Appendix for: Beyond the Static World: Continual Category Discovery under Visual Drift

Wei Feng[1,2] Yiwen Jiang[1] Sijin Zhou[1,2] Zongyuan Ge[1,2]

[1]AIM for Health Lab, Monash University [2]Airdoc–Monash Research, Monash University

{wf02429, sjzhou1995}@gmail.com, {yiwen.jiang, zongyuan.ge}@monash.edu

1. Theoretical Analysis

Dual Form of PUOT

We consider the Partial Unbalanced Optimal Transport (PUOT) objective J_{puot} as introduced in Eq.(2), which can be equivalently rewritten in its dual representation as follows:

$$\min_{f,g,\zeta}\left[\tau\sum_{i=1}^{N}a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right)-\sum_{j=1}^{H}b_j(g_j-\zeta)\right] \quad \text{s.t.} \quad f_i+g_j+s_{ij}=U_{ij}, \quad s_{ij}\geq 0 \tag{1}$$

Here, f, g, s, and ζ denote the dual variables (Lagrange multipliers), with the constraint $s_{ij} \geq 0$ ensured by the KKT conditions. Based on this formulation, PUOT can be equivalently reformulated as a reweighted OT problem:

$$\min_{\pi\geq 0}\langle U,\pi\rangle \quad \text{s.t.}\sum_{j=1}^{H}\pi_{ij}=a_i\exp\left(-\frac{f_i^*+\zeta^*}{\tau}\right), \quad \sum_{i=1}^{N}\pi_{ij}=b_j \tag{2}$$

Proof. We begin by introducing a marginal constraint $\pi\mathbf{1}_{C_{\text{kno}}}=\alpha$, where α does not need to be known beforehand. This assumption simplifies the derivation. The PUOT objective can thus be rewritten as:

$$\min_{\pi_{ij}\geq 0}\mathcal{J}_{\text{puot}}=\langle U,\pi\rangle+\tau\cdot\text{KL}(\pi\mathbf{1}_{C_{\text{kno}}}\,\|\,a) \quad \text{s.t.} \quad \pi^\top\mathbf{1}_N=b, \quad \pi\mathbf{1}_{C_{\text{kno}}}=\alpha \tag{3}$$

Its Lagrangian relaxation yields:

$$\max_{s\geq 0,f,g,\zeta}\min_{\pi\geq 0}J=\tau\cdot\text{KL}(\pi\mathbf{1}_{C_{\text{kno}}}\,\|\,a)+\langle f+\zeta,\pi\mathbf{1}_{C_{\text{kno}}}\rangle+\langle g-\zeta,b\rangle+C_{\text{SUOT}} \tag{4}$$

Here, ζ acts as a translation-invariant component. The term C_{SUOT} can be expanded as:

$$\begin{aligned}C_{\text{SUOT}}&=\sum_{i,j}(U_{ij}-f_i-g_j-s_{ij})\pi_{ij}\\&=\langle U-f\otimes\mathbf{1}_N^\top-\mathbf{1}_{C_{\text{kno}}}\otimes g^\top-s,\pi\rangle\end{aligned} \tag{5}$$

Here, we use the Karush–Kuhn–Tucker (KKT) conditions which imply $s_{ij} \geq 0$ and $s_{ij}\cdot\pi_{ij}=0$. Recall the definition of Kullback-Leibler divergence between two distributions x and z:

$$\text{KL}(x\,\|\,z)=\sum_{d=1}^{D}\left[x_d\log\frac{x_d}{z_d}-x_d+z_d\right] \tag{6}$$

Its derivative w.r.t. π_{ij} is:

$$\frac{\partial}{\partial\pi_{ij}}\text{KL}(\pi\mathbf{1}_{C_{\text{kno}}}\,\|\,a)=\log\frac{\sum_{j=1}^{C_{\text{kno}}}\pi_{ij}}{a_i} \tag{7}$$

Thus:

$$\frac{\partial J}{\partial\pi_{ij}}=U_{ij}+\tau\log\frac{\sum_{j=1}^{C_{\text{kno}}}\pi_{ij}}{a_i}+\zeta-g_j-s_{ij}=0 \tag{8}$$

We obtain the following conclusions:

$$\sum_{j=1}^{C_{\text{kno}}}\pi_{ij}=a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right), \quad \sum_{i=1}^{N}\pi_{ij}=b_j \tag{9}$$

We can verify that the resulting constraints satisfy the condition $U_{ij}-f_i-g_j-s_{ij}=0$. By incorporating the marginal distributions into the KL divergence term, the objective function can be expressed as:

$$\begin{aligned}\mathcal{L}&=\tau\cdot\text{KL}(\pi\mathbf{1}_{C_{\text{kno}}}\,\|\,a)+\langle f+\zeta,\pi\mathbf{1}_H\rangle\\&=\tau\cdot\text{KL}\left(\left\langle a,\exp\left(-\frac{f+\zeta}{\tau}\right)\right\rangle\middle\|\,a\right)\\&+\left\langle f+\zeta,\,a\cdot\exp\left(-\frac{f+\zeta}{\tau}\right)\right\rangle\\&=\sum_{i=1}^{N}\left(-\tau a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right)+\tau a_i\right)\end{aligned} \tag{10}$$

Discarding constants, the final dual becomes:

$$\max_{f,g,\zeta}\left[-\tau\sum_{i=1}^{N}a_i\exp(-\frac{f_i+\zeta}{\tau})+\sum_{j=1}^{C_{\text{kno}}}b_j(g_j-\zeta)\right] \quad \text{s.t.} \quad f_i+g_j+s_{ij}=U_{ij},\ s_{ij}\geq 0 \tag{11}$$

Thus, the proof is completed.

Optimization Strategy for PUOT

To optimize the PUOT objective, we focus on the variables f and ζ. The corresponding objective function is expressed as:

$$\begin{aligned}\min_{f,\zeta} L_F = \tau\sum_{i=1}^{N}a_i\exp(-\frac{f_i+\zeta}{\tau}) \\ -\sum_{j=1}^{C_{\text{kno}}}\left[\inf_k(U_{kj}-f_k)-\zeta\right]b_j\end{aligned} \tag{12}$$

To improve the efficiency of the optimization process, we introduce a smooth surrogate for the infimum operation. Specifically, the term $\inf_{k\in[N]}[U_{kj}-f_k]$ is approximated using a soft minimum:

$$\inf_{k\in[N]}[U_{kj}-f_k]\approx-\varsigma\log\left(\sum_{k=1}^{N}\exp\left(\frac{f_k-U_{kj}}{\varsigma}\right)\right), \tag{13}$$

where $\varsigma>0$ is a smoothing hyperparameter controlling the trade-off between numerical stability and approximation fidelity. In our implementation, we set $\varsigma=0.01$ to balance precision and smoothness. With this relaxation, the objective becomes:

$$\begin{aligned}\mathcal{A}_{\text{PUOT}} = \tau\exp\left(\frac{-\zeta}{\tau}\right)\sum_{i=1}^{N}a_i\exp\left(\frac{-f_i}{\tau}\right) \\ +\sum_{j=1}^{C_{\text{kno}}}\left[\varsigma\log\left(\sum_{k=1}^{N}\exp\left(\frac{f_k-U_{kj}}{\varsigma}\right)\right)+\zeta\right]b_j\end{aligned} \tag{14}$$

To facilitate the optimization procedure, we introduce a Wasserstein Fixed-Point Iteration (WFPI) method. The core idea is to derive the optimality condition by setting the partial derivative of the PUOT objective $\mathcal{A}_{\text{PUOT}}$ with respect to f_i to zero:

$$\begin{aligned}\exp\left(\frac{f_i}{\varsigma}\right)\sum_{j=1}^{C_{\text{kno}}}\left[\frac{b_j\exp\left(-\frac{U_{ij}}{\varsigma}\right)}{\sum_{k=1}^{N}\exp\left(\frac{f_k-U_{kj}}{\varsigma}\right)}\right] \\ =a_i\exp\left(\frac{-f_i+\zeta}{\tau}\right)\end{aligned} \tag{15}$$

The iterative update for f_s at iteration $\ell+1$ is then given by:

$$\begin{aligned}f_s^{(\ell+1)} &= \kappa\varsigma\left[\log(a_s)-\frac{\zeta}{\tau}-\log\left(\sum_{j=1}^{C_{\text{kno}}}\frac{b_j\exp\left(-\frac{U_{sj}}{\varsigma}\right)}{M_j}\right)\right] \\ &= \mathcal{U}_s(f_1^{(\ell)},\ldots,f_N^{(\ell)})\end{aligned} \tag{16}$$

where $\kappa=\frac{\tau}{\tau+\varsigma}$ and $M_j=\sum_{k=1}^{N}\exp\left(\frac{f_k^{(\ell)}-U_{kj}}{\varsigma}\right)$.

To guarantee the convergence of WFPI, we analyze the Jacobian matrix. For example, the partial derivative w.r.t. $f_1^{(\ell)}$ is:

$$\begin{aligned}\frac{\partial\mathcal{U}_s}{\partial f_1^{(\ell)}} &= -\frac{\kappa\varsigma}{\sum_{j=1}^{C_{\text{kno}}}\frac{b_j\exp\left(-\frac{U_{sj}}{\varsigma}\right)}{M_j}}\cdot\frac{\partial}{\partial f_1^{(\ell)}}\left(\sum_{j=1}^{C_{\text{kno}}}\frac{b_j\exp\left(-\frac{U_{sj}}{\varsigma}\right)}{M_j}\right) \\ &= \frac{\kappa}{\sum_{j=1}^{C_{\text{kno}}}\frac{b_j\exp\left(-\frac{U_{sj}}{\varsigma}\right)}{M_j}}\cdot\sum_{j=1}^{C_{\text{kno}}}\left[\frac{b_j\exp\left(-\frac{U_{sj}}{\varsigma}\right)}{M_j}\cdot\frac{\exp\left(\frac{f_1^{(\ell)}-U_{1j}}{\varsigma}\right)}{M_j}\right] \\ &< 1\end{aligned} \tag{17}$$

By further expanding the inner terms, the derivative is upper-bounded by:

$$\max_i\left\|\frac{\partial\mathcal{U}_s}{\partial f_i^{(\ell)}}\right\|_\infty=\frac{\tau}{\tau+\varsigma}<1 \tag{18}$$

indicating that the Jacobian's infinity norm is strictly less than 1, hence WFPI is contractive and converges.

Finally, we solve for ζ by enforcing $\nabla_\zeta\mathcal{L}_F=0$. This yields a closed-form solution:

$$\zeta=\tau\left[\log\left(\sum_{i=1}^{N}a_i\exp\left(-\frac{f_i}{\tau}\right)\right)-\log\left(\sum_{j=1}^{C_{\text{kno}}}b_j\right)\right] \tag{19}$$

Clearly, this ensures that the constraint $\sum_{j=1}^{N}b_j=\sum_{i=1}^{N}a_i\exp\left(-\frac{f_i+\zeta}{\tau}\right)$ is satisfied during each iteration. In this formulation, the term $\exp(-\frac{f_i+\zeta}{\tau})$ serves as a reweighting factor for unlabeled samples, which can be subsequently used to construct the Response Spectrum for discovering unknown categories.

2. Datasets

To comprehensively evaluate the proposed method under both domain shift and semantic shift conditions, we conduct experiments on two widely-used benchmarks: **DomainNet** [4] and **SSB-C** [5]. These datasets cover diverse visual domains and fine-grained recognition tasks, enabling a robust assessment of generalization capabilities.

Table 1. Number of classes at each incremental stage on the Corrupted SSB and DomainNet benchmarks. We report the cumulative class count for each dataset (**Original** vs. **Corrupted**) across four stages.

Stage	**CUB-C**		**Stanford Cars-C**		**FGVC-Aircraft-C**		**DomainNet**	
	Original	Corrupted	Original	Corrupted	Original	Corrupted	Real	Other Domains
0	140	N/A	130	N/A	70	N/A	225	N/A
1	160	160	152	152	80	80	265	265
2	180	180	174	174	90	90	305	305
3	200	200	196	196	100	100	345	345

Table 2. Class splits across labeled set ($\mathcal{D}_l$) and unlabeled streams ($\mathcal{D}_u^1$, $\mathcal{D}_u^2$, $\mathcal{D}_u^3$) for both known domain and unknown domain.

Class Range	Known Domain				Unknown Domain			
	$\mathcal{D}_l$	$\mathcal{D}_u^1$	$\mathcal{D}_u^2$	$\mathcal{D}_u^3$	$\mathcal{D}_l$	$\mathcal{D}_u^1$	$\mathcal{D}_u^2$	$\mathcal{D}_u^3$
$\{y_i \mid y_i \leq 0.7 \cdot \lvert\mathcal{Y}\rvert\}$	87%	7%	3%	3%	0%	7%	3%	3%
$\{y_i \mid 0.7 \cdot \lvert\mathcal{Y}\rvert < y_i \leq 0.8 \cdot \lvert\mathcal{Y}\rvert\}$	0%	70%	20%	10%	0%	70%	20%	10%
$\{y_i \mid 0.8 \cdot \lvert\mathcal{Y}\rvert < y_i \leq 0.9 \cdot \lvert\mathcal{Y}\rvert\}$	0%	0%	90%	10%	0%	0%	90%	10%
$\{y_i \mid 0.9 \cdot \lvert\mathcal{Y}\rvert < y_i \leq \lvert\mathcal{Y}\rvert\}$	0%	0%	0%	100%	0%	0%	0%	100%

Table 3. Clustering performance (**mean ± std**) on DomainNet benchmark. Real is used as known domain $\mathcal{T}_A$, and each of the remaining domains is used as unknown domain $\mathcal{T}_B$. We report average All / Old / New accuracy across all stages for both domains.

Method	Real → Painting						Real → Sketch						Real → Quickdraw						Real → Clipart						Real → Infograph					
	Real			Painting			Real			Sketch			Real			Quickdraw			Real			Clipart			Real			Infograph		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	51.3 ± 2.2	67.2 ± 2.1	45.4 ± 1.7	27.4 ± 1.4	26.7 ± 1.9	28.1 ± 1.0	52.3 ± 1.0	65.7 ± 1.1	41.7 ± 1.8	9.2 ± 1.8	14.5 ± 1.1	10.1 ± 1.6	38.7 ± 2.0	56.2 ± 2.0	29.6 ± 1.0	5.0 ± 1.5	4.7 ± 1.4	5.8 ± 2.1	46.7 ± 2.0	65.7 ± 0.8	40.1 ± 1.0	14.5 ± 1.9	21.2 ± 1.6	10.1 ± 1.4	39.8 ± 2.1	55.3 ± 0.9	32.4 ± 1.3	8.1 ± 2.3	9.8 ± 2.0	6.4 ± 1.5
SimGCD	48.4 ± 1.7	63.9 ± 1.3	41.3 ± 1.1	22.6 ± 0.9	22.4 ± 1.3	23.5 ± 1.3	48.5 ± 1.9	60.2 ± 1.2	36.5 ± 0.9	7.2 ± 0.9	11.3 ± 1.5	9.2 ± 1.2	32.4 ± 1.6	50.3 ± 2.1	23.5 ± 1.4	4.2 ± 2.1	4.0 ± 1.3	5.1 ± 1.7	40.2 ± 1.9	58.8 ± 0.9	33.5 ± 1.7	10.3 ± 1.9	18.8 ± 1.4	8.2 ± 2.0	33.6 ± 1.5	49.2 ± 1.4	27.8 ± 1.3	6.7 ± 1.0	7.8 ± 2.2	5.2 ± 1.4
SPTNet	49.8 ± 1.9	64.5 ± 2.2	42.5 ± 1.2	24.1 ± 1.1	23.5 ± 1.0	24.3 ± 1.5	49.9 ± 1.3	62.3 ± 1.1	37.8 ± 1.3	7.9 ± 2.2	11.7 ± 1.8	9.6 ± 1.7	34.8 ± 0.9	52.6 ± 1.3	24.8 ± 1.3	4.9 ± 1.1	4.6 ± 1.5	5.5 ± 0.9	43.1 ± 1.7	60.3 ± 1.0	35.9 ± 1.3	11.6 ± 2.1	19.3 ± 1.2	8.9 ± 1.9	35.9 ± 1.2	51.4 ± 1.1	29.8 ± 1.0	7.2 ± 1.9	8.0 ± 1.5	5.9 ± 1.1
HiLo	56.1 ± 1.2	70.9 ± 1.6	49.2 ± 1.3	31.0 ± 1.0	31.8 ± 1.0	30.8 ± 0.8	55.7 ± 1.5	71.3 ± 1.8	47.2 ± 1.2	13.1 ± 1.6	17.2 ± 1.2	12.4 ± 1.7	43.9 ± 1.2	61.2 ± 1.5	35.4 ± 1.0	5.9 ± 0.9	5.5 ± 1.2	6.7 ± 0.9	55.4 ± 1.3	70.8 ± 1.4	48.1 ± 1.2	18.2 ± 1.2	25.1 ± 1.3	13.0 ± 2.2	49.4 ± 2.0	65.2 ± 2.3	42.1 ± 0.9	9.8 ± 0.9	12.5 ± 1.0	8.1 ± 1.4
G&M	47.1 ± 1.3	62.3 ± 2.2	41.2 ± 1.8	26.3 ± 0.9	25.5 ± 1.0	26.2 ± 1.2	48.9 ± 2.1	63.4 ± 0.8	42.3 ± 2.0	10.9 ± 1.7	15.1 ± 1.1	10.5 ± 2.2	34.1 ± 2.0	50.2 ± 1.5	27.3 ± 1.1	3.5 ± 2.0	4.1 ± 1.0	5.2 ± 1.4	40.3 ± 2.3	61.1 ± 1.3	34.2 ± 1.5	12.3 ± 1.8	19.2 ± 1.1	8.8 ± 1.7	32.4 ± 2.2	50.1 ± 2.1	27.6 ± 1.3	7.5 ± 1.5	9.2 ± 2.1	5.5 ± 1.8
PA-CGCD	55.4 ± 1.6	70.3 ± 1.1	48.1 ± 1.5	30.1 ± 1.4	30.8 ± 1.3	30.2 ± 1.3	55.1 ± 1.9	70.7 ± 1.6	46.6 ± 1.2	12.3 ± 1.2	16.1 ± 1.5	11.2 ± 1.3	43.6 ± 1.7	60.4 ± 1.9	34.2 ± 1.5	5.1 ± 1.4	5.0 ± 1.1	6.0 ± 1.6	52.2 ± 1.8	70.3 ± 1.3	44.6 ± 1.5	17.8 ± 1.8	24.5 ± 1.9	12.3 ± 1.3	45.2 ± 1.4	61.3 ± 1.2	38.1 ± 1.5	9.0 ± 1.3	11.8 ± 1.8	7.1 ± 1.4
DEAN	56.0 ± 1.2	71.7 ± 1.9	47.9 ± 2.0	32.8 ± 1.1	34.4 ± 1.4	31.5 ± 2.1	56.7 ± 2.0	71.5 ± 2.2	47.6 ± 1.5	12.9 ± 1.5	16.8 ± 0.9	11.2 ± 1.1	43.6 ± 2.1	60.4 ± 1.7	34.2 ± 2.1	5.1 ± 1.2	5.0 ± 1.6	6.1 ± 2.2	55.1 ± 2.2	72.7 ± 1.2	47.5 ± 2.2	20.3 ± 1.9	26.7 ± 1.5	15.0 ± 1.2	46.7 ± 1.7	62.3 ± 1.5	40.8 ± 1.4	9.5 ± 1.4	12.5 ± 2.0	7.9 ± 1.9
PromptCCD	56.5 ± 1.1	71.2 ± 1.7	50.3 ± 1.8	31.5 ± 1.5	32.1 ± 1.9	31.2 ± 1.4	57.4 ± 1.3	73.6 ± 1.5	48.6 ± 1.2	13.4 ± 1.1	17.7 ± 1.3	12.1 ± 1.0	45.2 ± 1.3	62.3 ± 1.6	36.7 ± 1.0	5.8 ± 1.2	5.1 ± 1.1	6.5 ± 1.3	54.1 ± 1.3	71.2 ± 1.1	46.7 ± 1.4	19.8 ± 1.5	26.1 ± 1.6	14.4 ± 1.4	47.1 ± 1.5	63.1 ± 1.3	40.2 ± 1.2	9.2 ± 1.3	12.2 ± 1.6	7.8 ± 1.1
Ours	60.2 ± 1.4	73.3 ± 0.8	54.2 ± 1.3	38.9 ± 1.0	38.5 ± 1.4	37.7 ± 1.3	59.3 ± 1.2	73.4 ± 1.4	50.2 ± 1.2	16.3 ± 1.1	19.3 ± 1.2	15.2 ± 1.2	53.5 ± 1.5	73.1 ± 0.7	48.9 ± 1.2	6.9 ± 1.3	6.2 ± 1.0	7.1 ± 0.8	57.4 ± 1.3	72.5 ± 1.3	50.7 ± 1.5	23.5 ± 0.7	29.8 ± 1.4	18.6 ± 1.5	59.3 ± 1.1	73.5 ± 1.0	52.1 ± 0.8	10.7 ± 1.2	13.3 ± 0.9	9.2 ± 1.0

2.1. DomainNet

DomainNet is one of the largest and most diverse datasets in the domain adaptation and generalization literature. It contains approximately 600,000 images spanning 345 object categories and is distributed across six visually distinct domains, each characterized by unique styles or modalities:

- **Real**: photographs of real-world objects,
- **Clipart**: cartoon-style clipart images,
- **Sketch**: black-and-white hand-drawn sketches,
- **Painting**: artistic renditions including oil and watercolor paintings,
- **Infograph**: infographic-like images with symbolic content,
- **Quickdraw**: doodle-style line drawings from the Google QuickDraw dataset.

The diversity across these domains poses significant challenges for cross-domain recognition and makes DomainNet an ideal testbed for evaluating domain-invariant feature learning and category separation.

2.2. SSB-C

The SSB-C benchmark [5] is a corrupted extension of the Semantic Shift Benchmark (SSB), specifically designed to evaluate robustness under semantic and distributional shifts. The original SSB is constructed from three fine-grained visual categorization datasets:

- **CUB-200-2011 (CUB)**: 200 bird species with fine-grained visual distinctions,
- **Stanford Cars (SCAR)**: 196 car categories covering different brands and models,
- **FGVC-Aircraft (FGVC)**: 100 airplane categories with subtle structural variations.

SSB-C introduces controlled corruption to the SSB images by applying nine types of perturbations (e.g., Gaussian noise, Frosted blur, Impulse noise, etc.) at five severity levels, following the protocol of common corruption benchmarks. This results in a comprehensive evaluation set that is **45× larger** than the original SSB, significantly increasing its utility for studying robustness in fine-grained settings.

For each dataset, we begin by selecting a subset of categories as labeled known classes to construct the initial ses-

Table 4. Clustering performance (**mean** $\pm$ **std**) on SSB-C benchmarks. Each dataset includes both *Original* and *Corrupted* domains. We report the average accuracy for *All*, *Old*, and *New* classes across all stages in each domain.

Method	CUB-C						Stanford Cars-C						FGVC-Aircraft-C					
	Original			Corrupted			Original			Corrupted			Original			Corrupted		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	29.4 ± 1.4	47.7 ± 1.5	23.4 ± 1.5	26.8 ± 1.3	45.9 ± 1.5	20.1 ± 2.2	26.4 ± 1.0	56.1 ± 1.8	21.5 ± 1.7	22.3 ± 1.7	43.1 ± 1.0	11.2 ± 1.6	27.7 ± 1.0	33.6 ± 1.2	24.9 ± 2.1	28.8 ± 2.2	41.4 ± 1.4	28.8 ± 1.8
SimGCD	26.6 ± 1.5	44.5 ± 2.0	21.0 ± 2.1	23.4 ± 2.0	42.4 ± 1.9	17.7 ± 1.2	23.1 ± 1.6	52.5 ± 1.4	18.9 ± 1.1	19.3 ± 0.8	39.7 ± 2.2	9.8 ± 1.5	25.4 ± 2.1	30.1 ± 1.6	22.1 ± 2.3	25.2 ± 2.0	38.1 ± 1.0	25.8 ± 1.4
SPTNet	27.8 ± 2.3	45.2 ± 1.5	22.0 ± 1.8	25.1 ± 1.2	44.2 ± 1.2	18.1 ± 0.8	24.9 ± 1.7	55.0 ± 2.1	20.3 ± 1.3	21.1 ± 1.1	41.6 ± 2.0	9.9 ± 1.0	26.1 ± 1.7	31.2 ± 2.3	23.3 ± 1.2	26.9 ± 1.7	39.5 ± 1.6	26.7 ± 1.7
HiLo	32.3 ± 1.5	50.6 ± 1.6	27.5 ± 1.6	28.1 ± 1.0	48.1 ± 1.4	22.3 ± 1.1	30.5 ± 2.1	58.8 ± 0.9	25.2 ± 2.2	25.7 ± 2.2	46.8 ± 1.9	13.4 ± 1.1	32.9 ± 2.1	37.1 ± 1.5	28.4 ± 1.7	33.8 ± 2.0	45.6 ± 2.2	32.1 ± 1.3
G&M	16.4 ± 1.5	34.1 ± 1.3	10.5 ± 0.9	13.7 ± 2.1	32.1 ± 1.3	7.7 ± 1.5	15.7 ± 2.0	43.8 ± 1.9	12.3 ± 1.2	11.4 ± 1.7	30.5 ± 1.5	6.7 ± 2.1	20.5 ± 2.1	24.8 ± 0.8	17.9 ± 1.1	21.6 ± 2.1	32.7 ± 2.0	22.3 ± 1.3
PA-CGCD	28.3 ± 1.7	46.5 ± 1.6	22.7 ± 1.8	25.4 ± 1.2	44.7 ± 1.9	18.4 ± 1.6	25.2 ± 1.9	55.1 ± 2.2	20.9 ± 1.0	21.2 ± 1.1	41.5 ± 2.3	10.2 ± 1.2	26.4 ± 1.3	31.4 ± 1.7	23.7 ± 1.6	27.8 ± 2.2	40.1 ± 2.3	27.2 ± 1.2
DEAN	28.9 ± 1.2	47.1 ± 2.1	23.0 ± 1.1	26.3 ± 1.5	46.2 ± 2.3	18.2 ± 1.4	26.1 ± 1.7	58.1 ± 1.9	19.4 ± 0.9	22.1 ± 1.6	41.2 ± 1.2	12.9 ± 2.0	28.1 ± 1.3	32.8 ± 1.9	28.9 ± 1.7	29.1 ± 2.3	40.1 ± 2.2	30.3 ± 1.1
PromptCCD	30.1 ± 1.1	48.1 ± 1.3	24.5 ± 1.2	27.4 ± 1.6	46.1 ± 1.4	20.3 ± 1.5	27.4 ± 1.7	57.4 ± 2.0	22.1 ± 1.1	23.1 ± 1.6	44.4 ± 1.9	11.4 ± 1.3	29.9 ± 1.8	34.5 ± 1.2	26.4 ± 2.3	30.3 ± 1.7	42.9 ± 2.0	29.9 ± 1.3
Ours	52.9 ± 0.7	68.8 ± 1.0	47.1 ± 0.7	47.2 ± 1.0	64.4 ± 1.1	40.4 ± 1.3	40.1 ± 1.4	62.3 ± 1.2	32.4 ± 1.3	35.5 ± 1.4	59.8 ± 1.1	25.6 ± 0.9	43.3 ± 1.3	52.3 ± 1.3	42.2 ± 1.3	38.9 ± 0.8	49.5 ± 1.4	37.9 ± 1.4

sion dataset. In subsequent stages, additional novel classes are gradually introduced. The class distributions for each dataset are detailed in Table 1. Furthermore, we incorporate both known-class and unknown-class samples from unseen domains, with their proportions summarized in Table 2.

These splits are defined across stages and domains to simulate real-world deployment scenarios, where novel classes and domain shifts may emerge over time. All methods are evaluated based on their ability to recognize both known and unknown classes under such distributional shifts, with particular emphasis on generalization and semantic separation capabilities.

3. Detailed Clustering Performance Analysis

To thoroughly evaluate the robustness and effectiveness of different methods, we conduct multi-stage clustering experiments on both the **SSB-C** and **DomainNet** benchmarks. Tables 3 and 4 summarize the average performance and standard deviation (*mean* $\pm$ *std*) across multiple runs, demonstrating that our method achieves consistently superior and stable results compared to existing baselines.

Furthermore, Tables 5 and 6 provide a detailed stage-wise accuracy breakdown. We report clustering accuracy (%) on all classes (*All*), known classes (*Old*), and novel classes (*New*) at each incremental stage, along with the overall average. Specifically, Table 5 presents results on the DomainNet benchmark under various domain shifts, while Table 6 includes results on the FGVC-Aircraft-C, Stanford Cars-C, and CUB-C datasets. These detailed evaluations further validate the effectiveness of our method in discovering novel classes under distribution shifts and incremental stages.

4. More Ablation Study on the WSM Module

To further validate the effectiveness of the proposed WSM module, we conduct detailed ablation experiments by comparing three settings: (*i*) removing WSM entirely, (*ii*) using a single-scale margin threshold τ, and (*iii*) employing our multi-scale design. As shown in Table 7, removing WSM leads to a clear performance drop, demonstrating the necessity of adopting a divide-and-conquer strategy for known and unknown samples during clustering.

Moreover, when comparing single-scale and multi-scale settings, the multi-scale design consistently yields superior performance. This advantage stems from its ability to characterize the margin probability of unlabeled samples from multiple granularities, providing a more reliable estimation of semantic separability. Such multi-scale modeling becomes particularly important when the labeled and unlabeled domains exhibit non-identical distributions, effectively enhancing robustness against distributional mismatch.

These results collectively confirm that WSM plays a critical role in stabilizing pseudo-label quality and improving overall category discovery performance.

5. Attention Map Visualization

To understand the spatial reasoning of our model, we visualize the attention maps from the final transformer layer, focusing on the interaction between the `[CLS]` token and the patch tokens across all attention heads. For each input, we compute attention scores and mark the top 10% most attended spatial locations in red to highlight regions that the model deems most informative.

Figure 1 presents examples from both known and unknown domains, covering both known and unknown categories. Across all conditions, our model exhibits a stable focus on semantically coherent object parts, suggesting a reduced sensitivity to domain-specific variations or style shifts. Unlike methods that overfit to low-level patterns, our attention distributions consistently localize object-centric regions critical for recognition.

This robustness in attention alignment across diverse settings illustrates the effectiveness of our model in suppressing distractions from irrelevant backgrounds. By concentrating attention on task-relevant structures, the learned features become more generalizable, leading to improved performance under domain shift and category novelty.

6. Integrating Contemporary Domain Adaptation Methods for OCCD

To further benchmark the performance of our proposed method, we incorporate several state-of-the-art domain

Table 5. Stage-wise clustering accuracy (%) of all methods on DomainNet benchmark. We report the accuracy on all classes (All), known classes (Old), and novel classes (New) at each incremental stage, as well as the average.

Method	Stage 1			Stage 2			Stage 3			Average			Stage 1			Stage 2			Stage 3			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
	Real → Painting																							
	Real												*Painting*											
GCD	54.6	71.0	48.6	51.1	67.0	45.5	48.2	63.6	42.0	51.3	67.2	45.4	28.9	28.1	29.8	27.4	26.8	28.1	25.9	25.2	26.4	27.4	26.7	28.1
SimGCD	50.9	66.0	43.7	48.3	63.9	41.2	46.0	61.8	39.0	48.4	63.9	41.3	24.4	24.3	25.3	22.5	22.5	23.5	20.9	20.5	21.7	22.6	22.4	23.5
SPTNet	52.7	68.0	45.8	50.3	64.1	41.9	46.4	61.4	39.7	49.8	64.5	42.5	25.8	24.9	25.8	24.0	23.5	24.2	22.5	22.0	22.9	24.1	23.5	24.3
HiLo	58.8	74.1	52.5	56.1	70.6	49.0	53.5	68.0	46.1	56.1	70.9	49.2	32.8	33.6	32.5	30.9	31.8	30.6	29.4	29.9	29.3	31.0	31.8	30.8
G&M	50.0	65.0	43.8	47.0	62.4	41.6	44.3	59.6	38.3	47.1	62.3	41.2	28.2	26.8	27.9	26.3	25.5	26.0	24.4	24.2	24.7	26.3	25.5	26.2
PA-CGCD	57.9	73.1	50.7	55.4	70.6	48.4	52.9	67.2	45.2	55.4	70.3	48.1	31.8	32.7	31.8	30.0	30.7	30.1	28.5	29.0	28.7	30.1	30.8	30.2
DEAN	59.1	75.6	51.2	55.9	71.7	47.6	53.0	67.8	44.9	56.0	71.7	47.9	34.1	35.9	33.2	32.9	34.3	31.5	31.5	33.0	29.8	32.8	34.4	31.5
PromptCCD	60.2	73.8	53.6	56.5	71.3	49.7	52.8	68.5	47.6	56.5	71.2	50.3	33.3	34.0	33.0	31.6	32.0	31.3	29.6	30.2	29.3	31.5	32.1	31.2
Ours	64.1	76.7	57.5	60.2	73.5	54.5	56.3	69.8	50.6	60.2	73.3	54.2	40.3	40.4	39.4	38.8	38.4	37.8	37.6	36.7	35.9	38.9	38.5	37.7
	Real → Sketch																							
	Real												*Sketch*											
GCD	55.7	68.6	45.1	51.9	65.4	41.4	49.3	63.1	38.7	52.3	65.7	41.7	10.9	15.9	11.7	9.2	14.7	10.2	7.5	12.9	8.4	9.2	14.5	10.1
SimGCD	51.2	63.5	39.0	48.6	60.4	37.0	45.8	56.6	33.5	48.5	60.2	36.5	8.7	12.7	10.8	7.1	11.3	9.4	5.9	9.8	7.4	7.2	11.3	9.2
SPTNet	53.2	65.9	41.1	49.7	62.3	38.0	46.8	58.6	34.3	49.9	62.3	37.8	9.4	13.3	11.4	8.0	11.5	9.6	6.3	10.3	7.8	7.9	11.7	9.6
HiLo	58.7	74.1	50.4	56.1	71.4	47.6	52.3	68.4	43.6	55.7	71.3	47.2	14.7	19.1	13.8	13.0	17.2	12.4	11.6	15.4	10.9	13.1	17.2	12.4
G&M	51.3	66.1	45.6	49.0	64.0	42.0	46.4	60.1	39.3	48.9	63.4	42.3	12.7	16.7	12.4	11.0	15.0	10.5	9.0	13.6	8.6	10.9	15.1	10.5
PA-CGCD	57.4	73.0	49.7	55.2	70.9	46.8	52.7	68.2	43.3	55.1	70.7	46.6	13.7	17.7	12.8	12.5	15.9	11.4	10.8	14.7	9.5	12.3	16.1	11.2
DEAN	59.1	75.0	50.1	56.9	71.6	47.7	54.1	67.9	45.1	56.7	71.5	47.6	14.8	18.5	12.7	12.8	16.8	11.4	11.1	15.1	9.5	12.9	16.8	11.2
PromptCCD	60.5	76.5	51.8	57.4	73.4	48.5	54.3	70.9	45.4	57.4	73.6	48.6	15.0	19.5	13.7	13.5	17.7	12.0	11.7	15.9	10.5	13.4	17.7	12.1
Ours	62.2	76.4	53.6	59.0	73.8	49.9	56.8	70.1	47.1	59.3	73.4	50.2	18.1	20.7	16.6	16.2	19.3	15.3	14.6	17.8	13.7	16.3	19.3	15.2
	Real → Quickdraw																							
	Real												*Quickdraw*											
GCD	41.4	58.9	32.7	38.6	56.5	29.5	36.1	53.2	26.6	38.7	56.2	29.6	6.2	6.0	7.2	5.0	4.7	5.8	3.8	3.4	4.4	5.0	4.7	5.8
SimGCD	35.7	53.2	27.0	32.3	50.5	23.7	29.2	47.2	19.8	32.4	50.3	23.5	5.6	5.0	6.2	4.2	4.0	5.1	2.8	3.0	4.0	4.2	4.0	5.1
SPTNet	37.4	55.9	28.5	35.2	52.8	24.7	31.8	49.1	21.2	34.8	52.6	24.8	6.1	5.7	6.6	4.9	4.6	5.5	3.7	3.5	4.3	4.9	4.6	5.5
HiLo	46.4	63.7	38.1	44.1	61.1	35.8	41.1	58.9	32.3	43.9	61.2	35.4	7.0	6.7	8.1	6.0	5.5	6.8	4.7	4.3	5.3	5.9	5.5	6.7
G&M	37.8	53.4	30.0	34.0	50.1	27.0	30.5	47.1	24.9	34.1	50.2	27.3	4.8	5.3	6.4	3.5	4.2	5.2	2.2	2.8	4.0	3.5	4.1	5.2
PA-CGCD	46.8	63.1	37.6	43.6	60.9	34.3	40.3	57.2	30.7	43.6	60.4	34.2	6.2	6.4	7.3	5.1	5.0	5.9	4.0	3.6	4.8	5.1	5.0	6.0
DEAN	46.8	63.0	36.3	43.2	60.5	34.2	40.7	57.7	32.1	43.6	60.4	34.2	6.5	6.1	7.4	5.0	5.0	6.1	3.9	3.9	4.9	5.1	5.0	6.1
PromptCCD	48.0	64.6	38.9	45.0	62.3	36.6	42.6	60.1	34.6	45.2	62.3	36.7	7.2	6.3	7.6	5.7	5.0	6.5	4.6	4.0	5.4	5.8	5.1	6.5
Ours	56.6	76.4	52.1	53.0	72.6	48.8	50.9	70.3	45.7	53.5	73.1	48.9	8.2	7.4	8.4	7.0	6.2	7.1	5.5	5.0	5.9	6.9	6.2	7.1
	Real → Clipart																							
	Real												*Clipart*											
GCD	50.0	68.5	43.2	43.7	62.5	37.4	46.4	66.1	39.8	46.7	65.7	40.1	15.7	22.6	11.1	13.3	19.8	9.1	14.5	21.3	10.1	14.5	21.2	10.1
SimGCD	43.6	62.2	37.0	36.9	55.8	29.8	40.1	58.4	33.7	40.2	58.8	33.5	11.5	20.1	9.5	9.0	17.5	7.0	10.4	18.8	8.2	10.3	18.8	8.2
SPTNet	46.8	63.3	38.3	39.3	57.3	33.7	43.2	60.4	35.7	43.1	60.3	35.9	12.8	20.6	10.1	10.3	18.1	7.7	11.7	19.2	8.9	11.6	19.3	8.9
HiLo	58.5	73.3	50.6	52.6	68.1	45.6	55.1	71.0	48.2	55.4	70.8	48.1	19.6	26.4	14.1	16.8	23.8	11.9	18.2	25.1	13.0	18.2	25.1	13.0
G&M	43.8	64.0	36.9	36.5	58.5	31.8	40.5	60.8	33.9	40.3	61.1	34.2	13.7	20.6	10.1	10.9	17.8	7.4	12.3	19.2	8.9	12.3	19.2	8.8
PA-CGCD	55.4	73.4	47.9	49.3	67.6	42.0	52.0	70.0	44.0	52.2	70.3	44.6	18.9	25.6	13.6	16.7	23.3	11.1	17.8	24.6	12.2	17.8	24.5	12.3
DEAN	58.2	76.3	50.4	52.0	69.1	44.6	55.1	72.7	47.5	55.1	72.7	47.5	21.6	27.9	16.2	19.0	25.4	13.8	20.3	26.8	15.0	20.3	26.7	15.0
PromptCCD	56.6	73.9	49.5	51.3	68.0	44.0	54.5	71.7	46.6	54.1	71.2	46.7	21.1	27.3	15.8	18.5	24.9	13.1	19.8	26.1	14.3	19.8	26.1	14.4
Ours	60.3	76.1	53.5	54.4	69.2	48.0	57.5	72.2	50.6	57.4	72.5	50.7	24.8	31.0	19.7	22.2	28.7	17.6	23.5	29.7	18.6	23.5	29.8	18.6
	Real → Infograph																							
	Real												*Infograph*											
GCD	42.5	57.4	35.3	36.6	53.0	29.2	40.3	55.4	32.6	39.8	55.3	32.4	9.4	11.0	7.5	6.8	8.6	5.2	8.0	9.8	6.5	8.1	9.8	6.4
SimGCD	36.7	52.3	30.5	30.7	46.0	24.5	33.4	49.3	28.4	33.6	49.2	27.8	8.1	9.0	6.7	5.3	6.7	3.7	6.7	7.7	5.2	6.7	7.8	5.2
SPTNet	38.3	54.1	33.2	33.3	48.9	26.8	36.1	51.2	29.4	35.9	51.4	29.8	8.5	9.3	7.1	6.0	6.8	4.6	7.1	7.9	6.0	7.2	8.0	5.9
HiLo	52.5	67.6	45.8	46.1	62.9	38.6	49.7	65.2	41.9	49.4	65.2	42.1	11.2	13.7	9.2	8.4	11.2	7.0	9.8	12.6	8.1	9.8	12.5	8.1
G&M	36.1	52.7	31.1	28.6	47.2	24.0	32.5	50.5	27.7	32.4	50.1	27.6	8.8	10.3	6.8	6.3	8.0	4.3	7.4	9.2	5.4	7.5	9.2	5.5
PA-CGCD	48.0	64.7	41.7	42.2	57.8	34.7	45.4	61.4	37.9	45.2	61.3	38.1	10.3	12.8	8.2	7.6	10.8	5.8	9.0	11.8	7.2	9.0	11.8	7.1
DEAN	49.9	65.8	43.3	43.4	59.0	38.1	46.9	62.2	41.0	46.7	62.3	40.8	10.7	13.6	9.0	8.2	11.4	6.7	9.6	12.5	7.9	9.5	12.5	7.9
PromptCCD	50.4	66.1	43.7	43.5	60.2	37.0	47.4	63.0	39.9	47.1	63.1	40.2	10.4	13.5	9.0	8.0	10.8	6.6	9.2	12.3	7.7	9.2	12.2	7.8
Ours	61.6	76.5	55.4	56.8	70.7	48.6	59.5	73.4	52.3	59.3	73.5	52.1	11.9	14.5	10.4	9.4	12.0	8.1	10.8	13.4	9.1	10.7	13.3	9.2

adaptation (DA) techniques into the OCCD setting and compare them against our approach. Specifically, we integrate the following state-of-the-art DA methods into our evaluation pipeline: Mixstyle [6], Unknown-Aware Domain Adversarial Learning (cUADAL) [1], Unknown Oriented Learning (UOL) [3], and Adjustment and Alignment

Table 6. Stage-wise clustering accuracy (%) of all methods on FGVC-Aircraft-C, Stanford Cars-C, and CUB-C datasets. We report the accuracy on all classes (All), known classes (Old), and novel classes (New) at each incremental stage, as well as the average.

Method	Stage 1			Stage 2			Stage 3			Average			Stage 1			Stage 2			Stage 3			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
	FGVC-Aircraft-C																							
	Original												*Corrupted*											
GCD	30.6	37.3	28.3	24.5	30.4	21.5	27.9	33.2	25.0	27.7	33.6	24.9	31.8	45.1	32.0	26.0	37.5	25.3	28.6	41.7	29.0	28.8	41.4	28.8
SimGCD	29.2	33.8	25.9	21.8	26.6	18.4	25.2	29.9	22.0	25.4	30.1	22.1	28.3	42.3	29.1	21.5	34.0	22.6	25.8	38.0	25.7	25.2	38.1	25.8
SPTNet	29.7	35.5	27.1	22.3	26.8	19.2	26.3	31.3	23.6	26.1	31.2	23.3	31.0	43.3	29.7	23.1	36.3	23.4	26.7	38.9	27.0	26.9	39.5	26.7
HiLo	36.5	40.1	31.5	29.2	34.3	25.2	33.1	36.9	28.5	32.9	37.1	28.4	37.1	48.7	35.6	30.5	42.5	28.5	33.8	45.6	32.1	33.8	45.6	32.1
G&M	24.1	28.5	21.7	17.1	21.6	14.2	20.2	24.2	17.8	20.5	24.8	17.9	25.4	35.6	25.7	17.7	29.6	19.2	21.7	32.9	22.0	21.6	32.7	22.3
PA-CGCD	29.4	34.9	27.5	23.1	28.2	20.4	26.6	31.1	23.1	26.4	31.4	23.7	31.5	43.9	31.6	24.4	36.2	22.8	27.5	40.1	27.2	27.8	40.1	27.2
DEAN	31.2	35.3	32.8	25.1	30.3	25.3	28.0	32.8	28.7	28.1	32.8	28.9	32.5	42.9	34.1	25.4	37.2	27.1	29.5	40.2	29.7	29.1	40.1	30.3
PromptCCD	32.8	38.2	29.8	27.1	30.7	23.4	29.8	34.6	26.0	29.9	34.5	26.4	33.7	45.9	33.1	26.4	39.5	26.8	30.8	43.3	29.8	30.3	42.9	29.9
Ours	47.1	55.8	45.6	39.5	49.2	38.9	43.3	51.9	42.2	43.3	52.3	42.2	41.9	53.0	41.1	35.9	45.9	34.5	38.9	49.6	38.1	38.9	49.5	37.9
	Stanford Cars-C																							
	Original												*Corrupted*											
GCD	30.3	59.4	25.0	26.1	56.3	21.6	22.7	52.7	17.9	26.4	56.1	21.5	26.1	46.1	14.5	22.5	43.1	11.6	18.3	40.1	7.5	22.3	43.1	11.2
SimGCD	26.7	55.7	22.8	22.6	52.2	19.1	20.0	49.6	14.7	23.1	52.5	18.9	23.0	42.9	13.1	19.6	39.7	9.4	15.3	36.5	6.9	19.3	39.7	9.8
SPTNet	28.3	58.5	24.0	24.9	54.8	20.3	21.6	51.7	16.6	24.9	55.0	20.3	24.6	45.4	13.7	20.7	41.9	9.6	18.0	37.5	6.4	21.1	41.6	9.9
HiLo	34.7	62.2	28.1	30.5	59.1	25.6	26.3	55.1	21.9	30.5	58.8	25.2	28.5	50.5	17.6	25.8	47.0	13.4	22.8	42.9	9.2	25.7	46.8	13.4
G&M	18.6	48.1	16.3	15.9	43.9	12.5	12.5	39.5	8.1	15.7	43.8	12.3	15.2	33.3	10.0	10.9	30.7	6.6	8.1	27.5	3.6	11.4	30.5	6.7
PA-CGCD	27.9	58.7	24.4	25.3	55.1	21.0	22.4	51.5	17.3	25.2	55.1	20.9	25.6	45.2	13.9	21.1	41.5	10.2	16.9	37.8	6.5	21.2	41.5	10.2
DEAN	29.1	61.8	23.1	26.0	58.5	19.1	23.2	54.1	16.0	26.1	58.1	19.4	25.6	45.3	16.3	22.3	41.3	12.7	18.4	37.0	9.7	22.1	41.2	12.9
PromptCCD	31.7	61.2	26.3	27.2	56.9	22.2	23.2	54.1	17.8	27.4	57.4	22.1	27.3	48.2	15.3	23.1	44.1	11.3	18.9	40.8	7.6	23.1	44.4	11.4
Ours	43.1	65.9	35.4	40.3	62.4	32.5	36.9	58.6	29.3	40.1	62.3	32.4	39.1	63.6	28.3	35.2	59.8	25.7	32.3	56.0	22.9	35.5	59.8	25.6
	CUB-C																							
	Original												*Corrupted*											
GCD	32.6	51.2	27.2	29.5	47.7	23.7	26.1	44.2	19.4	29.4	47.7	23.4	30.5	49.6	23.8	26.3	46.2	19.9	23.7	41.9	16.6	26.8	45.9	20.1
SimGCD	29.5	47.7	24.9	26.6	44.8	20.5	23.7	41.0	17.6	26.6	44.5	21.0	27.1	46.1	22.0	23.2	42.8	17.6	19.9	38.4	13.5	23.4	42.4	17.7
SPTNet	31.2	48.5	25.5	28.0	45.0	22.5	24.2	42.1	18.0	27.8	45.2	22.0	28.4	47.5	21.0	24.9	44.7	18.2	22.0	40.4	15.1	25.1	44.2	18.1
HiLo	35.4	54.0	30.5	32.9	50.2	27.4	28.6	47.6	24.6	32.3	50.6	27.5	32.0	51.8	25.2	27.6	48.3	22.1	24.7	44.2	19.5	28.1	48.1	22.3
G&M	19.1	38.4	14.6	16.4	34.0	10.4	13.7	29.9	6.5	16.4	34.1	10.5	17.5	36.2	10.5	13.2	32.2	7.9	10.4	27.9	4.7	13.7	32.1	7.7
PA-CGCD	32.4	50.1	26.0	28.0	46.2	22.7	24.5	43.2	19.4	28.3	46.5	22.7	28.3	48.3	21.9	25.4	44.8	18.3	22.5	41.0	15.0	25.4	44.7	18.4
DEAN	32.3	51.4	27.2	28.5	47.1	23.0	25.9	42.9	18.8	28.9	47.1	23.0	30.4	49.4	21.4	26.0	46.8	18.3	22.5	42.3	14.9	26.3	46.2	18.2
PromptCCD	33.3	51.0	28.3	30.6	48.3	24.7	26.4	45.0	20.5	30.1	48.1	24.5	31.4	49.6	23.6	27.4	46.2	20.1	23.4	42.5	17.2	27.4	46.1	20.3
Ours	56.7	71.7	49.8	52.3	69.0	47.3	49.7	65.7	44.2	52.9	68.8	47.1	49.9	68.1	43.9	47.1	64.2	40.2	44.6	60.9	37.1	47.2	64.4	40.4

Table 7. Ablation study on WSM module.

Method	Real			Painting		
	All	Old	New	All	Old	New
w/o WSM	54.6	68.7	46.5	28.7	28.1	27.9
0.01	56.7	71.3	50.4	34.9	34.5	35.8
0.05	56.8	72.0	50.3	36.2	35.0	36.1
0.1	57.6	71.0	52.7	36.8	36.0	36.8
0.5	57.4	69.0	53.2	36.5	35.6	36.1
1	57.2	69.8	52.8	36.4	35.3	36.8
5	58.0	71.1	53.0	35.4	35.0	35.4
10	57.3	70.2	52.4	35.0	34.6	35.9
Ours	**60.2**	**73.3**	**54.2**	**38.9**	**38.5**	**37.7**

(ANNA) [2].

The experimental results are summarized in Table 8. Although these approaches have shown strong results in standard DA scenarios, they fall short in the context of OCCD, where continual discovery of unknown categories and robust known-class recognition under distributional shifts are essential. As shown in the table, most of these methods provide only marginal gains and sometimes even exhibit unstable performance under evolving streams. This is likely due to the fact that OCCD imposes a more demanding setting, where the model must adapt not only to domain shifts but also to uncover new classes.

In contrast, our method offers a unified and tailored solution for OCCD by explicitly addressing unknown category emergence, dynamic distributional change, and cross-domain generalization. The superior performance across all metrics validates the effectiveness and robustness of our approach in this challenging environment.

7. Empirical Study with DINOv2 Backbone

To evaluate the robustness and generality of our proposed framework, we further conducted experiments using a stronger visual encoder, DINOv2, to replace the original DINO backbone. We re-ran all benchmark settings with this modification, and the results are summarized in Table 9. As expected, the enhanced feature capacity of DINOv2 leads

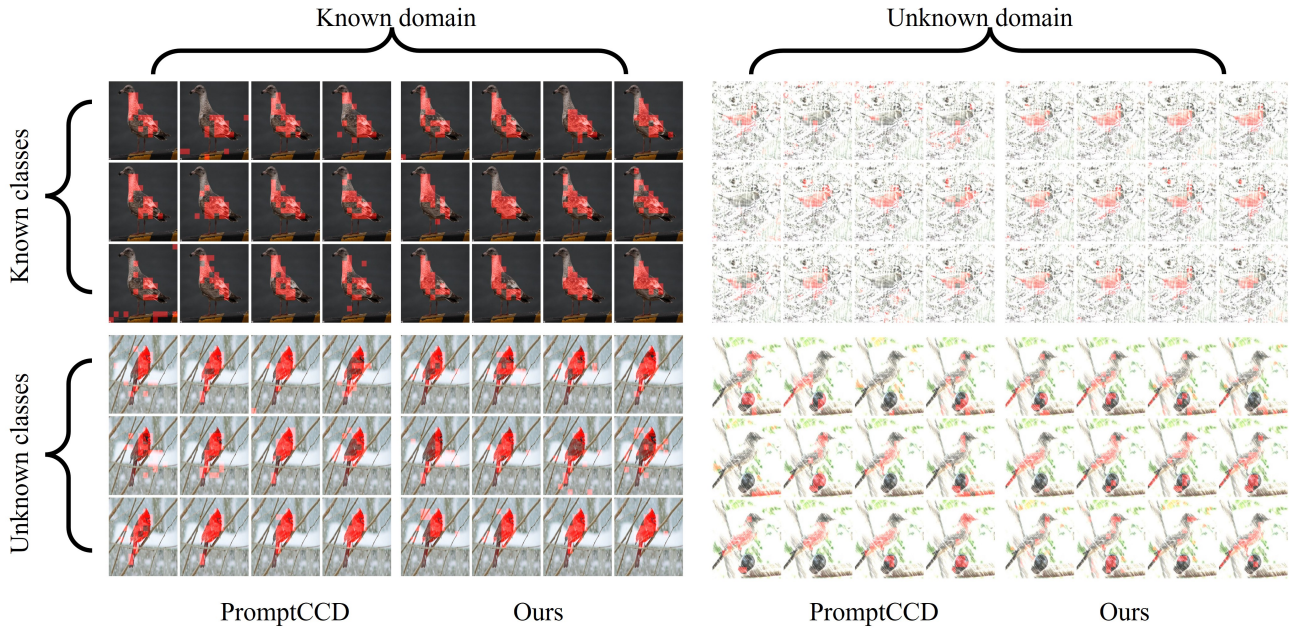

Figure 1. Attention visualization from the final block of the ViT backbone on the CUB-C dataset. The top 10% most attended patches from different heads are highlighted in red. Our method shows stronger focus on foreground regions across both known and unknown domains, demonstrating improved robustness to background style variations.

Table 8. Clustering performance of other DA methods.

Method	Real			Painting		
	All	Old	New	All	Old	New
UOL	55.1	69.0	45.6	29.0	29.3	27.6
Mixstyle	53.2	67.1	44.0	26.0	25.2	25.1
cUADAL	55.7	70.1	46.4	29.9	28.8	26.3
ANNA	54.8	68.4	48.7	30.7	28.9	26.8
Ours	**60.2**	**73.3**	**54.2**	**38.9**	**38.5**	**37.7**

to overall performance improvements. Notably, our method continues to outperform all competing approaches under this more powerful configuration. This suggests that the effectiveness of our framework is not tied to a specific backbone design and can generalize well across different feature extractors. The consistent advantage across architectures demonstrates the scalability and adaptability of our approach.

8. Evaluation under Multiple Unknown Domains

To assess the adaptability of our method in more complex settings, we consider a challenging configuration where the unknown domain comprises a combination of several distinct subdomains. Specifically, we construct a new unknown domain $\mathcal{T}_B$ by merging all five non-*real* domains from the DomainNet dataset. This setup introduces greater diversity and simulates practical scenarios where the model encounters samples from multiple unseen distributions simultaneously.

The comparative results, summarized in Table 10, reveal that our method consistently surpasses all baselines across this multi-domain setup. The stable performance in such a diverse and unstructured environment underscores the strong generalization ability of our framework, demonstrating its effectiveness not only under conventional domain shift but also in more intricate unknown domain compositions.

9. Evaluation under More-Stage Setting

In our primary experiments, we adopt a three-stage setting for the online category discovery process to evaluate the ef-

Table 9. Clustering performance on SSB-C benchmarks using DINOv2 as the backbone. Each dataset includes both Original and Corrupted settings, and we report the average All / Old / New accuracy across all stages for both domains.

Method	CUB-C						Stanford Cars-C						FGVC-Aircraft-C					
	Original			Corrupted			Original			Corrupted			Original			Corrupted		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	41.2	55.6	35.9	38.5	53.9	31.5	46.3	65.4	43.4	42.3	53.7	32.3	38.1	39.8	36.1	39.4	49.3	40.4
SimGCD	39.3	52.4	33.6	35.3	51.4	30.5	42.5	63.3	41.0	39.5	50.0	30.9	36.1	37.4	34.1	36.1	44.2	34.7
SPTNet	40.5	52.4	34.3	36.9	51.5	30.6	45.0	64.5	41.4	40.4	51.7	32.1	36.0	36.9	31.3	37.8	48.4	36.4
HiLo	45.2	58.8	39.5	39.4	55.9	34.6	51.4	69.1	47.0	44.8	56.7	34.7	42.7	42.3	40.0	42.9	51.3	44.1
G&M	28.2	42.6	22.4	25.3	39.3	20.7	36.1	52.8	34.6	30.9	41.0	29.5	31.4	30.9	28.6	31.8	40.3	31.7
PA-CGCD	39.7	53.8	35.2	38.0	53.2	31.1	44.2	65.6	43.1	40.7	51.6	31.3	37.3	40.2	35.0	37.0	45.7	37.3
DEAN	41.7	54.5	35.5	37.8	53.6	29.6	45.8	67.4	41.7	41.7	50.8	35.7	39.0	40.2	38.1	38.8	45.6	41.1
PromptCCD	41.2	55.6	36.0	39.1	54.3	33.2	48.4	67.8	43.7	43.3	53.7	34.1	39.4	42.4	35.9	40.0	48.3	41.0
Ours	**65.9**	**76.3**	**58.2**	**59.1**	**73.3**	**52.5**	**59.2**	**71.7**	**53.4**	**56.4**	**70.4**	**46.8**	**53.9**	**59.6**	**52.4**	**49.4**	**56.0**	**46.0**

Table 10. Clustering performance when $\mathcal{T}_B$ contains multiple unknown domains. We construct $\mathcal{T}_B$ by combining the five domains from DomainNet excluding the Real domain, and report clustering accuracy separately for each domain.

Methods	Real			Painting			Sketch			Quickdraw			Clipart			Infograph		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	50.9	66.7	44.9	26.9	26.3	27.7	8.8	14.0	9.6	4.7	2.3	3.6	12.2	20.8	9.7	7.6	9.4	6.0
SimGCD	48.0	63.5	40.8	22.1	22.1	23.2	6.7	10.9	8.9	3.7	1.6	2.7	8.0	18.3	7.8	6.4	7.4	4.8
SPTNet	49.4	64.1	42.1	23.7	23.2	24.0	7.6	11.2	9.2	4.5	2.4	3.0	9.3	18.9	8.5	6.9	7.5	5.5
HiLo	55.7	70.6	48.9	30.5	31.5	30.4	12.6	16.7	11.9	5.4	2.7	**5.1**	16.0	24.7	12.6	9.3	12.1	7.8
G&M	46.7	61.9	40.8	25.9	25.1	25.8	10.5	14.7	10.2	3.1	1.8	2.8	10.1	18.7	8.4	7.1	8.7	5.1
PA-CGCD	55.0	69.9	47.6	29.6	30.4	29.7	11.9	15.7	10.8	4.8	2.5	3.7	15.4	24.1	12.0	8.6	11.4	6.8
DEAN	55.5	71.2	47.5	32.3	33.9	31.1	12.5	16.4	10.8	4.7	2.6	3.9	18.0	26.3	14.6	9.1	**12.9**	7.4
PromptCCD	56.2	70.7	49.8	31.1	31.8	30.8	13.0	17.3	11.7	5.3	2.7	4.2	17.6	25.8	14.0	8.8	11.8	7.4
Ours	**59.8**	**72.9**	**53.7**	**38.6**	**38.2**	**37.4**	**15.9**	**18.9**	**14.8**	**6.4**	**4.0**	4.8	**21.1**	**29.4**	**18.2**	**10.4**	12.8	**8.9**

Table 11. Clustering performance on the DomainNet benchmark under extended-stage online discovery. We use Real as the known domain $\mathcal{T}_A$ and treat each of the other domains as the unknown domain $\mathcal{T}_B$. Results are averaged over all discovery stages (including the extended 4-stage setting), and we report the All / Old / New classification accuracy for each domain pair.

Method	Real → Painting						Real → Sketch						Real → Quickdraw						Real → Clipart						Real → Infograph					
	Real			Painting			Real			Sketch			Real			Quickdraw			Real			Clipart			Real			Infograph		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	50.9	66.8	44.9	27.8	27.2	28.5	52.8	66.3	42.2	8.7	13.9	9.7	38.1	55.9	29.2	5.4	5.3	6.3	47.0	66.2	40.6	13.9	20.7	9.5	39.4	55.0	31.9	8.6	10.2	6.8
SimGCD	47.9	63.5	40.8	22.9	23.0	24.0	49.0	60.5	36.9	6.7	10.9	8.7	32.1	49.8	23.2	4.5	4.5	5.6	40.6	59.3	34.1	9.7	18.4	7.8	33.1	48.8	27.3	7.1	8.2	5.7
SPTNet	49.3	64.1	42.0	24.5	23.9	24.9	50.4	62.8	38.3	7.6	11.1	9.3	34.3	52.3	24.4	5.3	5.0	6.0	43.4	60.6	36.2	11.1	19.0	8.4	35.4	50.9	29.3	7.5	8.5	6.5
HiLo	55.6	70.5	48.6	31.6	32.4	31.1	56.1	**74.6**	45.6	12.6	16.7	11.8	43.3	60.7	34.9	6.4	5.8	7.2	55.9	71.2	48.5	17.7	24.6	12.6	48.8	64.9	41.6	10.2	13.1	8.4
G&M	46.7	62.0	40.8	26.7	26.1	26.7	49.3	63.8	42.9	10.5	14.6	10.2	33.6	49.8	26.9	3.9	4.7	5.7	40.8	61.5	34.8	11.8	18.7	8.5	31.9	49.7	27.1	7.9	9.8	5.9
PA-CGCD	55.1	70.0	47.7	30.4	31.3	30.6	55.7	71.1	47.1	11.7	15.6	10.8	43.0	59.9	33.7	5.6	5.3	6.5	52.6	**74.8**	42.1	17.2	24.1	12.0	44.6	61.0	37.5	9.4	12.3	7.7
DEAN	55.6	71.2	47.6	33.2	34.9	31.9	57.1	72.0	48.2	12.4	16.4	10.6	43.0	60.1	33.7	5.6	5.4	6.6	55.7	73.3	48.0	19.9	26.3	14.7	46.2	62.0	40.3	10.1	**14.5**	7.4
PromptCCD	56.1	70.8	49.9	31.9	32.6	31.7	58.0	73.9	49.2	13.1	17.4	11.7	44.7	61.9	36.3	6.2	5.5	7.0	54.7	71.5	47.0	19.4	25.5	14.0	46.7	62.6	39.7	9.7	12.6	8.2
Ours	**60.1**	**73.0**	**53.9**	**39.3**	**39.1**	**38.1**	**59.6**	73.8	**50.7**	**15.9**	**19.0**	**14.7**	**52.9**	**72.6**	**48.4**	**7.3**	**6.6**	**7.6**	**57.7**	73.1	**51.0**	**23.1**	**29.3**	**18.0**	**58.7**	**73.1**	**51.6**	**11.3**	13.8	**9.6**

fectiveness of the proposed method. To further assess its scalability and robustness in more complex discovery scenarios, we conduct an additional experiment with four sequential discovery stages, simulating a more gradual and challenging emergence of unknown categories.

As reported in Table 11, our method maintains consistent superiority across all metrics and category splits (*All*, *Old*, *New*) compared to existing baselines. This observation confirms the method's capability to handle more fragmented and progressive category emergence, demonstrating its strong adaptability in dynamic, multi-stage discovery environments.

10. Computational Complexity Analysis

We analyze the computational overhead introduced by each component of our framework using an RTX 4090 GPU with a batch size of 128. The *Instance Probability Modeling (IPM)* module, which solves a relaxed partial unbalanced optimal transport (PUOT) problem, requires approximately

0.6 GFLOPs and takes around 0.10 seconds per iteration. The *Response Spectrum Quantization (RSQ)* module evaluates marginal probabilities under multiple τ settings and performs K-means clustering in a low-dimensional space, incurring about 0.2 GFLOPs and 0.05 seconds per iteration. The *Cross-Domain Semantic Alignment (CDSA)* module introduces a lightweight domain discriminator with adversarial optimization over prototypes, contributing roughly 0.5 GFLOPs and 0.08 seconds per iteration. The *Category Topology Consistency Constraint (CTCC)* involves FFT-based frequency transformations, amplitude augmentation, and Gaussian potential-based topology regularization, leading to approximately 0.8 GFLOPs and 0.12 seconds per iteration.

In total, our full framework adds about 2.1 GFLOPs and increases the per-iteration training time by approximately 0.35 seconds. Thanks to GPU-accelerated operations such as `torch.fft`, momentum-based prototype updates, and efficient OT solvers, the computational overhead remains manageable, making our method scalable to large-scale open-world settings.

11. Experiments on Dynamic Domain-Incremental Setting

To further assess the adaptability of our method under dynamic domain shifts, we conducted a domain-incremental evaluation on the CUB-C dataset across three progressive stages. At each stage, additional corruption types were introduced to emulate evolving data distributions. Specifically, Stage 1 contained Gaussian, Shot, and Impulse Noise; Stage 2 added Zoom Blur, Snow, and Frost; and Stage 3 further incorporated Fog, Speckle, and Spatter. This setup effectively mimics a non-stationary data stream in which new domains continuously emerge. As summarized in Table 12, our approach consistently achieved strong gains over competing baselines across all stages, demonstrating its robustness and sustained effectiveness under dynamic domain evolution.

References

[1] JoonHo Jang, Byeonghu Na, Dong Hyeok Shin, Mingi Ji, Kyungwoo Song, and Il-Chul Moon. Unknown-aware domain adversarial learning for open-set domain adaptation. *Advances in Neural Information Processing Systems*, 35:16755–16767, 2022. 5

[2] Wuyang Li, Jie Liu, Bo Han, and Yixuan Yuan. Adjustment and alignment for unbiased open set domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24110–24119, 2023. 6

[3] Jie Liu, Xiaoqing Guo, and Yixuan Yuan. Unknown-oriented learning for open set domain adaptation. In *European Conference on Computer Vision*, pages 334–350. Springer, 2022. 5

[4] 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. 2

[5] Hongjun Wang, Sagar Vaze, and Kai Han. Hilo: A learning framework for generalized category discovery robust to domain shifts. *arXiv preprint arXiv:2408.04591*, 2024. 2, 3

[6] Renjun Xu, Pelen Liu, Yin Zhang, Fang Cai, Jindong Wang, Shuoying Liang, Heting Ying, and Jianwei Yin. Joint partial optimal transport for open set domain adaptation. In *IJCAI*, pages 2540–2546, 2020. 5

Table 12. Stage-wise clustering accuracy (%) of different methods on the CUB-C dataset under the dynamic domain-incremental setting. At each stage, new domains are progressively introduced (Stage 1: Gaussian, Shot, Impulse Noise; Stage 2: Zoom Blur, Snow, Frost; Stage 3: Fog, Speckle, Spatter). We report accuracies on all (All), known (Old), and novel (New) classes at each stage, as well as the average across all stages.

Method	Stage 1			Stage 2			Stage 3			Average			Stage 1			Stage 2			Stage 3			Average		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
	CUB-C																							
	Original												*Corrupted*											
GCD	31.1	49.4	24.1	28.6	46.2	22.1	24.9	43.2	20.1	28.2	46.3	22.1	23.8	43.0	17.2	26.6	46.0	20.0	25.5	44.5	19.0	25.3	44.5	18.7
SimGCD	28.2	45.1	22.3	24.8	43.0	20.0	22.6	40.9	17.2	25.2	43.0	19.8	20.7	39.8	15.1	23.3	42.4	17.7	22.0	40.9	16.4	22.0	41.0	16.4
SPTNet	29.5	46.0	23.3	26.8	43.9	20.3	23.5	41.1	18.5	26.6	43.7	20.7	22.4	41.3	15.1	25.1	44.1	18.5	23.6	43.0	16.8	23.7	42.8	16.8
HiLo	32.1	48.9	26.7	28.4	45.7	23.1	24.5	41.9	20.3	28.4	45.8	22.9	23.9	43.0	16.9	26.7	45.9	19.7	25.6	44.7	18.5	25.3	44.9	18.8
G&M	17.8	36.4	12.8	14.9	32.7	9.2	12.9	29.3	5.6	15.2	32.8	9.2	10.6	29.6	5.1	13.9	31.8	7.4	12.4	30.7	6.4	12.3	30.7	6.3
PA-CGCD	29.4	48.1	23.9	26.9	44.7	21.6	24.6	42.2	18.7	27.0	45.0	21.4	22.5	41.8	15.7	25.3	44.6	18.5	24.2	43.5	16.8	24.0	43.3	17.0
DEAN	29.9	48.9	24.5	28.2	45.7	21.2	24.8	42.5	19.4	27.6	45.7	21.7	23.7	43.2	15.8	26.1	46.4	18.1	24.8	44.7	16.8	24.9	44.8	16.9
PromptCCD	33.2	49.8	26.4	28.7	46.5	22.8	26.3	43.5	20.1	28.7	46.6	23.1	24.6	43.2	17.4	27.3	46.3	20.4	26.2	44.6	18.8	26.0	44.7	18.9
Ours	**51.6**	**67.0**	**48.0**	**49.2**	**64.9**	**43.6**	**47.8**	**62.8**	**40.9**	**49.9**	**64.6**	**44.3**	**42.3**	**59.5**	**35.7**	**45.6**	**61.3**	**38.6**	**43.9**	**60.6**	**36.8**	**43.7**	**60.9**	**36.8**