Supplementary Materials: Unknown Prompt, the only Lacuna: Unveiling CLIP's Potential in Open Domain Generalization

Mainak Singha[1†] Ankit Jha[2] Shirsha Bose[3] Ashwin Nair[4]
Moloud Abdar[5] Biplab Banerjee[2]
[1] Aisin Corporation, Japan [2] IIT Bombay, India
[3] TU Munich, Germany [4] IISER Thiruvananthapuram, India [5] Deakin University, Australia
{mainaksingha.iitb, ankitjha16, shirshabosecs, ashwin9084yt, m.abdar1987, getbiplab}@gmail.com

1. Contents of the supplementary materials

In this supplementary document, we present detailed information and further experimental results, including:

1. **Dataset Splits for ODG Settings**: Table 1 lists the dataset splits for PACS, VLCS, OfficeHome, DigitDG, Multi-Dataset, and Mini-DomainNet.

2. **Extended Literature Survey on Prompt Learning**: An expanded review of prompt learning in CLIP is available in Section 3.

3. **Implementation Details of Competitors**: Section 4 elaborates on how competitor models were implemented.

4. **Analysis of Fréchet Distance**: In Table 2, we analyze the Fréchet distance [9] between each source and target domain in the PACS dataset to evaluate domain alignment.

5. **Model Complexity Comparison (GFLOPS)**: Figure 1 compares different models based on their GFLOPS calculation during training.

6. **Ablation Studies**: These include an examination of the domain token position in prompts (Table 3), context length for prompts (Table 4), and cosine-similarity of $\hat{x}$ features for pseudo-unknown-class samples across domains (Table 5).

7. **Qualitative Analysis**: Figure 2 highlights the effect of utilizing negative prompts for creating pseudo-open images. Additionally, Figure 3 presents a t-SNE visualization, contrasting our method's latent visual space representation with the traditional hand-crafted $\hat{x}$ for class embeddings. Furthermore, Figure 4 offers a comparative analysis of open samples generated by Cumix [30], OpenGAN [23], and our diffusion model [38] within the embedding space.

8. **Model Ablation Results**: Table 6 shows results for ODG-CLIP using ViT/B-16 and ResNet-50-based CLIP visual encoders.

9. **Extended Results with Unknown-Class Prompts**: Table 7 extends the (model+SD) results from Table 1 in the main paper.

10. **ODG Results on Full DomainNet**: Table 8 provides detailed results and comparisons for the full DomainNet dataset [34].

11. **Individual Domain Combination Results**: Detailed results for individual domain combinations of open and closed-set DG, supplementing Tables 1 and 2 in the main paper, are presented in Tables 9 through 16.

[†]This work is partially done while studying at IIT Bombay, India

2. Datasets descriptions

Office-Home Dataset [45]: Comprising 15,500 images, this dataset is divided into 65 classes across four domains: Art, Clipart, Product, and Real. **PACS** Dataset [27]: The PACS dataset includes 9,991 images, categorized into seven classes and spread over four domains: Artpaint, Cartoon, Sketch, and Photo. **VLCS** Dataset [11]: This dataset amalgamates images from four classification datasets (PASCAL VOC 2007 [10], Caltech [12], LabelMe [39], Sun [48]) and consists of images across five classes: Bird, Car, Chair, Dog, and Person. **Digits-DG** Dataset [52]: Digits-DG is an aggregation of several handwritten digit recognition datasets, including MNIST [24], MNIST-M [13], SVHN [31], and SYN [13]. **Multi-dataset** [42]: This dataset combines various public datasets such as Office-31 [40], STL-10 [8], and Visda2017 [35], including four domains from DomainNet [34]. It features 20 open classes not present in the source domains' joint label set. **Mini-domainnet** [34]: This dataset features four domains, each comprising images from 125 categories. **Domainnet** [34]: Comprising six domains, this dataset includes images from 345 categories. The class splits for all five datasets used in ODG are detailed in Table 1, with classes arranged in alphabetical order.

Table 1. Dataset splits for the ODG settings: PACS, VLCS, OfficeHome (O.H.), DigitDG (D-DG), Multi-dataset(M.Data), Mini-DomainNet (M.DNet) and DomainNet datasets.

Domain	**PACS**	**VLCS**	**OfficeHome**	**Digits-DG**	**Multi-Datasets**	**Mini-DomainNet**	**DomainNet**
Source 1	3, 0, 1	0, 1	0 - 14,21 - 31	0, 1, 2	0 - 30	0 - 19, 40 - 59	0 - 19, 30 - 59, 70 - 99
Source 2	4, 0, 2	1, 2	0 - 8, 15 - 20, 32 - 42	2, 3, 4	1, 31 - 41	0 - 9, 20 - 39, 80 - 89	10 - 49, 90 - 129
Source 3	5, 1, 2	2, 3	0 - 2, 9 - 20, 43 - 53	4, 5, 6	31, 33, 34, 41 - 47	10 - 19, 40 - 49, 60 - 79	60 - 79, 140 - 164 180 - 194, 210 - 229
Source 4	-	-	-	-	-	-	130 - 139, 160 - 184, 195 - 219, 250 - 269
Source 5	-	-	-	-	-	-	20 - 39, 220 - 249, 270 - 299
Target	0-6	0-4	0, 3 - 4, 9 - 10, 15 - 16, 21 - 23, 32 - 34, 43 - 45, 54 - 64	0-9	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	0 - 9, 70 - 79 120 - 129, 180 - 189 230 - 239, 280 - 289 300 - 344

3. Extended literature survey of prompt learning using CLIP

Vision-Language Models (VLMs) have garnered significant interest across language processing and computer vision fields [3, 5, 16, 25, 37, 43, 44]. These models typically employ task-specific textual descriptions to interpret and analyze visual data [17, 19]. While early prompting strategies relied on manual definitions, more recent developments have shifted towards automated prompt learning. CoOp [51] introduces an approach to optimize both unified and class-specific prompts via back-propagation. CoCoOp [50] further expands on CoOp by incorporating input-conditioned prompt learning, thus addressing issues related to generalization. The CLIP-adapter [15] innovates by fine-tuning feature adapters within both the visual and language branches of the model. ProGrad [54] is designed to prevent the forgetting of foundational knowledge within these models. TPT [41] leverages the consistency between multiple views of an image for supervision. Probabilistic and variational models such as Prod [28] and Varprompt [29] focus on learning prompt distributions that align with the spread of visual features. LASP [6] enhances the quality of learned prompts through a text-to-text cross-entropy loss. Meanwhile, MaPLe [21] works on improving the compatibility between different levels of CLIP encoders. However, a notable limitation of these approaches is their lack of specialization in handling multi-domain data, a crucial aspect for broader applicability in diverse real-world scenarios.

In the realm of domain generalization, several researchers have investigated the concept of domain invariant prompts. For instance, [32] and [26] focus on harnessing text-based source domain knowledge or utilizing image patches as prompt inputs in Vision Transformer (ViT) models. This approach is akin to the methodology used in VPT [20], where prompts are adapted based on specific image features, aiming to achieve a more domain-agnostic model performance. DPL [49] employs CLIP [36] for multi-source Domain Generalization (DG) by deducing domain information from visual features on a batch-wise basis. However, DPL does not fully exploit CLIP's capability to discern domain-specific details. Additionally, it is prone to overfitting when dealing with small batches, as accurately estimating unbiased style characteristics becomes challenging.

As can be observed, our prompt learning technique stands out from all the previous literature.

4. Additional implementation details of the competitor models

In the CLIP+OpenMax configuration, we have developed a $\mathcal{C}+1$-class, threshold-free classifier using CLIP features to form a unified classifier. For the CLIP+OSDA variant, we incorporate a trainable linear layer on top of the pre-trained CLIP features, which acts as the generator. This is complemented by distinct discriminators for both source-specific classification and domain alignment. The adversarial aspect of this setup is implemented through a gradient-reversal layer, following the methodology outlined in [14].

Regarding other prompt learning techniques, our implementation is faithful to the procedures described in the original works. For the CLIPN+StyLIP model, we divide the tokens into two separate categories. One category is shaped by the token learning strategy of StyLIP, and the other consists of specialized tokens that are modified in line with CLIPN's framework. This bifurcated token strategy effectively combines the strengths of both StyLIP and CLIPN, ensuring a harmonious and potent integration of these methodologies.

5. Analysis of domain alignment using the Fréchet distance [9]

Table 2 presents the source-to-target domain alignment in various PACS dataset combinations, using the Fréchet distance as a metric. A lower Fréchet distance denotes better domain alignment. In these evaluations, ODG-CLIP demonstrates significant superiority over two main competitors: DAML [42], employing a traditional CNN backbone, and the combined model of CLIPN + StyLIP, using baseline CLIP [36] features. This advantage of ODG-CLIP is evidenced by its smaller Fréchet distances, indicating more effective domain alignment. Additionally, the impact of excluding the consistency loss $\mathcal{L}_{sem}$ from ODG-CLIP is shown, revealing a decrease in alignment quality compared to the complete ODG-CLIP model.

Table 2. Ablation study on Fréchet distance between each of the source and target domains on PACS dataset using the visual features for domain alignment.

Methods	Cr→Ar	Ph→Ar	Sk→Ar	Ar→Cr	Ph→Cr	Sk→Cr	Ar→Ph	Cr→Ph	Sk→Ph	Ar→Sk	Cr→Sk	Ph→Sk
DAML [42]	256.41	278.35	224.13	235.89	240.14	197.34	301.56	296.31	283.27	200.37	178.92	235.28
CLIP [36]	231.43	217.75	230.32	224.51	234.17	207.21	267.56	275.32	258.48	160.31	180.46	218.35
CLIPN [46] + StyLIP [4]	200.67	195.70	180.35	198.21	204.21	180.25	247.89	263.19	240.38	149.39	160.86	198.37
ODG-CLIP w/o $\mathcal{L}_{sem}$	140.22	135.68	120.75	105.43	145.90	125.22	187.33	189.45	178.88	121.22	142.67	150.40
ODG-CLIP	**112.56**	**120.48**	**95.26**	**87.32**	**103.78**	**105.47**	**140.26**	**132.58**	**146.52**	**105.37**	**124.50**	**131.41**

6. Comparison of model complexity for different CLIP based techniques for ODG

In Fig. 1, we present a comparison of the model complexity of ODG-CLIP with its competitors. ODG-CLIP exhibits a level of complexity that is on par with most other models, yet it notably surpasses more complex alternatives like StyLIP + SD or CLIPN by a considerable margin. Importantly, when it comes to the H-Score, a key metric of performance, ODG-CLIP consistently outperforms all its counterparts, demonstrating its efficacy despite having comparable complexity.

7. Additional ablation studies

Position of the dom **token in the prompts**: In Table 3, we present an ablation study that varies the position of domain tokens in $\mathcal{P}_{dom,class}$ and $\mathcal{P}_{dom}$, demonstrated across four datasets.

Sensitivity of ODG-CLIP to the context lengths of the prompts: Table 4 illustrates how ODG-CLIP's performance is affected by the context length in both $\mathcal{P}_{dom,cls}$ and $\mathcal{P}_{dom}$. Generally, a context length of four yields the best outcomes, though a length of 16 also shows comparable results in most cases.

Cosine similarity measurements of latent features $\hat{x}$ **for pseudo-unknown class images**: Building on the findings presented in Fig. 3 (Top) of the main paper, where we explored the impact of $\mathcal{L}_{sem}$ on the cosine similarity of the $\hat{x}$ tensor for

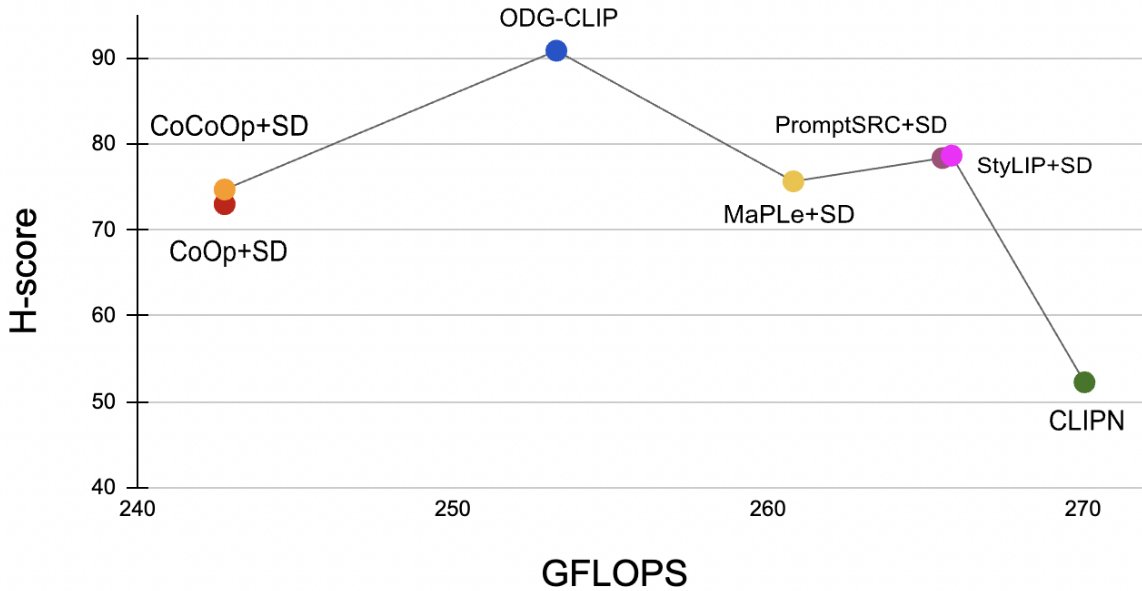

Figure 1. GFLOPs comparison of different methods.

Table 3. Ablation on the position of the domain tokens in the prompts.

position	PACS		O.H.		M.Data		M.DNet	
	Acc	H	Acc	H	Acc	H	Acc	H
front	**99.53**	**99.70**	**98.32**	**96.08**	84.60	90.00	**95.68**	**94.48**
middle	98.40	98.35	98.15	**96.08**	**84.63**	**90.08**	95.51	93.87
end	**99.53**	**99.70**	98.27	**96.08**	**84.63**	90.00	**95.68**	**94.48**

Table 4. Ablation on context lengths. $(\mathcal{M}, \mathcal{N})$ depicts the context length of $\mathcal{P}_{dom,cls}$ and $\mathcal{P}_{dom}$. We consider the case when `Art` serves as the target domain in Office-Home.

token length	(4,4)	(4,28)	(8,24)	(12,20)	(16,16)	(20,12)	(24,8)	(28,4)
H-score	**95.88**	93.78	94.80	94.80	**95.88**	92.83	92.81	91.81

closed classes, Table 5 extends this analysis by demonstrating the effects of $\mathcal{L}_{sem}$ on the $\hat{x}$ information for pseudo-unknown images.

Table 5. Cosine similarity in terms of $\hat{x}$ features with and without $\mathcal{L}_{sem}$ for the unknown-class samples averaged over all the domains.

Configuration	**PACS**	**VLCS**	**Office-Home**	**M-Dataset**	**M-DomainNet**
With $\mathcal{L}_{sem}$	0.81	0.82	0.76	0.78	0.79
Without $\mathcal{L}_{sem}$	0.31	0.30	0.32	0.37	0.35

8. Qualitative analysis

Effects of NP prompts for pseudo-open image generation: In Fig. 2, we note that using only the positive prompt, stable diffusion continues to produce images of known classes. For instance, in the PACS dataset, a positive prompt (`PP`) repeatedly generates images of 'Person' and 'Guitar', which are inlier classes.

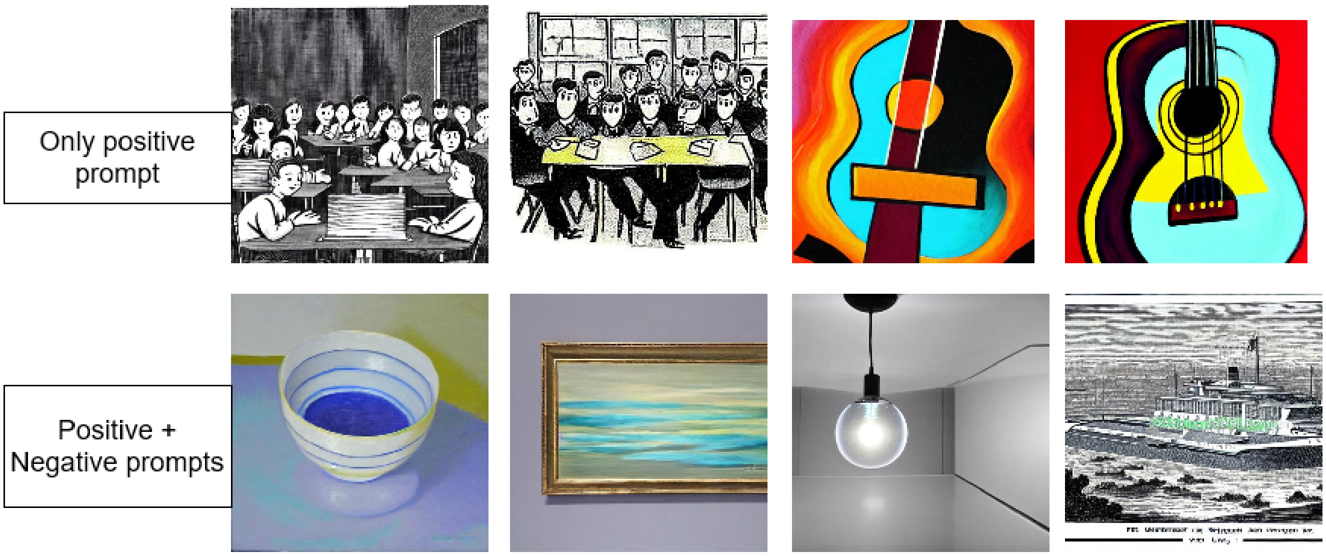

Figure 2. Images generated with only positive prompts vs. both the positive and negative prompts together by stable diffusion.

Analysis of the generated visual latent space: Figure 3 demonstrates that our method for generating $\tilde{x}$ provides greater discriminability compared to manually defining $\hat{x}$ from static class embeddings.

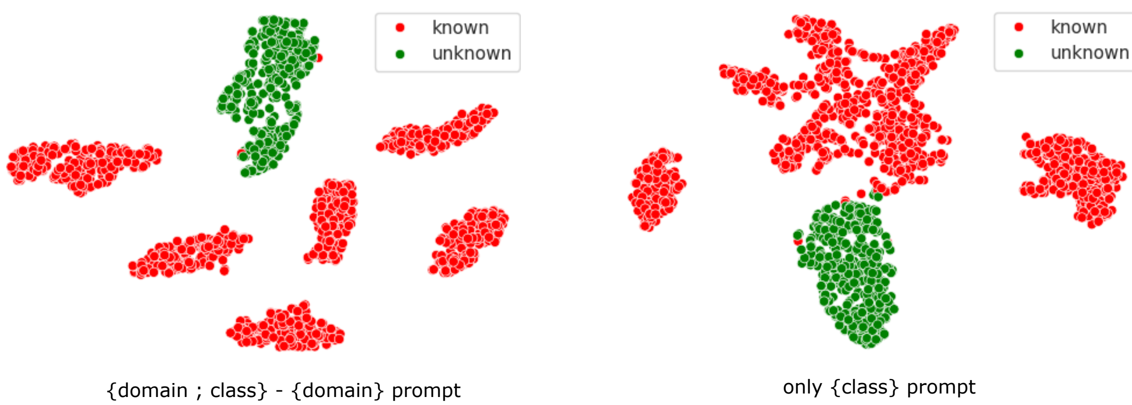

Figure 3. t-SNE of pseudo-open and closed image features produced using a manual $\hat{x}$ and by our proposed approach in ODG-CLIP.

t-SNE of open images produced by different methods: In Figure 4, we show the t-SNE plots of the CLIP features of the pseudo-open images produced by CuMix, OpenGAN, and the stable diffusion model, which clearly shows that the diffusion based model can better cover the open space.

9. Model ablation analysis

Table 6 presents the performance outcomes of ODG-CLIP using various CLIP visual encoders.

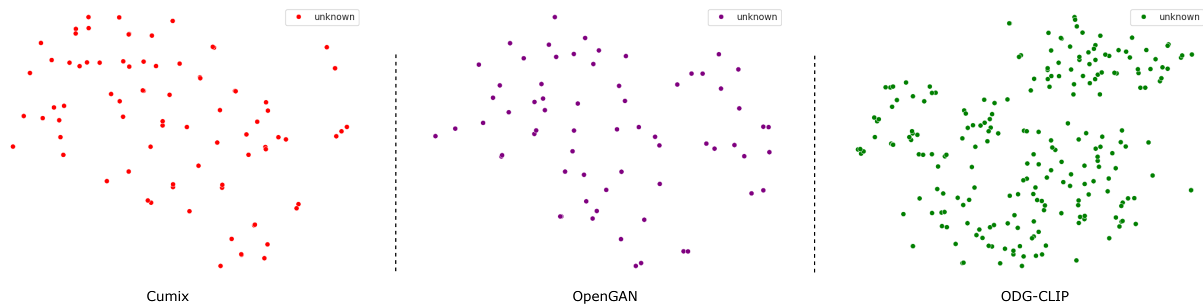

Figure 4. t-SNE of pseudo-open image features produced by Cumix, OpenGAN and ODG-CLIP.

Table 6. Ablation with ResNet-50 and ViT/B-16 based CLIP encoders.

Methods	PACS		VLCS		OfficeHome		Digits-DG		Multi-Dataset		Mini DomainNet		Average	
	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
RN-50	94.30	90.76	88.21	79.36	91.13	88.28	89.76	72.50	75.15	70.27	88.70	84.34	87.88	80.92
ViT-B/16	98.64	97.23	94.95	86.24	97.85	95.73	91.44	77.85	82.38	87.62	94.50	94.11	93.29	89.80
ViT-B/32	**99.53**	**99.70**	**95.71**	**86.53**	**98.32**	**96.08**	**91.53**	**78.27**	**84.60**	**90.00**	**95.68**	**94.48**	**94.23**	**90.84**

10. Additional results of using `unknown`-class prompts into existing models

In Table 7, we show further comparisons to the existing prompting techniques, equipped with the `unknown`-class prompts for the open samples, where the stable-diffusion model [38] was used to generate the training pseudo-open images for this prompt.

Table 7. Extended comparisons with respect to the prompting techniques coupled with the `unknown`-class prompts.

Methods	PACS		VLCS		OfficeHome		Digits-DG		Multi-Dataset		Mini DomainNet		Average	
	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
CoOp [51] + SD [38]	92.53	79.27	92.24	70.52	84.63	75.34	80.36	62.78	78.10	71.48	83.25	78.55	85.19	72.99
CoCoOp [50] + SD [38]	92.65	81.45	92.51	72.00	82.35	79.53	80.58	62.95	78.24	73.29	83.50	78.93	84.97	74.69
MaPLe [21] + SD [38]	91.47	82.60	91.70	72.67	85.02	80.60	79.92	65.82	77.62	72.83	83.79	79.30	84.92	75.64
LASP [6] + SD [38]	90.32	82.44	90.37	71.19	81.56	80.42	80.55	62.50	75.89	70.04	82.82	79.46	83.59	74.34
PromptSRC [22] + SD [38]	93.21	87.95	90.34	72.62	84.60	83.31	80.92	65.37	78.44	77.89	83.87	82.95	85.23	78.35
StyLIP [4] + SD [38]	91.78	87.42	92.11	73.34	85.51	81.22	81.45	68.10	79.05	78.52	84.12	83.21	85.67	78.64
ODG-CLIP	**99.53**	**99.70**	**95.71**	**86.53**	**98.32**	**96.08**	**91.53**	**78.27**	**84.60**	**90.00**	**95.68**	**94.48**	**94.23±0.19**	**90.84±0.26**

11. ODG results on full DomainNet

In Table 8, we show the ODG results on the full DomainNet dataset for all the domain combinations. The dataset splits are mentioned in Table 1.

12. Complete results on the all the datasets for ODG and closed-set DG

Please refer to Tables 9-14 for the detailed ODG results and Table 15-16 for the closed-set DG results, respectively.

References

[1] Devansh Arpit, Huan Wang, Yingbo Zhou, and Caiming Xiong. Ensemble of averages: Improving model selection and boosting performance in domain generalization. *Advances in Neural Information Processing Systems*, 35:8265–8277, 2022. 10

Table 8. Comparative analysis for DomainNet in ODG setting on average Acc and H-score over all the domain combinations following leave-one-domain-out protocol.

	Methods	Clipart		Painting		Real		Infograph		Quickdraw		Sketch		Average	
		Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
CNN-based	Cumix [30]	40.29	36.23	29.45	27.72	55.76	44.63	18.67	19.53	6.78	5.96	27.43	28.38	29.73	27.08
	MixStyle [53]	44.24	38.85	33.81	29.68	58.29	46.47	24.18	21.31	8.34	8.62	34.56	32.50	33.90	29.57
	DAML [42]	48.59	46.31	38.40	35.25	59.47	54.49	25.63	25.17	10.57	13.00	35.77	35.15	36.41	34.90
	MEDIC [47]	54.32	49.33	40.22	35.73	64.60	53.33	27.32	25.27	9.25	10.95	38.12	37.16	38.97	35.29
CLIP-based	CLIP [36]	67.98	52.20	61.76	54.00	82.92	61.54	43.50	43.47	13.87	14.46	55.58	49.05	54.27	45.79
	CLIP + OpenMax [2]	68.05	36.43	60.02	39.85	80.28	52.94	42.41	38.81	12.48	13.87	52.43	47.97	52.61	38.31
	CLIP + OSDA [33]	67.45	37.12	62.84	39.02	82.34	55.04	43.84	39.65	12.07	13.66	53.95	47.72	53.75	38.70
	CoOp [51]	68.77	31.42	58.94	26.17	72.58	34.11	45.26	29.89	14.71	10.55	56.81	30.93	52.85	27.18
	CoCoOp [50]	66.65	32.14	59.94	20.15	77.32	37.01	46.33	32.87	16.82	13.05	60.90	32.53	54.66	27.96
	MaPLe [21]	74.56	38.47	67.06	30.38	78.14	42.21	56.33	33.55	12.94	13.16	65.97	38.45	59.17	32.70
	LASP [6]	68.20	36.19	61.38	34.08	75.29	43.08	49.81	34.41	15.37	15.13	62.36	37.05	55.40	33.32
	PromptSRC [22]	76.43	42.55	66.25	32.33	79.17	43.98	58.29	36.56	15.78	13.93	66.45	40.83	60.40	35.03
	CLIPN [46]	75.51	53.40	62.64	41.21	82.49	56.08	55.28	45.37	17.54	15.89	64.58	48.30	59.67	43.37
	StyLIP [4]	79.14	48.23	64.80	46.39	86.52	53.07	56.12	42.74	18.65	16.85	68.14	45.48	62.23	42.13
	CLIPN + StyLIP	78.67	57.41	65.22	46.73	84.20	57.20	53.48	38.22	18.78	17.75	67.93	49.95	61.38	44.54
	MaPLe + SD	75.22	66.86	64.21	56.40	79.27	69.10	55.25	53.77	13.46	13.95	66.15	58.95	58.93	53.17
	PromptSRC + SD	75.39	68.92	62.48	60.34	79.93	70.76	57.82	57.01	16.38	15.89	69.37	61.50	60.23	55.74
	StyLIP + SD	79.25	71.60	65.04	59.14	85.19	74.45	56.73	55.16	17.32	17.18	68.93	62.60	62.08	56.69
	ODG-CLIP	**90.41**	**85.07**	**79.28**	**75.19**	**92.38**	**87.63**	**65.34**	**66.80**	**25.41**	**25.47**	**78.46**	**73.65**	**71.88**	**68.97**

Table 9. Comparative analysis for PACS in ODG setting on average Acc and H-score over all the domain combinations following leave-one-domain-out protocol.

	Methods	Art		Sketch		Photo		Cartoon		Average	
		Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
CNN-based	Cumix [30]	53.85	38.67	37.70	28.71	65.67	49.28	74.16	47.53	57.85	41.05
	MixStyle [53]	53.41	39.33	56.10	54.44	72.37	47.21	71.54	52.22	63.36	48.30
	DAML [42]	54.10	43.02	58.50	56.73	75.69	53.29	73.65	54.47	65.49	51.88
	MEDIC [47]	91.62	81.61	84.61	78.35	96.37	94.75	86.65	77.39	89.81	83.03
CLIP-based	CLIP [36]	96.87	73.50	85.38	70.90	99.75	92.83	98.65	69.85	95.16	76.77
	CLIP + OpenMax [2]	95.25	76.19	85.27	72.15	96.18	95.60	97.10	72.59	93.45	79.13
	CLIP + OSDA [33]	93.48	73.38	85.46	67.64	95.26	92.29	96.28	68.30	92.62	75.40
	CoOp [51]	96.23	29.60	83.05	21.91	89.04	34.78	46.77	21.20	78.77	26.87
	CoCoOp [50]	95.17	30.81	84.77	22.54	90.30	40.15	72.80	38.23	85.76	32.93
	MaPLe [21]	95.70	37.89	85.69	26.42	99.03	68.46	95.46	61.12	93.97	48.47
	LASP [6]	95.34	28.45	86.38	22.56	93.48	36.29	78.61	34.19	88.45	30.37
	PromptSRC [22]	96.05	30.14	87.23	23.49	98.6	62.36	96.24	57.27	94.53	43.32
	CLIPN [46]	97.27	32.50	91.71	20.80	98.15	66.17	97.83	60.52	96.24	45.00
	StyLIP [4]	96.93	40.74	92.34	28.51	96.38	70.43	95.79	63.26	95.36	50.74
	CLIPN + StyLIP	97.05	59.27	91.86	42.78	98.44	77.65	98.13	78.13	96.37	64.46
	MaPLe + SD	94.35	84.79	84.42	74.13	95.25	85.76	91.87	85.70	91.47	82.60
	PromptSRC + SD	94.84	88.51	89.30	83.59	94.28	90.35	94.43	89.36	93.21	87.95
	StyLIP + SD	95.27	87.48	87.25	81.38	91.65	90.93	92.95	89.90	91.78	87.42
	ODG-CLIP	**99.42**	**99.58**	**99.17**	**99.67**	**100.00**	**100.00**	**99.52**	**99.54**	**99.53**	**99.70**

[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. 7, 8, 9, 10

[3] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021. 2

[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. 3, 6, 7, 8, 9, 10

[5] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101–mining discriminative components with random forests. In *European conference on computer vision*, pages 446–461. Springer, 2014. 2

[6] Adrian Bulat and Georgios Tzimiropoulos. Lasp: Text-to-text optimization for language-aware soft prompting of vision & language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23232–23241, 2023. 2, 6, 7, 8, 9, 10

Table 10. Comparative analysis for VLCS in ODG setting on average Acc and H-score over all the domain combinations following leave-one-domain-out protocol.

	Methods	Caltech		LabelMe		Pascal VOC		Sun		Average	
		Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
CNN-based	Cumix [30]	66.21	63.76	46.72	45.59	50.54	45.78	46.38	45.32	52.46	50.11
	MixStyle [53]	66.11	63.19	46.72	46.22	49.75	46.19	46.62	46.85	52.30	50.61
	DAML [42]	69.18	64.65	48.22	47.71	49.87	47.22	46.87	46.78	53.54	51.59
	MEDIC [47]	76.47	69.90	52.47	55.27	52.91	50.43	47.25	47.32	57.28	55.73
CLIP-based	CLIP [36]	97.32	83.33	92.54	73.03	86.28	62.93	91.20	72.48	91.84	72.94
	CLIP + OpenMax [2]	97.92	85.25	93.67	76.51	85.98	62.34	90.78	70.57	92.09	73.67
	CLIP + OSDA [33]	96.53	80.36	90.23	72.43	82.45	60.55	91.64	70.23	90.21	70.89
	CoOp [51]	98.17	38.00	91.74	36.64	87.37	34.79	90.79	47.60	92.02	39.26
	CoCoOp [50]	96.86	30.70	87.11	37.78	87.52	34.30	86.40	45.27	89.47	37.01
	MaPLe [21]	93.72	45.92	90.53	43.18	86.07	48.50	88.46	35.71	89.70	43.33
	LASP [6]	95.37	39.54	88.62	39.47	89.40	47.12	89.28	31.51	90.67	39.41
	PromptSRC [22]	94.92	40.47	91.37	44.27	86.66	51.37	87.55	35.00	90.13	42.78
	CLIPN [46]	92.47	59.36	84.19	50.59	80.48	59.20	82.13	33.73	84.82	50.72
	StyLIP [4]	96.26	70.35	92.48	68.25	87.22	65.32	87.05	58.71	90.75	65.66
	CLIPN + StyLIP	92.31	73.68	85.50	71.46	80.42	68.79	80.35	58.15	84.65	68.02
	MaPLe + SD	96.45	79.26	93.24	75.24	88.82	70.25	88.30	65.94	91.70	72.67
	PromptSRC + SD	96.02	79.94	90.66	73.10	88.20	70.94	86.47	66.50	90.34	72.62
	StyLIP + SD	97.64	80.82	94.25	75.95	90.23	70.20	86.30	66.37	92.11	73.34
	ODG-CLIP	**98.35**	**90.75**	**96.35**	**89.45**	**94.65**	**88.05**	**93.48**	**77.85**	**95.71**	**86.53**

Table 11. Comparative analysis for Office-Home in ODG setting on average Acc and H-score over all the domain combinations following leave-one-domain-out protocol.

	Methods	Clipart		Real-World		Product		Art		Average	
		Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
CNN-based	Cumix [30]	41.54	43.07	64.63	58.02	57.74	55.79	42.76	40.72	51.67	49.40
	MixStyle [53]	42.28	41.15	61.78	60.23	59.92	53.97	50.11	42.78	53.52	49.53
	DAML [42]	45.13	43.12	65.99	60.13	61.54	59.00	53.13	51.11	56.45	53.34
	MEDIC [47]	48.96	49.39	67.42	61.00	65.20	66.09	59.46	55.17	60.26	57.91
CLIP-based	CLIP [36]	68.07	64.02	90.02	67.35	86.79	57.77	80.82	65.34	81.43	63.62
	CLIP + OpenMax [2]	68.44	63.41	89.10	62.30	85.25	55.32	81.20	65.12	81.00	61.54
	CLIP + OSDA [33]	69.76	67.93	91.67	70.65	84.60	61.53	84.29	69.28	82.58	67.35
	CoOp [51]	65.28	39.54	82.07	36.04	79.02	30.91	69.03	38.55	73.85	36.26
	CoCoOp [50]	68.21	33.05	81.62	39.41	80.92	30.19	70.77	34.86	75.38	34.38
	MaPLe [21]	79.48	36.57	85.44	31.42	77.11	28.23	75.83	36.00	79.47	33.06
	LASP [6]	72.36	32.75	80.50	37.78	76.37	31.38	75.27	36.15	76.13	34.52
	PromptSRC [22]	80.27	38.26	86.25	36.27	78.30	32.47	76.01	38.58	80.21	36.40
	CLIPN [46]	84.18	86.54	89.47	28.53	85.45	28.20	79.10	28.05	84.55	42.83
	StyLIP [4]	86.32	45.56	88.35	65.38	84.92	65.62	79.33	67.32	84.73	60.97
	CLIPN + StyLIP	85.97	84.71	85.69	70.31	84.10	72.02	78.92	78.94	83.67	76.50
	MaPLe + SD	87.23	81.32	89.34	80.79	84.15	81.00	79.34	79.30	85.02	80.60
	PromptSRC + SD	87.37	83.27	89.37	84.26	85.40	82.45	80.24	83.25	85.60	83.31
	StyLIP + SD	90.36	83.02	89.26	80.93	83.92	80.47	78.50	80.44	85.51	81.22
	ODG-CLIP	**97.84**	**96.33**	**98.74**	**95.36**	**99.50**	**96.74**	**97.18**	**95.88**	**98.32**	**96.08**

[7] Junbum Cha, Sanghyuk Chun, Kyungjae Lee, Han-Cheol Cho, Seunghyun Park, Yunsung Lee, and Sungrae Park. Swad: Domain generalization by seeking flat minima. *Advances in Neural Information Processing Systems*, 34:22405–22418, 2021. 10

[8] 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. 2

[9] DC Dowson and BV666017 Landau. The fréchet distance between multivariate normal distributions. *Journal of multivariate analysis*, 12(3):450–455, 1982. 1, 3

[10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010. 2

[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)*, December 2013. 2

[12] Li Fei-Fei, Rob Fergus, and Pietro 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. IEEE, 2004. 2

Table 12. Comparative analysis for Digits-DG in ODG setting on average Acc and H-score over all the domain combinations following leave-one-domain-out protocol.

	Methods	MNIST		MNIST-M		SVHN		SYN		Average	
		Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
CNN-based	Cumix [30]	72.10	67.52	45.88	43.74	52.22	47.22	62.33	58.33	58.13	54.20
	MixStyle [53]	76.56	70.56	47.81	45.66	54.97	47.24	61.80	61.96	60.29	56.36
	DAML [42]	73.98	69.88	46.49	45.62	53.34	47.72	64.22	59.23	59.51	55.61
	MEDIC [47]	97.89	83.20	71.14	60.98	76.00	58.77	88.11	62.24	83.29	66.30
CLIP-based	CLIP [36]	80.35	73.73	67.83	53.82	70.83	59.62	89.31	60.63	77.08	61.95
	CLIP + OpenMax [2]	79.28	76.32	63.49	51.18	74.30	60.83	90.65	62.78	76.93	62.78
	CLIP + OSDA [33]	81.54	79.51	71.50	54.21	78.91	64.11	90.17	64.95	80.53	65.70
	CoOp [51]	72.98	48.06	44.29	30.09	47.02	29.67	69.88	31.43	58.54	34.81
	CoCoOp [50]	45.24	41.01	50.60	28.96	49.29	31.42	65.95	32.62	52.77	33.50
	MaPLe [21]	77.74	55.19	58.21	37.35	61.67	43.52	84.52	39.25	70.54	43.83
	LASP [6]	61.43	42.65	51.32	29.30	51.33	38.70	79.48	30.27	60.89	35.23
	PromptSRC [22]	85.31	57.20	63.32	40.22	63.95	43.87	88.79	35.72	75.34	44.25
	CLIPN [46]	93.80	58.37	70.18	42.49	72.47	45.91	90.35	35.46	81.70	45.56
	StyLIP [4]	94.29	70.51	70.03	50.37	68.50	61.12	89.54	50.61	80.59	58.15
	CLIPN + StyLIP	93.87	71.43	69.74	51.28	74.52	60.84	90.43	53.42	82.14	59.24
	MaPLe + SD	91.44	77.19	67.92	59.59	73.33	66.28	86.97	60.21	79.92	65.82
	PromptSRC + SD	92.80	75.24	67.13	57.70	75.11	66.50	88.63	62.03	80.92	65.37
	StyLIP + SD	93.68	78.73	69.84	60.35	75.23	68.21	87.05	65.12	81.45	68.10
	ODG-CLIP	**96.24**	**87.14**	**86.23**	**72.10**	**87.41**	**79.34**	**96.24**	**74.51**	**91.53**	**78.27**

Table 13. Comparative analysis for Multi Dataset in ODG setting on average Acc and H-score over all the domain combinations following leave-one-domain-out protocol.

	Methods	Clipart		Real		Painting		Sketch		Average	
		Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
CNN-based	Cumix [30]	30.03	40.18	64.61	65.07	44.37	48.70	29.72	33.70	42.18	46.91
	MixStyle [53]	31.24	38.56	65.32	66.25	44.72	47.32	27.43	35.49	42.18	46.91
	DAML [42]	37.62	44.27	66.54	67.80	47.80	52.93	34.48	41.82	46.61	51.71
	MEDIC [47]	43.13	36.74	68.87	68.14	50.93	55.21	40.02	52.41	50.74	53.13
CLIP-based	CLIP [36]	81.00	74.13	84.02	72.31	69.53	68.77	76.98	73.55	77.88	72.19
	CLIP + OpenMax [2]	81.45	75.32	84.68	73.69	70.21	69.19	77.03	74.83	78.34	73.26
	CLIP + OSDA [33]	75.21	78.41	80.29	76.56	68.92	66.32	73.37	79.57	74.45	75.22
	CoOp [51]	66.00	51.65	63.11	38.72	69.90	45.97	65.10	41.03	66.03	44.34
	CoCoOp [50]	68.76	55.99	60.18	44.52	67.86	47.01	62.57	42.77	64.84	47.57
	MaPLe [21]	72.42	67.51	65.49	56.00	73.20	64.35	66.25	60.93	69.34	62.20
	LASP [6]	71.90	56.20	62.06	49.15	69.25	48.73	63.90	46.78	66.78	50.22
	PromptSRC [22]	73.15	64.29	61.75	55.39	64.72	60.45	62.41	57.67	65.51	59.45
	CLIPN [46]	80.39	68.43	71.99	58.40	80.61	65.10	75.63	58.46	77.16	62.60
	StyLIP [4]	83.25	80.30	74.32	70.52	82.89	76.83	79.07	60.32	79.88	71.99
	CLIPN + StyLIP	82.36	81.57	70.79	68.84	80.47	77.50	74.10	60.67	76.93	72.15
	MaPLe + SD	82.93	82.43	71.55	69.31	81.59	78.21	74.39	61.36	77.62	72.83
	PromptSRC + SD	83.54	87.31	72.40	76.09	84.13	84.58	73.68	63.59	78.44	77.89
	StyLIP + SD	84.30	87.78	73.75	76.56	85.92	86.38	72.21	63.34	79.05	78.52
	ODG-CLIP	**90.65**	**94.43**	**80.39**	**88.31**	**90.47**	**91.53**	**76.89**	**85.72**	**84.60**	**90.00**

[13] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015. 2

[14] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016. 3

[15] Peng Gao, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li, and Yu Qiao. Clip-adapter: Better vision-language models with feature adapters. *arXiv preprint arXiv:2110.04544*, 2021. 2

[16] Patrick Helber, Benjamin Bischke, Andreas Dengel, and Damian Borth. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7):2217–2226, 2019. 2

[17] Olivier Henaff. Data-efficient image recognition with contrastive predictive coding. In *International conference on machine learning*, pages 4182–4192. PMLR, 2020. 2

[18] Lanqing Hu, Meina Kan, Shiguang Shan, and Xilin Chen. Dandelionnet: Domain composition with instance adaptive classification for domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19050–19059,

Table 14. Comparative analysis for Mini-DomainNet in ODG setting on average Acc and H-score over all the domain combinations following leave-one-domain-out protocol.

	Methods	Clipart		Real		Painting		Sketch		Average	
		Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score	Acc	H-score
CNN-based	Cumix [30]	46.48	30.50	62.13	53.58	54.02	47.54	38.46	25.00	50.27	39.16
	MixStyle [53]	46.59	31.39	63.56	55.69	55.15	48.45	36.42	25.45	50.43	40.25
	DAML [42]	47.39	36.21	67.37	58.21	60.37	50.58	36.11	29.52	52.81	43.63
	MEDIC [47]	51.98	38.36	67.53	60.12	65.32	51.78	36.32	32.56	55.29	45.71
CLIP-based	CLIP [36]	88.00	69.35	90.50	68.84	80.00	66.72	79.50	70.85	84.50	68.94
	CLIP + OpenMax [2]	85.36	71.47	89.44	67.47	77.20	68.21	75.56	70.46	81.89	69.40
	CLIP + OSDA [33]	86.32	76.32	88.57	70.31	81.34	74.59	71.77	73.25	82.00	73.62
	CoOp [51]	64.50	75.53	75.00	77.68	57.50	70.70	47.50	49.50	61.13	68.35
	CoCoOp [50]	47.50	51.68	76.50	68.63	58.50	57.28	60.00	47.59	60.63	56.30
	MaPLe [21]	86.00	61.47	86.67	51.39	74.67	76.22	51.33	53.20	74.67	60.57
	LASP [6]	49.21	63.13	78.34	65.36	60.28	63.23	61.52	54.52	62.34	61.56
	PromptSRC [22]	87.33	63.28	87.17	65.06	67.60	67.56	52.30	54.35	73.60	62.56
	CLIPN [46]	88.64	66.21	88.35	70.32	73.24	71.02	59.28	60.14	77.38	66.92
	StyLIP [4]	89.18	68.93	89.84	74.27	76.69	71.58	65.15	61.66	80.22	69.11
	CLIPN + StyLIP	88.67	70.48	88.39	80.32	85.34	77.40	83.97	76.50	86.59	76.18
	MaPLe + SD	88.73	78.50	85.60	78.47	80.60	79.80	80.22	80.43	83.79	79.30
	PromptSRC + SD	89.03	80.29	86.04	84.96	80.11	82.35	80.30	84.21	83.87	82.95
	StyLIP + SD	89.67	83.13	86.39	85.12	80.20	83.04	80.23	81.53	84.12	83.21
	ODG-CLIP	**97.55**	**94.50**	**96.40**	**95.60**	**95.33**	**95.45**	**93.44**	**92.35**	**95.68**	**94.48**

Table 15. Comparative analysis for PACS, VLCS and Office-Home in closed-set setting over all the domain combinations following leave-one-domain-out protocol.

	Methods	PACS					VLCS					Office-Home				
		Art	Cartoon	Photo	Sketch	Avg	Caltech	LabelMe	Sun	P-VOC	Avg	Art	Clipart	Product	R-World	Avg
CNN	SWAD [7]	89.3	83.4	97.3	82.5	88.1	98.8	63.3	75.3	79.2	79.1	66.1	57.7	78.4	80.2	70.6
	EoA [1]	90.5	83.4	98.0	82.5	88.6	99.1	63.1	75.9	78.3	79.1	69.1	59.8	79.5	81.5	72.5
	DandelionNet [18]	87.8	86.5	96.8	85.8	89.2	99.1	70.2	77.2	80.0	81.6	65.8	58.6	78.0	79.7	70.5
CLIP-based	CLIP [36]	96.21	98.07	98.65	86.62	94.89	98.73	69.05	82.56	78.23	82.14	74.58	67.94	84.85	86.21	78.40
	CoOp [51]	97.85	98.64	99.70	92.23	97.11	98.58	70.20	84.28	80.31	83.34	77.32	72.10	88.43	87.46	81.33
	CoCoOp [50]	97.42	98.18	99.54	91.02	96.54	98.93	73.18	85.21	82.76	85.02	77.45	72.03	87.92	86.81	81.05
	MaPLe [21]	98.84	98.90	99.75	93.40	97.72	99.12	75.66	86.43	85.80	86.75	78.50	76.23	89.95	89.40	83.52
	LASP [6]	98.10	98.34	99.27	92.35	97.02	99.45	76.54	86.98	86.02	87.25	79.24	76.75	90.14	90.37	84.13
	PromptSRC [22]	98.79	99.02	99.50	94.76	98.02	99.61	75.30	85.39	85.07	86.34	78.97	75.82	90.31	90.44	83.89
	StyLIP [4]	98.73	99.15	99.97	94.82	98.17	99.70	75.84	87.08	86.22	87.21	81.54	78.78	91.67	91.75	85.94
	ODG-CLIP	**99.93**	**99.87**	**100.00**	**99.51**	**99.83**	**100.00**	**92.63**	**95.71**	**94.60**	**95.74**	**96.38**	**92.35**	**99.52**	**99.37**	**96.91**

Table 16. Comparative analysis for Digits-DG and Mini-DomainNet in closed-set setting over all the domain combinations following leave-one-domain-out protocol.

Methods	Digits-DG					Mini-DomainNet				
	MNIST	MNIST-M	SVHN	SYN	Average	Clipart	Real	Painting	Sketch	Average
CLIP [36]	83.48	58.41	46.64	69.82	64.59	85.25	66.84	95.13	67.71	78.73
CoOp [51]	93.11	71.32	61.28	82.73	77.11	82.49	61.34	92.94	64.42	75.30
CoCoOp [50]	93.56	74.90	64.51	84.45	79.36	77.38	59.75	88.57	60.34	71.51
MaPLe [21]	94.25	75.68	66.72	84.33	80.25	81.27	62.58	88.29	63.32	73.87
LASP [6]	95.87	75.61	65.91	82.28	79.92	80.51	58.30	85.14	58.72	70.67
PromptSRC [22]	96.24	78.94	68.04	86.36	82.40	87.63	62.45	89.52	64.80	76.10
StyLIP [4]	96.39	78.53	66.35	85.20	81.62	89.36	67.63	94.57	70.14	80.43
ODG-CLIP	**99.48**	**96.38**	**91.22**	**98.42**	**96.38**	**98.54**	**92.37**	**99.42**	**96.25**	**96.65**

2023. 10

[19] Dat Huynh and Ehsan Elhamifar. Fine-grained generalized zero-shot learning via dense attribute-based attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4483–4493, 2020. 2

[20] Menglin Jia, Luming Tang, Bor-Chun Chen, Claire Cardie, Serge Belongie, Bharath Hariharan, and Ser-Nam Lim. Visual prompt tuning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*, pages 709–727. Springer, 2022. 2

[21] Muhammad Uzair Khattak, Hanoona Rasheed, Muhammad Maaz, Salman Khan, and Fahad Shahbaz Khan. Maple: Multi-modal prompt learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19113–19122, June 2023. 2, 6, 7, 8, 9, 10

[22] 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. 6, 7, 8, 9, 10

[23] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 813–822, 2021. 1

[24] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 2

[25] J Devlin M Chang K Lee and K Toutanova. Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 2

[26] Aodi Li, Liansheng Zhuang, Shuo Fan, and Shafei Wang. Learning common and specific visual prompts for domain generalization. In *Proceedings of the Asian Conference on Computer Vision*, pages 4260–4275, 2022. 2

[27] 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. 2

[28] Yuning Lu, Jianzhuang Liu, Yonggang Zhang, Yajing Liu, and Xinmei Tian. Prompt distribution learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5206–5215, 2022. 2

[29] Mohammad Mahdi Derakhshani, Enrique Sanchez, Adrian Bulat, Victor Guilherme Turrisi da Costa, Cees GM Snoek, Georgios Tzimiropoulos, and Brais Martinez. Variational prompt tuning improves generalization of vision-language models. *arXiv e-prints*, pages arXiv–2210, 2022. 2

[30] 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. 1, 7, 8, 9, 10

[31] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011. 2

[32] Hongjing Niu, Hanting Li, Feng Zhao, and Bin Li. Domain-unified prompt representations for source-free domain generalization. *arXiv preprint arXiv:2209.14926*, 2022. 2

[33] Pau Panareda Busto and Juergen Gall. Open set domain adaptation. In *Proceedings of the IEEE international conference on computer vision*, pages 754–763, 2017. 7, 8, 9, 10

[34] 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

[35] 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. 2

[36] 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. 2, 3, 7, 8, 9, 10

[37] 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

[38] 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. 1, 6

[39] 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(1):157–173, 2008. 2

[40] 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. 2

[41] Manli Shu, Weili Nie, De-An Huang, Zhiding Yu, Tom Goldstein, Anima Anandkumar, and Chaowei Xiao. Test-time prompt tuning for zero-shot generalization in vision-language models. *arXiv preprint arXiv:2209.07511*, 2022. 2

[42] Yang Shu, Zhangjie Cao, Chenyu Wang, Jianmin Wang, and Mingsheng Long. Open domain generalization with domain-augmented meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9624–9633, 2021. 2, 3, 7, 8, 9, 10

[43] Mainak Singha, Ankit Jha, Bhupendra Solanki, Shirsha Bose, and Biplab Banerjee. Applenet: Visual attention parameterized prompt learning for few-shot remote sensing image generalization using clip. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2023–2033, June 2023. 2

[44] Mainak Singha, Harsh Pal, Ankit Jha, and Biplab Banerjee. Ad-clip: Adapting domains in prompt space using clip. *arXiv preprint arXiv:2308.05659*, 2023. 2

[45] 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. 2

[46] 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. 3, 7, 8, 9, 10

[47] 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

[48] 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. IEEE, 2010. 2

[49] Xin Zhang, Yusuke Iwasawa, Yutaka Matsuo, and Shixiang Shane Gu. Amortized prompt: Lightweight fine-tuning for clip in domain generalization. *arXiv preprint arXiv:2111.12853*, 2021. 2

[50] 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. 2, 6, 7, 8, 9, 10

[51] 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. 2, 6, 7, 8, 9, 10

[52] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *European conference on computer vision*, pages 561–578. Springer, 2020. 2

[53] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. *arXiv preprint arXiv:2104.02008*, 2021. 7, 8, 9, 10

[54] Beier Zhu, Yulei Niu, Yucheng Han, Yue Wu, and Hanwang Zhang. Prompt-aligned gradient for prompt tuning. *arXiv preprint arXiv:2205.14865*, 2022. 2