	64.02	**68.26**	**76.86**	**76.86**	**76.33**	**64.84**	**79.04**	**77.34**

Table 13. Accuracy and H-score across target domains in the PACS dataset. The 1-shot results are shown at the top and the 5-shot results are at the bottom. For each case, the other domains are considered as the source domains.

Method	Art Painting		Photo		Sketch		Cartoon		Avg Acc	Avg H-score
	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score		
CLIP + OpenMax [2]	33.83	49.78	14.70	25.63	5.00	9.52	27.44	42.95	20.24	31.97
CLIPN [33]	63.34	54.15	65.46	53.62	63.80	56.84	63.50	58.55	64.03	55.79
MORGAN [18]	28.92	0.94	44.35	21.44	32.22	31.96	44.13	21.91	37.40	19.06
STYLIP [4]	73.17	78.96	73.10	77.68	73.49	43.85	79.78	43.47	74.89	60.99
PROMPTSRC [13]	30.15	17.47	35.83	30.98	37.42	33.05	39.47	26.84	35.72	27.09
2LM [22]	32.53	9.57	36.37	24.89	31.54	30.88	40.45	20.35	35.22	21.42
ODG-Net [3]	36.67	9.00	29.73	12.49	32.56	35.56	40.32	29.61	34.82	21.67
MEDIC [34]	29.47	11.12	34.33	20.61	35.54	28.84	36.28	25.03	33.91	21.40
SCI-PD [5]	26.30	28.75	19.50	22.07	22.31	24.88	25.47	27.65	23.40	25.84
ODG-CLIP [30]	51.34	62.53	59.05	73.28	82.30	87.93	82.88	78.51	68.89	75.56
OSLOPROMPT	**91.61**	**93.43**	**99.43**	**99.71**	**82.75**	**92.70**	**97.06**	**93.59**	**92.71**	**94.86**
CLIP + OpenMax [2]	63.79	77.55	74.56	85.26	60.47	74.83	76.17	86.29	68.75	80.98
CLIPN [33]	78.10	69.36	78.27	71.89	77.41	71.88	78.36	71.42	78.04	71.14
MORGAN [18]	50.39	33.03	38.52	29.40	45.84	8.95	50.34	24.86	46.27	24.06
STYLIP [4]	75.24	79.67	87.26	88.31	74.45	50.78	83.45	61.27	80.10	70.01
PROMPTSRC [13]	50.71	36.59	48.53	32.53	41.35	22.46	46.84	29.32	46.86	30.23
2LM [22]	51.79	27.75	42.98	28.11	41.79	12.15	50.25	28.23	46.70	24.06
ODG-Net [3]	42.55	32.99	49.07	21.24	49.63	17.72	45.37	31.71	46.66	25.92
MEDIC [34]	48.11	30.37	46.49	27.34	39.65	15.50	45.28	27.00	44.88	25.05
SCI-PD [5]	35.16	36.79	32.48	30.70	35.73	34.45	37.28	36.18	35.16	34.53
ODG-CLIP [30]	82.23	87.13	93.46	96.29	72.09	81.03	86.80	88.19	83.65	88.16
OSLOPROMPT	**92.18**	**94.26**	**99.60**	**99.80**	**85.41**	**93.50**	**97.67**	**92.49**	**93.72**	**95.01**

References

[1] AF Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018. 5

[2] Abhijit Bendale and Terrance E Boult. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1563–1572, 2016. 5, 7, 8, 9, 10, 11

[3] Shirsha Bose, Ankit Jha, Hitesh Kandala, and Biplab Banerjee. Beyond boundaries: A novel data-augmentation discourse for open domain generalization. *Transactions on Machine Learning Research*, 2023. 2, 5, 7, 8, 9, 10, 11

[4] Shirsha Bose, Ankit Jha, Enrico Fini, Mainak Singha, Elisa Ricci, and Biplab Banerjee. Stylip: Multi-scale style-conditioned prompt learning for clip-based domain generalization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5542–5552, 2024. 4, 5, 7, 8, 9, 10, 11

[5] Zining Chen, Weiqiu Wang, Zhicheng Zhao, Fei Su, Aidong Men, and Hongying Meng. Practicaldg: Perturbation distillation on vision-language models for hybrid domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23501–23511, 2024. 7, 8, 9, 10, 11

[6] Adam Coates, Andrew Ng, and Honglak Lee. An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 215–223. JMLR Workshop and Conference Proceedings, 2011. 1, 2

[7] DC Dowson and BV666017 Landau. The fréchet distance between multivariate normal distributions. *Journal of multivariate analysis*, 12(3):450–455, 1982. 5

[8] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, 2015. 1, 2

[9] Lijie Fan, Dilip Krishnan, Phillip Isola, Dina Katabi, and Yonglong Tian. Improving clip training with language rewrites. *Advances in Neural Information Processing Systems*, 36, 2024. 2

[10] Chen Fang, Ye Xu, and Daniel N Rockmore. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1657–1664, 2013. 2, 4, 5

[11] Chen Fang, Ye Xu, and Daniel N. Rockmore. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2013. 1, 2

[12] Li Fei-Fei, R. Fergus, and P. Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pages 178–178, 2004. 1, 2

[13] Muhammad Uzair Khattak, Syed Talal Wasim, Muzammal Naseer, Salman Khan, Ming-Hsuan Yang, and Fahad Shahbaz Khan. Self-regulating prompts: Foundational model adaptation without forgetting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15190–15200, 2023. 4, 7, 8, 9, 10, 11

[14] Gahyeon Kim, Sohee Kim, and Seokju Lee. Aapl: Adding attributes to prompt learning for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1572–1582, 2024. 2

[15] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M Hospedales. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE international conference on computer vision*, pages 5542–5550, 2017. 1

[16] Massimiliano Mancini, Zeynep Akata, Elisa Ricci, and Barbara Caputo. Towards recognizing unseen categories in unseen domains. In *European Conference on Computer Vision*, pages 466–483. Springer, 2020. 5

[17] Liliane Momeni, Mathilde Caron, Arsha Nagrani, Andrew Zisserman, and Cordelia Schmid. Verbs in action: Improving verb understanding in video-language models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 15579–15591, 2023. 2

[18] Debabrata Pal, Shirsha Bose, Biplab Banerjee, and Yogananda Jeppu. Morgan: Meta-learning-based few-shot open-set recognition via generative adversarial network. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6295–6304, 2023. 5, 7, 8, 9, 10, 11

[19] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*, 2017. 1, 2

[20] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1406–1415, 2019. 1, 2

[21] Sarah Pratt, Ian Covert, Rosanne Liu, and Ali Farhadi. What does a platypus look like? generating customized prompts for zero-shot image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15691–15701, 2023. 2

[22] Xiaorong Qin, Xinhang Song, and Shuqiang Jiang. Bi-level meta-learning for few-shot domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15900–15910, 2023. 5, 7, 8, 9, 10, 11

[23] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019. 2

[24] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021. 5

[25] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. 7

[26] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman. Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision*, 77:157–173, 2008. 1, 2

[27] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010. 1, 2

[28] 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

[29] Yang Shu, Zhangjie Cao, Chenyu Wang, Jianmin Wang, and Mingsheng Long. Open domain generalization with domain-augmented meta-learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9619–9628, 2021. 1, 2

[30] Mainak Singha, Ankit Jha, Shirsha Bose, Ashwin Nair, Moloud Abdar, and Biplab Banerjee. Unknown prompt the only lacuna: Unveiling clip's potential for open domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13309–13319, 2024. 2, 5, 6, 7, 8, 9, 10, 11

[31] Xinyu Tian, Shu Zou, Zhaoyuan Yang, and Jing Zhang. Argue: Attribute-guided prompt tuning for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28578–28587, 2024. 2

[32] 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*, pages 5018–5027, 2017. 1

[33] Hualiang Wang, Yi Li, Huifeng Yao, and Xiaomeng Li. Clipn for zero-shot ood detection: Teaching clip to say no. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1802–1812, 2023. 7, 8, 9, 10, 11

[34] Xiran Wang, Jian Zhang, Lei Qi, and Yinghuan Shi. Generalizable decision boundaries: Dualistic meta-learning for open set domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11564–11573, 2023. 7, 8, 9, 10, 11

[35] Jianxiong Xiao, James Hays, Krista A. Ehinger, Aude Oliva, and Antonio Torralba. Sun database: Large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3485–3492, 2010. 1, 2

[36] Yue Yang, Artemis Panagopoulou, Shenghao Zhou, Daniel Jin, Chris Callison-Burch, and Mark Yatskar. Language in a bottle: Language model guided concept bottlenecks for interpretable image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19187–19197, 2023. 2

[37] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16816–16825, 2022. 5

[38] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9):2337–2348, 2022. 5