Appendix for: Seeing Through the Shift: Causality-Inspired Robust Generalized Category Discovery

Wei Feng[1,2] Yiwen Jiang[1] Sijin Zhou[1,2] Zhuang Qi[3]
Zhongxing Xu[1] Zhonghua Wang[1] Feilong Tang[1] Zongyuan Ge[1,2]

[1]AIM for Health Lab, Monash University [2]Airdoc–Monash Research, Monash University
[3]School of Software, Shandong University
{wf02429, sjzhou1995}@gmail.com {yiwen.jiang, zongyuan.ge}@monash.edu z_qi@mail.sdu.edu.cn

1. Theoretical Proofs

1.1. Proof of Theorem 1

Consider two domains: a source domain $\mathbb{P}$ and a target domain $\mathbb{Q}$. Let $G(\cdot)$ denote an inference encoder, $G'(\cdot)$ an intervention encoder. With the observational and interventional inference, we define:

$$\begin{aligned}\mathrm{CDR}_{\mathbb{Q}}(u,u') &= \mathcal{O}_{\mathbb{Q}}(u) - \mathcal{I}_{\mathbb{Q}}(u') \\ &= \mathbb{E}_{(x,y)\sim\mathbb{Q}}\Big[\mathbb{E}_{u\sim P^{G}_{\mathbb{Q}}(U|X=x)} P^{g}(Y\neq y \mid U=u)\Big] - \mathbb{E}_{(x,y)\sim\mathbb{Q}}\Big[\mathbb{E}_{u'\sim P^{G'}_{\mathbb{Q}}(U'|X=x)} P^{g}(Y\neq y \mid U'=u')\Big].\end{aligned} \tag{1}$$

Decomposing $\mathbb{Q}$ with respect to the labeled known and novel label sets gives

$$\mathrm{CDR}_{\mathbb{Q}}(u,u') = \mathbb{E}_{(x,y)\sim\mathbb{Q}}\big[\mathbb{I}(y\in\mathcal{Y}_l)\,\Delta(x,y)\big] + \mathbb{E}_{(x,y)\sim\mathbb{Q}}\big[\mathbb{I}(y\in\mathcal{Y}_u^{\text{novel}})\,\Delta(x,y)\big], \tag{2}$$

where the unified discrepancy term is defined as

$$\Delta(x,y) = \mathbb{E}_{u\sim P^{G}_{\mathbb{Q}}(U|X=x)} P(Y\neq y \mid U=u) - \mathbb{E}_{u'\sim P^{G'}_{\mathbb{Q}}(U'|X=x)} P(Y\neq y \mid U'=u'). \tag{3}$$

We denote the two expectations as $\pi_{\mathbb{Q}\wedge\mathbb{P}}(X,Y)$ and $\pi_{\mathbb{Q}\setminus\mathbb{P}}(X,Y)$, respectively. Thus,

$$\mathrm{CDR}_{\mathbb{Q}}(u,u') = \pi_{\mathbb{Q}\wedge\mathbb{P}}(X,Y) + \pi_{\mathbb{Q}\setminus\mathbb{P}}(X,Y). \tag{4}$$

This completes the proof of Theorem 1.

1.2. Proof of Theorem 2

Let $\mathbb{Q}$ and $\mathbb{P}$ be the same domains as above. For any probability distribution ς over g, we define the expected joint error and classifier disagreement under $\mathbb{Q}$ as:

$$\begin{aligned}\varepsilon_{\mathbb{Q}}(u) &= \mathbb{E}_{g_1,g_2\sim\varsigma}\mathbb{E}_{(x,y)\sim\mathbb{Q}}\mathbb{I}[g_1(u)\neq y]\,\mathbb{I}[g_2(u)\neq y], \\ \delta_{\mathbb{Q}}(u) &= \mathbb{E}_{g_1,g_2\sim\varsigma}\mathbb{E}_{x\sim\mathbb{Q}}\mathbb{I}[g_1(u)\neq g_2(u)].\end{aligned} \tag{5}$$

Following the symmetrization step in Appendix 1.1, one can express

$$\begin{aligned}\mathrm{CDR}_{\mathbb{Q}}(u,u') &= \mathcal{O}_{\mathbb{Q}}(u) - \mathcal{I}_{\mathbb{Q}}(u') \\ &= \tfrac{1}{2}\mathbb{E}_{(x,y)\sim\mathbb{Q}}\mathbb{E}_{u\sim P^{G}_{\mathbb{Q}}(U|X=x)}\mathbb{E}_{g_1,g_2\sim\varsigma}[\mathbb{I}(g_1(u)\neq y) + \mathbb{I}(g_2(u)\neq y)] \\ &\quad - \tfrac{1}{2}\mathbb{E}_{(x,y)\sim\mathbb{Q}}\mathbb{E}_{u'\sim P^{G'}_{\mathbb{Q}}(U'|X=x)}\mathbb{E}_{g_1,g_2\sim\varsigma}[\mathbb{I}(g_1(u')\neq y) + \mathbb{I}(g_2(u')\neq y)].\end{aligned} \tag{6}$$

Using the Boolean relation

$$\mathbb{I}[A] + \mathbb{I}[B] = \mathbb{I}[A \triangle B] + 2\mathbb{I}[A \wedge B], \tag{7}$$

where $\triangle$ denotes exclusive-or and substituting into the above, we obtain the decomposition

$$\begin{aligned}
\mathrm{CDR}_{\mathbb{Q}}(u, u') &= \mathbb{E}_{(x,y)\sim\mathbb{Q}} \mathbb{E}_{u\sim P_{\mathbb{Q}}^{G}(U|X=x)} \left[\frac{\mathbb{E}_{g_1,g_2\sim\varsigma}\big[\mathbb{I}\big(g_1(u) \neq g_2(u)\big) + 2\,\mathbb{I}\big(g_1(u) \neq y \wedge g_2(u) \neq y\big)\big]}{2} \right] \\
&\quad - \mathbb{E}_{(x,y)\sim\mathbb{Q}} \mathbb{E}_{u'\sim P_{\mathbb{Q}}^{G'}(U'|X=x)} \left[\frac{\mathbb{E}_{g_1,g_2\sim\varsigma}\big[\mathbb{I}\big(g_1(u') \neq g_2(u')\big) + 2\,\mathbb{I}\big(g_1(u') \neq y \wedge g_2(u') \neq y\big)\big]}{2} \right] \\
&= \tfrac{1}{2}\big(\delta_{\mathbb{Q}}(u) - \delta_{\mathbb{Q}}(u')\big) + \big(\varepsilon_{\mathbb{Q}}(u) - \varepsilon_{\mathbb{Q}}(u')\big).
\end{aligned} \tag{8}$$

Invoking Hölder's inequality,

$$\int |fg|\, d\mu \;\leq\; \Big(\int |f|^q\, d\mu\Big)^{1/q} \Big(\int |g|^p\, d\mu\Big)^{1/p}, \qquad \frac{1}{p} + \frac{1}{q} = 1, \tag{9}$$

with $g = 1$, $q \to \infty$, and $p \to 1$, we obtain the following upper bound:

$$\mathbb{E}_{(x,y)\sim\mathbb{Q}}\big[f(x,y)\big] \;=\; \mathbb{E}_{(x,y)\sim\mathbb{P}}\left[\frac{\mathbb{Q}(x,y)}{\mathbb{P}(x,y)}\, f(x,y)\right] \;\leq\; \lim_{q\to\infty} \beta_q(\mathbb{Q}\|\mathbb{P})\, \mathbb{E}_{(x,y)\sim\mathbb{P}}\big[f(x,y)\big], \tag{10}$$

where the β-divergence term [2] is defined as

$$\beta_q(\mathbb{Q}\|\mathbb{P}) = \Big(\mathbb{E}_{(x,y)\sim\mathbb{P}}\big[\big(\tfrac{\mathbb{Q}(x,y)}{\mathbb{P}(x,y)}\big)^q\big]\Big)^{1/q}. \tag{11}$$

Substituting this inequality into our error decomposition yields

$$\begin{aligned}
\varepsilon_{\mathbb{Q}}(u) &= \mathbb{E}_{g_1,g_2\sim\varsigma} \mathbb{E}_{(x,y)\sim\mathbb{Q}} \mathbb{I}[g_1(u) \neq y]\, \mathbb{I}[g_2(u) \neq y] \\
&= \mathbb{E}_{(x,y)\sim\mathbb{P}}\left[\frac{\mathbb{Q}(x,y)}{\mathbb{P}(x,y)}\right] \mathbb{E}_{u\sim P_{\mathbb{Q}}^{G}(U|X=x)} \mathbb{E}_{g_1,g_2\sim\varsigma} \mathbb{I}[g_1(u) \neq y] \mathbb{I}[g_2(u) \neq y] + r\, \varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u) \\
&\leq \lim_{q\to\infty} \beta_q(\mathbb{Q}\,\|\,\mathbb{P})\, \varepsilon_{\mathbb{P}}(u) + r\, \varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u),
\end{aligned} \tag{12}$$

where $r = \mathbb{E}_{(x,y)\sim\mathbb{Q}}[\mathbb{I}(y \in \mathcal{Y}_u^{\text{novel}})]$. And similarly

$$\varepsilon_{\mathbb{Q}}(u') \leq \lim_{q\to\infty} \beta_q(\mathbb{Q}\,\|\,\mathbb{P})\, \varepsilon_{\mathbb{P}}(u') + r\, \varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u').$$

Denoting $r(\varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u) - \varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u'))$ as $\pi_{\mathbb{Q}\setminus\mathbb{P}}(X, Y)$, we obtain the overall bound:

$$\mathrm{CDR}_{\mathbb{Q}}(u, u') < \tfrac{1}{2}(\delta_{\mathbb{Q}}(u) - \delta_{\mathbb{Q}}(u')) + \lim_{q\to\infty} \beta_q(\mathbb{Q}\,\|\,\mathbb{P})(\varepsilon_{\mathbb{P}}(u) - \varepsilon_{\mathbb{P}}(u')) + \pi_{\mathbb{Q}\setminus\mathbb{P}}(X, Y). \tag{13}$$

This completes the proof of Theorem 2.

1.3. Proof of Corollary 1

PAC–Bayes foundation. Following the PAC–Bayesian theorem of [1], let $\mathbb{P}$ and $\mathbb{Q}$ denote the source and target distributions respectively, any set of voters $\mathcal{H}$, any prior distribution ν over $\mathcal{H}$, and ς a posterior distribution over classifiers $g \in \mathcal{H}$. For any distribution $\mathbb{M} \in \{\mathbb{P}, \mathbb{Q}\}$, any risk function $R : \mathcal{H} \to [0, 1]$, and any real number $c > 0$, with probability at least $1 - \delta$ over samples $\{(x_i, y_i)\}_{i=1}^n \sim \mathbb{M}$, the following inequality holds for all posterior distributions ς:

$$\mathbb{E}_{(x,y)\sim\mathbb{M}} \mathbb{E}_{g\sim\varsigma} R(\varsigma; x, y) \;\leq\; \frac{c}{1 - e^{-c}} \left[\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{g\sim\varsigma} R(\varsigma; x_i, y_i) + \frac{\mathrm{KL}(\varsigma\|\nu) + \ln(\frac{1}{\delta})}{n\, c}\right]. \tag{14}$$

Applying to disagreement and joint error. Let $\mathcal{D}_u = \{x_i\}_{i=1}^{N_u}$ and $\mathcal{D}_l = \{(x_i, y_i)\}_{i=1}^{N_l}$ be target and source empirical samples respectively. By applying (14), with constants $c' = \frac{c}{1-e^{-c}}$, we have:

$$\begin{aligned}
&\forall \varsigma \in \mathcal{H}, \quad \delta_{\mathbb{Q}}(u) \le c'\left(\delta_{\mathcal{D}_u}(u) + \frac{2\,\mathrm{KL}(\varsigma\|\nu_U) + \ln\frac{1}{\delta}}{N_u\, c}\right), \quad \varepsilon_{\mathbb{P}}(u) \le c'\left(\varepsilon_{\mathcal{D}_l}(u) + \frac{2\,\mathrm{KL}(\varsigma\|\nu_U) + \ln\frac{1}{\delta}}{N_l\, b}\right), \\
&\forall \varsigma \in \mathcal{H}, \quad \delta_{\mathbb{Q}}(u') \le c'\left(\delta_{\mathcal{D}_u}(u') + \frac{2\,\mathrm{KL}(\varsigma\|\nu_{U'}) + \ln\frac{1}{\delta}}{N_u\, c}\right), \quad \varepsilon_{\mathbb{P}}(u') \le c'\left(\varepsilon_{\mathcal{D}_l}(u') + \frac{2\,\mathrm{KL}(\varsigma\|\nu_{U'}) + \ln\frac{1}{\delta}}{N_l\, b}\right).
\end{aligned} \tag{15}$$

CDM decomposition. From the binary decomposition of $\mathrm{CDR}_{\mathbb{Q}}(u, u')$, we recall:

$$\mathrm{CDR}_{\mathbb{Q}}(u, u') \;\le\; \tfrac{1}{2}\big(\delta_{\mathbb{Q}}(u) - \delta_{\mathbb{Q}}(u')\big) + \big(\varepsilon_{\mathbb{Q}}(u) - \varepsilon_{\mathbb{Q}}(u')\big). \tag{16}$$

Using the β–divergence bound. Following [2], the β–divergence between $\mathbb{Q}$ and $\mathbb{P}$ is defined as

$$\beta_q(\mathbb{Q}\,\|\,\mathbb{P}) = \left(\mathbb{E}_{(x,y)\sim\mathbb{P}}\left[\left(\tfrac{\mathbb{Q}(x,y)}{\mathbb{P}(x,y)}\right)^q\right]\right)^{1/q}. \tag{17}$$

Applying Hölder's inequality $(\int |fg|\,d\mu) \le (\int |f|^q d\mu)^{1/q} (\int |g|^p d\mu)^{1/p}$ with $g = 1$, $q \to \infty$, $p \to 1$, and defining $r = \mathbb{E}_{(x,y)\sim\mathbb{Q}}[\mathbb{I}(y \in \mathcal{Y}_u^{\text{novel}})]$, we obtain

$$\begin{aligned}
\varepsilon_{\mathbb{Q}}(u) &\le \lim_{q\to\infty} \beta_q(\mathbb{Q}\,\|\,\mathbb{P})\, \varepsilon_{\mathbb{P}}(u) + r\, \varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u), \\
\varepsilon_{\mathbb{Q}}(u') &\le \lim_{q\to\infty} \beta_q(\mathbb{Q}\,\|\,\mathbb{P})\, \varepsilon_{\mathbb{P}}(u') + r\, \varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u').
\end{aligned} \tag{18}$$

Let $\pi_{\mathbb{Q}\setminus\mathbb{P}}(X, Y) := r\,(\varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u) - \varepsilon_{\mathbb{Q}\setminus\mathbb{P}}(u'))$ denote the novel categories residual term.

Combining results. Substituting (15)–(18) into (16), let $b' = \frac{b}{1-e^{-b}}\,\beta_\infty(T\,\|\,S)$, we have that with probability at least $1-\delta$, $\forall \varsigma \in \mathcal{H}$:

$$\begin{aligned}
\mathrm{CDR}_{\mathbb{Q}}(u, u') \le\; & \tfrac{c'}{2}\big(\delta_{\mathcal{D}_u}(u) - \delta_{\mathcal{D}_u}(u')\big) + b' \lim_{q\to\infty} \beta_q(\mathbb{Q}\|\mathbb{P})\big(\varepsilon_{\mathcal{D}_l}(u) - \varepsilon_{\mathcal{D}_l}(u')\big) + \pi_{\mathbb{Q}\setminus\mathbb{P}}(X, Y) \\
& + \Big(\tfrac{c'}{N_u c} + \tfrac{b'}{N_l b}\Big)\Big(2\,\mathrm{KL}(\varsigma\|\nu_U) + \ln\tfrac{2}{\delta}\Big) - \Big(\tfrac{c'}{N_u c} + \tfrac{b'}{N_l b}\Big)\Big(2\,\mathrm{KL}(\varsigma\|\nu_{U'}) + \ln\tfrac{2}{\delta}\Big).
\end{aligned} \tag{19}$$

Grouping the KL terms yields the symmetric regularizer:

$$\mathrm{Reg}_{\mathrm{KL}} = \Big(\tfrac{c'}{N_u c} + \tfrac{b'}{N_l b}\Big)\big(\mathrm{KL}(\varsigma\|\nu_U) + \mathrm{KL}(\nu_{U'}\|\varsigma)\big), \tag{20}$$

Therefore, the gap between the expected causal-dependency risk $\mathrm{CDR}_{\mathbb{Q}}(u, u')$ and its empirical counterpart $\mathrm{CDR}_{\mathcal{D}_u}(u, u')$ can be reduced by minimizing the regularization term: $\mathrm{KL}(\varsigma\|\nu_U) + \mathrm{KL}(\nu_{U'}\|\varsigma)$.

Assume a non-informative Gaussian prior $\nu := \mathcal{N}(\mu_\nu, \sigma_\nu^2)$. For the Gaussian posterior $\mathcal{N}(\mu_u, \sigma_u^2)$ produced by the encoder, the KL term satisfies

$$\begin{aligned}
\min_{\mu_u,\sigma_u}\left\{\lim_{\sigma_\nu\to\infty} \mathrm{KL}\big(\mathcal{N}(\mu_u, \sigma_u^2)\,\|\,\mathcal{N}(\mu_\nu, \sigma_\nu^2)\big)\right\} &= \min_{\mu_u,\sigma_u}\left\{\lim_{\sigma_\nu\to\infty}\left(\log\frac{\sigma_\nu}{\sigma_u} + \frac{\sigma_u^2 + (\mu_u - \mu_\nu)^2}{2\sigma_\nu^2} - \frac{1}{2}\right)\right\} \\
&= \min_{\sigma_u}\left\{\lim_{\sigma_\nu\to\infty} \log\frac{\sigma_\nu}{\sigma_u}\right\} = \min_{\sigma_u}(-\log\sigma_u).
\end{aligned} \tag{21}$$

Finally, by incorporating the deviation bound in Corollary 1, the expected causal-dependency risk $\mathrm{CDR}_{\mathbb{Q}}(u, u')$ can be further minimized through jointly tightening its empirical counterpart $\mathrm{CDR}_{\mathcal{D}_u}(u, u')$ and the KL-based representation variability penalty, which completes the proof.

1.4. Proof of Exogeneity

According to the definitions of KL divergence and conditional mutual information,

$$\mathbb{E}_{P(U,S)}\big[\mathrm{KL}\big(P(Y \mid U,S) \,\|\, P(Y \mid U)\big)\big] = I(Y;S \mid U). \tag{22}$$

Hence,

$$P(Y \mid U,S) = P(Y \mid U) \iff I(Y;S \mid U) = 0, \tag{23}$$

which is the exogeneity condition (domain S carries no residual information about Y given U).

By Bayes' rule,

$$P(Y \mid U,S) \propto P(U \mid Y,S)\,\pi_Y. \tag{24}$$

Therefore, a sufficient way to enforce (23) is to align the class-conditional distributions across domains:

$$P(U \mid Y{=}k, S{=}\mathbb{P}) = P(U \mid Y{=}k, S{=}\mathbb{Q}) \quad (\forall k). \tag{25}$$

Let $U \mid (Y{=}k, S{=}s)$ follow a location family with domain-invariant shape,

$$U \mid (Y{=}k, S{=}s) \sim \mathcal{F}(e_s^k, \Sigma_k), \qquad e_s^k := \mathbb{E}[U \mid Y{=}k, S{=}s], \tag{26}$$

so that aligning prototypes implies class-conditional alignment:

$$e_{\mathbb{P}}^k = e_{\mathbb{Q}}^k \ (\forall k) \implies P(U \mid Y{=}k, S{=}\mathbb{P}) = P(U \mid Y{=}k, S{=}\mathbb{Q}). \tag{27}$$

Finally, prototype equality can be promoted by maximizing cross-domain prototype mutual information (with bounded prototype marginals):

$$\max_G \sum_{k=1}^K I(e_{\mathbb{P}}^k;\, e_{\mathbb{Q}}^k) \iff \min_G \sum_{k=1}^K H(e_{\mathbb{P}}^k \mid e_{\mathbb{Q}}^k) \implies e_{\mathbb{P}}^k = e_{\mathbb{Q}}^k. \tag{28}$$

Combining (27) with (25)–(23) yields $I(Y;S \mid U) = 0$, i.e., exogeneity holds.

2. More Empirical Results

2.1. Robustness and Consistency Across Methods

To compare the stability of different clustering approaches under distribution shift, we report mean accuracy and standard deviation in Table 1 and Table 2. For SSB-C, the clean datasets (CUB, SCAR, FGVC) are used as the *known source domain*, while their corrupted versions serve as the *unknown target domain*. For DomainNet, the Real domain is taken as the known source domain, and each of the remaining domains (Painting, Sketch, Quickdraw, Clipart, Infograph) is treated as a separate unknown target domain.

Clustering accuracy is evaluated on both the source and target domains, and results are averaged across multiple runs. Overall, our method demonstrates strong consistency and robustness, maintaining stable performance across diverse domain shifts and corruption levels.

Table 1. Clustering performance of different methods on the SSB-C benchmark. Results are reported as mean ± std.

Methods	CUB-C						Scars-C						FGVC-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
RankStats+	19.3±2.3	22.0±2.7	15.4±1.8	13.6±2.4	23.9±2.1	4.5±0.6	14.8±2.3	20.8±2.2	7.8±2.0	11.5±2.6	22.6±2.3	1.0±0.2	14.4±2.3	16.4±2.2	14.5±1.9	8.3±1.1	15.6±2.4	5.0±0.7
UNO+	25.9±2.5	40.1±2.3	21.3±2.0	21.5±2.3	33.4±2.4	8.6±1.2	22.0±2.2	41.8±2.1	7.0±1.8	16.9±2.5	29.8±2.0	4.5±0.6	22.0±2.4	33.4±2.1	15.8±2.3	16.5±2.6	25.2±2.0	8.8±1.3
ORCA	18.2±1.6	22.8±2.4	14.5±1.9	21.5±2.0	23.1±2.3	18.9±2.2	19.1±2.2	28.7±2.0	11.2±1.6	15.0±2.4	22.4±2.1	8.3±1.0	17.6±1.9	19.3±2.2	16.1±2.1	13.9±1.4	17.3±1.8	10.1±2.2
GCD	26.6±1.1	27.5±1.7	25.7±1.3	25.1±1.0	28.7±1.4	22.0±1.2	22.1±1.6	35.2±1.3	20.5±1.1	21.6±1.5	29.2±1.1	10.5±1.4	25.2±1.5	28.7±1.2	23.0±1.4	21.0±1.3	23.1±1.0	17.3±1.6
SimGCD	31.9±2.3	33.9±1.9	29.0±2.0	28.8±2.4	31.6±2.1	25.0±2.0	26.7±2.2	39.6±2.1	25.6±1.9	22.1±2.5	30.5±2.1	14.1±2.4	26.1±2.3	28.9±2.0	25.1±1.9	22.3±2.1	23.2±2.3	21.4±2.0
SPTNet	33.0±1.7	34.5±1.1	31.2±1.9	30.1±2.0	33.1±1.4	26.1±2.2	28.0±1.5	40.2±2.0	27.9±1.6	24.2±2.3	32.1±1.8	16.3±1.3	28.7±2.0	30.2±1.7	27.9±2.2	24.8±1.5	25.7±1.1	23.9±2.4
RLCD	35.9±1.9	35.1±1.2	33.2±2.1	32.3±1.4	34.8±2.0	28.5±1.1	29.8±1.8	41.2±2.2	30.4±1.5	25.3±1.0	33.4±2.1	18.1±1.6	27.9±1.3	30.1±2.3	26.8±1.4	24.4±2.2	26.8±1.2	22.7±2.0
CDAD-Net	40.4±1.8	38.9±1.3	39.3±2.2	37.7±1.9	39.1±1.5	34.2±2.3	32.1±1.4	42.9±2.0	32.2±1.6	28.8±2.1	35.6±1.7	21.4±1.9	33.8±2.0	35.5±1.1	31.2±2.2	27.8±1.6	29.6±2.4	25.6±1.5
HiLo	56.8±1.6	54.0±1.3	60.3±1.5	52.0±1.6	53.6±1.5	50.5±1.6	39.5±1.2	44.8±1.7	37.0±1.3	35.6±1.8	42.9±1.2	28.4±1.5	44.2±1.7	50.6±1.6	47.4±1.3	31.2±1.4	29.0±1.5	33.4±1.6
FREE	60.4±1.9	58.5±2.1	63.2±2.0	55.7±2.2	57.1±1.8	53.7±1.4	43.6±1.9	48.1±2.1	40.8±1.7	38.9±1.5	46.1±1.8	32.6±1.2	48.5±2.0	54.9±1.9	51.2±1.6	35.0±1.3	32.4±1.7	38.9±1.5
CausalGCD	62.2±1.4	60.1±1.5	64.9±1.5	57.8±1.7	56.6±1.2	56.9±1.4	45.8±1.8	49.1±1.3	42.2±1.5	41.5±1.1	45.6±1.3	36.0±1.0	49.8±1.9	56.1±1.6	53.1±1.4	37.2±1.1	33.4±1.0	40.2±1.2

Table 2. Clustering performance of different methods on the DomainNet benchmark. Results are reported as mean ± std.

Methods	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
RankStats+	34.1±2.1	62.0±2.8	19.7±2.4	29.7±2.2	49.7±2.3	9.6±0.9	34.2±2.5	62.0±2.9	19.8±2.3	17.1±2.6	31.1±2.8	6.8±1.1	34.1±2.7	62.5±2.5	19.5±2.0	4.1±1.2	4.4±0.7	3.9±0.8	34.0±2.2	62.4±2.8	19.4±2.1	24.1±2.4	45.1±2.2	6.2±2.0	34.2±2.3	62.4±2.5	19.6±2.7	12.5±1.9	21.9±2.5	6.3±1.0
UNO+	44.2±2.4	72.2±2.8	29.7±2.0	30.1±2.3	45.1±2.5	17.2±1.1	43.7±2.2	72.5±2.8	28.9±2.5	12.5±1.9	17.0±2.4	9.2±1.3	31.1±2.5	60.0±2.7	16.1±2.2	6.3±1.4	5.8±0.6	6.8±0.6	44.5±2.1	66.1±2.9	33.3±2.3	21.9±2.4	35.6±2.7	10.1±2.0	42.8±2.3	69.4±2.5	29.0±2.6	10.9±2.0	15.2±2.5	8.0±1.3
ORCA	31.9±1.5	49.8±1.7	23.5±1.9	28.7±2.3	38.5±2.2	7.1±0.9	32.5±1.1	50.0±1.3	23.9±2.1	11.4±1.8	14.5±1.3	7.2±1.0	19.2±2.0	39.1±2.0	15.3±2.4	3.4±1.1	3.5±0.6	3.2±0.9	32.0±0.9	49.7±1.5	23.9±1.3	19.1±1.2	31.8±1.0	4.3±0.5	29.1±2.2	47.7±1.8	20.1±2.4	8.6±1.3	13.7±1.9	7.1±1.3
GCD	47.3±0.6	53.6±1.5	44.1±1.2	32.9±1.4	41.8±1.7	23.0±0.8	48.0±1.3	53.8±1.4	45.3±1.2	16.6±0.7	22.4±1.0	11.1±1.0	37.6±1.4	41.0±1.6	35.2±1.0	5.7±0.8	4.2±0.9	6.9±1.0	47.7±1.6	53.8±1.3	44.3±1.1	22.4±1.0	34.4±1.0	16.0±1.8	41.9±1.1	46.1±1.2	39.0±1.8	10.9±1.4	17.1±1.3	8.8±1.1
SimGCD	61.3±1.1	77.8±1.4	52.9±2.1	34.5±1.9	35.6±2.1	33.5±1.1	62.4±1.7	77.6±2.4	54.6±1.5	16.4±2.0	20.2±1.0	13.6±2.5	47.4±1.8	64.5±1.2	37.4±1.1	6.6±0.7	5.8±0.7	7.5±0.9	61.6±2.5	77.2±2.4	53.6±1.6	23.9±1.0	31.5±2.2	17.3±1.6	52.7±2.0	67.0±1.7	44.8±1.3	11.6±2.2	15.4±2.5	9.1±1.2
SPTNet	61.6±1.5	76.9±2.3	54.7±2.0	35.2±1.8	35.9±1.2	35.1±1.2	63.3±1.1	77.8±2.2	55.3±1.8	16.7±2.0	26.0±1.0	11.3±2.4	47.1±2.2	65.6±1.3	35.4±1.3	6.9±1.3	5.7±0.4	7.7±1.3	62.5±1.6	76.5±1.4	55.4±1.9	24.7±1.2	30.9±1.4	18.8±1.5	54.5±1.6	67.9±2.1	46.2±1.3	11.9±1.7	19.4±1.8	7.9±1.1
RLCD	62.1±1.9	78.3±1.2	53.8±1.1	36.9±2.3	35.7±2.4	36.2±2.1	62.8±1.4	77.4±1.1	55.7±2.0	17.0±1.6	20.4±1.2	15.2±1.7	49.1±1.0	67.8±2.3	38.0±1.4	7.0±0.9	5.8±0.6	7.8±1.2	62.3±1.8	77.1±1.3	54.7±2.4	24.5±2.1	38.0±2.3	13.9±2.3	57.2±1.8	68.3±2.3	48.1±1.1	12.0±1.3	15.9±1.1	9.8±0.9
CDAD-Net	63.6±1.5	77.8±1.4	56.3±2.2	38.4±1.5	38.4±1.4	37.5±1.8	61.9±1.2	76.3±2.1	52.1±1.1	17.3±2.4	20.9±2.1	15.9±1.3	48.5±1.0	66.5±2.1	36.7±2.0	6.4±1.0	5.6±0.7	7.3±1.1	61.3±1.5	77.0±1.3	53.1±1.8	25.2±1.6	31.9±1.4	19.0±2.2	56.5±2.0	68.0±1.4	47.1±1.3	11.8±1.2	15.6±1.7	9.4±0.8
HiLo	64.4±1.5	77.6±1.5	57.5±1.6	42.1±1.6	42.9±1.7	41.3±1.4	63.3±1.7	77.9±1.8	55.9±1.5	19.4±1.1	22.4±1.7	17.1±1.4	58.6±1.6	76.4±1.1	52.5±1.4	7.4±1.1	6.9±0.5	8.0±1.6	63.8±1.0	77.6±1.0	56.6±1.8	27.7±1.2	34.6±1.3	21.7±1.6	64.2±1.8	78.1±1.0	57.0±1.0	13.7±1.3	16.4±1.3	11.9±1.1
FREE	67.7±1.6	78.1±1.5	61.2±1.3	45.6±1.4	46.1±1.6	44.8±1.2	67.8±1.8	78.2±1.5	61.6±1.4	22.5±1.1	25.8±1.3	20.9±1.1	61.4±1.7	78.1±1.6	55.1±1.3	8.9±1.0	7.8±0.4	9.0±1.2	66.4±1.8	78.1±1.6	60.1±1.3	29.3±1.5	37.2±1.6	26.3±1.0	68.1±1.9	78.9±1.5	60.2±1.4	16.1±1.1	18.6±1.3	13.4±1.0
CausalGCD	69.9±1.0	77.9±1.2	64.1±1.4	48.0±1.1	47.3±1.1	46.4±1.5	69.3±1.0	78.7±1.3	63.5±1.4	24.3±1.2	25.1±1.0	23.2±1.3	62.9±1.1	77.7±1.4	57.6±1.2	9.3±1.1	7.6±0.9	9.9±1.0	68.0±1.5	77.9±1.4	62.6±1.4	31.2±1.2	37.0±1.3	28.4±1.4	70.0±1.0	78.2±1.5	62.1±1.3	17.5±1.0	19.9±1.2	15.0±1.1

Table 3. Detailed clustering performance on CUB-C. We report the clustering accuracy on each corrupted domain.

Methods	Gaussian Noise			Shot Noise			Impulse Noise			Zoom Blur			Snow			Frost			Fog			Speckle			Spatter		
	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
RankStats+	13.6	20.9	4.5	12.7	28.4	5.1	12.3	27.4	5.4	15.2	33.7	4.9	16.0	34.7	5.6	17.5	38.4	4.8	18.7	40.7	4.9	16.8	36.5	5.3	22.3	48.1	4.7
UNO+	18.5	32.4	7.6	17.2	30.5	7.2	17.1	31.1	6.2	20.4	35.7	8.4	20.7	35.6	7.0	20.7	35.2	7.4	30.2	52.2	10.5	22.9	42.0	8.4	29.7	52.7	11.2
ORCA	21.5	23.1	19.9	21.2	23.7	18.8	21.1	23.1	19.2	20.4	22.0	18.9	20.1	22.1	18.3	22.0	25.5	18.5	19.2	20.4	18.0	22.4	20.8	19.1	24.8	31.3	18.3
GCD	23.4	22.7	20.0	22.7	20.4	31.0	21.9	20.3	19.6	25.1	25.3	21.0	23.6	22.9	20.2	23.9	23.1	20.8	29.7	31.1	24.4	27.6	26.7	24.6	35.2	36.2	30.3
SimGCD	23.8	26.6	22.0	21.6	23.8	20.4	20.4	22.5	19.4	30.5	35.8	26.2	29.0	34.3	24.9	29.1	32.6	26.7	33.0	36.9	30.1	27.3	29.6	26.1	41.5	47.0	37.0
SPTNet	25.5	28.3	23.8	23.2	25.7	22.2	22.4	24.3	21.2	32.4	37.6	28.0	30.9	36.4	26.6	30.7	34.6	28.4	34.8	38.7	31.9	28.9	31.4	27.7	43.2	48.9	38.6
RLCD	26.5	29.4	24.9	24.0	26.8	23.2	23.4	25.5	22.6	33.3	39.0	29.4	32.0	37.1	28.1	31.2	35.3	29.1	35.5	39.7	33.0	29.5	32.2	28.6	44.4	50.1	39.9
CDAD-Net	31.9	35.2	29.6	30.5	33.1	28.4	28.2	30.4	26.8	38.3	44.0	33.5	37.6	42.5	34.1	36.9	41.0	34.4	39.7	43.9	37.4	34.5	37.2	33.0	49.7	**55.6**	44.6
HiLo	41.8	39.8	43.9	41.0	38.7	43.3	42.2	39.8	44.5	47.9	43.9	51.8	49.3	45.8	52.8	48.5	45.5	51.4	50.6	46.8	54.3	47.9	45.4	50.2	50.9	47.2	54.7
FREE	45.7	**44.9**	48.2	46.5	43.2	47.2	47.1	43.5	47.8	50.3	48.1	54.2	53.4	**49.7**	55.9	51.2	48.7	54.4	53.4	49.8	57.1	50.3	48.7	53.1	**53.8**	51.2	56.9
CausalGCD	**47.1**	43.5	**51.4**	**48.0**	**44.3**	**49.0**	**49.4**	**45.0**	**48.5**	**53.1**	**49.4**	**56.3**	**55.0**	49.1	**57.2**	**54.0**	**49.0**	**56.1**	**55.2**	**49.9**	**59.7**	**51.9**	**49.8**	**54.5**	53.0	50.1	**57.0**

Table 4. Detailed clustering performance on Scars-C. We report the clustering accuracy on each corrupted domain.

Methods	Gaussian Noise			Shot Noise			Impulse Noise			Zoom Blur			Snow			Frost			Fog			Speckle			Spatter		
	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
RankStats+	8.5	16.6	1.6	8.9	16.7	1.7	7.2	13.8	1.5	11.7	22.9	0.5	8.9	17.0	1.3	11.4	21.9	0.7	16.8	32.6	1.2	12.7	24.1	1.6	17.3	34.1	1.6
UNO+	13.9	24.8	6.5	14.0	25.0	6.9	11.2	20.4	6.4	17.1	33.2	2.6	13.3	24.0	4.5	17.3	29.9	6.3	22.4	39.8	3.8	18.6	33.1	7.1	21.8	38.4	4.0
ORCA	12.0	31.4	9.3	13.2	31.8	9.7	11.8	29.2	9.2	14.5	38.2	7.9	12.5	32.6	9.5	15.7	36.4	10.0	20.3	47.7	5.8	17.0	39.4	10.5	21.6	48.8	10.6
GCD	17.6	24.2	10.8	17.1	24.6	11.2	14.4	20.9	11.0	23.2	31.8	8.0	18.5	25.5	8.4	23.2	31.1	10.2	27.1	40.8	5.7	22.6	30.1	12.4	31.0	43.1	7.1
SimGCD	18.1	23.5	15.7	18.3	23.5	15.5	15.2	19.0	15.4	24.4	32.7	13.1	19.7	26.4	12.9	23.9	31.9	13.3	28.0	38.6	12.7	23.4	30.6	16.4	32.4	45.4	13.1
SPTNet	19.8	25.3	17.5	20.3	25.1	17.1	17.1	20.8	17.2	26.0	34.8	14.9	21.5	28.3	14.6	25.7	33.9	14.9	30.0	40.2	14.8	25.3	32.7	18.3	34.2	47.6	15.1
RLCD	20.9	26.3	18.7	21.4	26.8	18.5	17.8	21.7	18.4	27.0	35.9	16.2	22.9	29.5	15.8	27.3	35.1	15.9	31.2	42.0	15.7	26.4	33.6	19.2	35.6	49.0	16.3
CDAD-Net	22.6	28.3	20.3	23.2	29.0	20.0	19.3	24.0	20.1	28.7	38.0	17.7	24.4	31.4	17.2	29.1	37.4	17.4	33.5	45.3	17.1	28.1	35.5	20.6	37.4	49.0	18.1
HiLo	31.0	38.0	24.3	31.5	38.3	24.9	30.2	36.6	23.9	38.4	45.1	31.9	36.8	44.9	29.0	36.5	43.8	29.5	40.7	49.5	32.2	37.1	37.1	29.6	37.9	45.4	30.6
FREE	35.6	**42.8**	27.9	35.4	42.0	27.8	34.9	41.9	27.6	42.2	50.2	35.8	41.1	49.6	34.5	41.2	47.2	33.5	45.4	53.1	35.9	41.2	41.3	33.7	42.1	**49.9**	34.8
CausalGCD	**37.1**	42.5	**29.5**	**38.0**	**42.3**	**29.9**	**36.3**	**43.0**	**28.4**	**45.0**	**51.2**	**37.0**	**43.0**	**50.3**	**36.8**	**43.3**	**48.2**	**36.6**	**47.1**	**55.0**	**36.4**	**42.5**	**43.8**	**36.0**	**44.2**	48.0	**36.9**

Table 5. Detailed clustering performance on FGVC-C. We report the clustering accuracy on each corrupted domain.

Methods	Gaussian Noise			Shot Noise			Impulse Noise			Zoom Blur			Snow			Frost			Fog			Speckle			Spatter		
	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
RankStats+	7.3	13.6	5.0	6.3	10.7	5.8	6.0	10.7	5.3	10.1	19.9	4.3	6.2	12.5	3.8	8.9	17.7	4.1	12.5	24.4	4.5	7.6	14.0	5.2	10.5	20.1	4.9
UNO+	15.5	25.2	5.8	13.5	22.1	4.9	13.2	20.1	6.2	20.1	28.2	12.0	15.6	21.3	9.9	17.6	25.2	9.9	19.3	26.9	11.7	16.5	27.2	5.6	20.9	29.6	12.3
ORCA	11.9	17.3	11.1	11.1	15.6	11.3	10.9	15.9	10.3	15.2	24.3	8.5	11.3	15.4	9.1	12.6	22.1	9.3	16.7	28.9	9.1	12.2	18.8	10.4	15.0	25.1	9.3
GCD	16.0	20.1	14.3	13.8	19.1	11.5	12.3	16.0	13.4	27.7	25.4	24.1	19.1	17.7	15.2	23.9	24.0	18.2	31.8	30.1	24.7	16.1	27.0	14.9	28.7	30.7	25.9
SimGCD	16.3	16.2	18.4	14.2	14.5	16.0	13.7	13.0	16.5	28.9	31.4	28.4	20.0	22.4	19.5	24.5	29.2	21.9	31.9	**37.8**	28.0	16.8	18.0	17.7	29.8	32.9	28.6
SPTNet	17.6	17.5	19.9	15.6	15.7	17.7	14.9	14.5	18.2	30.4	33.2	30.0	21.8	24.1	21.3	25.7	30.5	23.7	33.2	36.1	29.3	18.3	19.4	19.0	31.0	34.5	30.3
RLCD	18.7	18.6	20.9	16.8	17.2	18.9	15.9	15.5	19.4	31.4	34.3	31.1	22.4	25.0	22.3	27.1	31.5	25.0	34.3	36.3	30.5	19.6	20.8	20.2	32.1	35.1	31.4
CDAD-Net	21.3	21.5	23.8	19.6	20.0	21.4	18.6	18.5	22.0	34.5	**37.2**	34.3	25.5	27.6	26.0	30.1	32.7	29.4	31.7	36.0	27.3	22.4	24.0	23.5	31.4	34.8	31.3
HiLo	28.6	25.2	32.0	26.8	24.4	29.2	27.9	24.5	31.4	36.8	34.2	39.4	27.8	27.9	27.8	33.4	30.4	36.4	35.8	34.1	37.5	30.4	30.4	32.7	33.4	32.4	34.4
FREE	33.7	28.4	34.5	31.2	27.2	32.5	32.0	27.9	34.6	39.9	36.3	41.3	30.4	30.5	30.9	36.1	33.1	38.7	37.9	37.6	40.9	34.5	34.2	36.8	37.3	35.6	37.9
CausalGCD	**35.1**	**29.2**	**36.1**	**33.0**	**29.0**	**34.1**	**33.3**	**28.1**	**35.7**	**42.1**	36.4	**43.0**	**32.8**	**33.0**	**32.9**	**38.4**	**35.3**	**39.2**	**39.0**	37.3	**42.6**	**36.2**	**35.8**	**38.2**	**39.7**	**37.9**	**39.4**

2.2. Comprehensive Results on the SSB-C Benchmark

To obtain a more complete understanding of model behavior under corruption-induced distribution shifts, we additionally evaluate our method across all corrupted settings included in the SSB-C benchmark. The corresponding results are listed in Table 3, Table 4, and Table 5. Across nearly all corruption types and severity levels, our approach achieves consistently strong accuracy and frequently surpasses alternative methods. These findings demonstrate that the proposed framework remains reliable even when the input distribution undergoes substantial degradation.

2.3. Extended Cross-Domain Analysis on DomainNet

To further examine the generalization capability of our framework, we conduct extensive experiments on the DomainNet benchmark under multiple cross-domain configurations. In each evaluation, the Real domain is used as the *known source domain*, while each of the other five domains is treated in turn as an *unknown target domain* during training. After training on a particular source–target combination, the resulting model is then tested on the remaining four domains that were not involved in training. Detailed quantitative results are included in Table 6 through Table 13.

Across this broad collection of transfer scenarios, our approach maintains a clear performance advantage over competitive baselines. These experiments confirm that the proposed method effectively captures semantic information that is invariant to visual style, enabling robust transfer to previously unseen domains.

Table 6. Clustering results when using Real as the known source domain and Sketch as the unknown target domain. Performance is evaluated not only on these two domains but also on the remaining unseen domains.

Methods	Real			Sketch			others		
	All	Old	New	All	Old	New	All	Old	New
RankStats+	34.2	62.0	19.8	17.1	**31.1**	6.8	17.3	**30.0**	6.1
UNO+	43.7	72.5	28.9	12.5	17.0	9.2	17.4	26.4	9.5
ORCA	32.5	50.0	23.9	11.4	14.5	7.2	13.3	23.1	9.1
GCD	48.0	53.8	45.3	16.6	22.4	11.1	20.7	25.8	15.8
SimGCD	62.4	77.6	54.6	16.4	20.2	13.6	20.4	25.4	16.1
SPTNet	62.7	77.8	54.9	16.9	20.8	13.9	21.2	25.9	16.9
RLCD	63.0	77.6	55.6	17.4	20.3	15.6	21.4	26.4	16.7
CDAD-Net	62.5	77.4	55.1	16.6	20.2	14.1	22.1	27.1	16.9
HiLo	63.3	77.9	55.9	19.4	22.4	17.1	21.3	25.8	17.4
FREE	65.8	**78.4**	57.2	22.5	25.1	19.8	24.0	27.0	19.8
CausalGCD	**67.4**	78.1	**59.1**	**24.7**	27.1	**22.0**	**26.9**	26.4	**23.2**

Table 7. Clustering performance on the remaining domains when the Real domain is used as the known source domain and Sketch is selected as the unknown target domain.

Methods	Painting			Quickdraw			Clipart			Infograph		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New
RankStats+	29.7	**49.2**	10.2	2.3	2.1	2.4	24.6	**45.9**	5.9	12.5	22.6	5.9
UNO+	30.8	44.0	17.6	2.4	2.4	2.3	23.1	38.0	10.1	13.2	21.2	7.9
ORCA	23.1	39.1	17.2	2.5	**3.0**	2.0	19.7	33.1	10.0	8.9	18.1	7.0
GCD	32.6	40.1	31.5	1.6	1.9	1.5	24.1	31.1	14.9	14.1	16.2	10.2
SimGCD	38.7	44.7	32.7	1.9	1.2	2.5	25.2	35.3	16.3	15.8	20.3	12.8
SPTNet	37.9	44.2	32.2	2.0	1.5	2.1	25.7	35.8	16.9	16.2	20.9	13.1
RLCD	39.0	44.4	33.1	1.6	1.1	2.2	26.2	35.8	17.2	16.2	20.8	13.2
CDAD-Net	38.0	44.1	32.8	2.0	1.3	2.3	25.8	36.1	16.7	17.1	21.3	14.8
HiLo	39.8	44.7	34.9	1.9	2.0	1.7	27.2	35.9	19.6	16.2	20.5	13.4
FREE	41.9	45.8	37.2	2.5	2.9	2.3	29.8	36.5	21.8	18.1	22.8	15.5
CausalGCD	**44.5**	47.1	**39.4**	**3.8**	2.5	**2.8**	**32.8**	35.6	**25.6**	**19.3**	**24.0**	**17.2**

2.4. Extended Hyperparameter Sensitivity Analysis

To further understand the influence of key hyperparameters in our framework, we conduct an extended sensitivity study focusing on two critical components: (1) the number of top eigenvectors m used in the CGMC module, and (2) the intervention

Table 8. Clustering performance when Real is used as the known source domain and Quickdraw as the unknown target domain, along with results on the remaining unseen domains.

Methods	Real			Quickdraw			others		
	All	Old	New	All	Old	New	All	Old	New
RankStats+	34.1	62.5	19.5	4.1	4.4	3.9	21.0	**37.4**	7.2
UNO+	31.1	60.0	16.1	6.3	5.8	6.8	18.6	32.2	7.0
ORCA	19.2	39.1	15.3	3.4	3.5	3.2	15.6	28.4	8.1
GCD	37.6	41.0	35.2	5.7	4.2	6.9	21.9	34.3	12.2
SimGCD	47.4	64.5	37.4	6.6	5.8	7.5	22.9	33.8	13.8
SPTNet	47.8	64.9	37.6	6.8	5.9	7.8	23.1	33.6	14.5
RLCD	49.2	67.1	38.2	6.9	5.6	8.5	25.1	34.3	15.1
CDAD-Net	51.3	66.7	49.4	7.1	6.2	7.9	25.3	35.8	15.9
HiLo	58.6	76.4	52.5	7.4	6.9	8.0	25.9	32.5	20.4
FREE	62.3	**78.9**	56.3	8.1	8.1	8.9	27.6	33.9	22.1
CausalGCD	**64.1**	78.5	**58.1**	**9.0**	**8.7**	**9.7**	**29.1**	34.0	**23.4**

Table 9. Detailed clustering performance on the remaining domains when Real is used as the known source domain and Quickdraw as the unknown target domain.

Methods	Painting			Sketch			Clipart			Infograph		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New
RankStats+	29.6	**49.0**	10.2	17.1	**32.1**	6.1	24.8	**45.4**	6.7	12.6	**23.1**	5.7
UNO+	26.8	43.7	9.9	14.7	25.6	6.6	20.7	38.4	5.1	12.2	21.0	6.4
ORCA	22.2	40.9	10.1	11.9	22.4	7.1	17.5	35.6	5.7	10.3	18.7	6.6
GCD	32.9	45.7	21.4	18.5	30.5	10.8	23.5	39.0	10.7	13.8	22.1	7.6
SimGCD	33.8	45.1	22.5	19.4	30.1	11.5	24.0	38.5	11.4	14.5	21.6	9.8
SPTNet	34.5	46.1	24.2	21.3	32.1	11.7	26.2	39.5	11.8	15.6	23.1	9.9
RLCD	36.2	46.0	25.1	21.8	31.8	12.5	26.0	39.1	11.7	16.5	22.4	10.9
CDAD-Net	36.8	46.9	24.3	21.9	30.8	14.6	26.8	38.7	14.2	15.7	22.3	11.8
HiLo	38.6	45.1	32.2	22.9	28.8	18.5	26.0	36.4	16.9	16.2	19.8	13.9
FREE	42.1	46.5	36.5	25.1	31.9	20.1	29.2	38.9	20.7	19.1	21.5	15.7
CausalGCD	**44.4**	46.7	**38.4**	**27.6**	31.8	**22.4**	**31.6**	38.9	**22.4**	**21.3**	22.1	**17.5**

strength parameter κ that controls the magnitude of feature intervention. As shown in Fig. 1, using approximately $m \approx 5$ top eigenvectors is sufficient to capture the intrinsic geometric structure of each category. This observation indicates that the dominant semantic manifold can be effectively represented by the subspace spanned by the first few largest eigenvalues and their corresponding eigenvectors. We additionally examine the parameter κ, which governs the degree of intervention applied to the latent representation. A larger κ introduces stronger perturbations that help disentangle causal semantic factors from domain-specific noise. When κ is too small, the intervention becomes insufficient to induce meaningful semantic shifts, potentially leading to unstable behavior and weaker causal separation. Conversely, appropriately increasing κ enhances the model's ability to isolate semantically relevant variations and prevents learning degenerate representations. Overall, the results show that our model is robust to a wide range of hyperparameter configurations, and the framework maintains stable performance as long as m and κ fall within reasonable ranges.

2.5. Pseudocode

The pseudocode of CausalGCD, summarizing its key causal components and overall training procedure, is provided in Algorithm 1.

Algorithm 1: Training Procedure of the CausalGCD Framework

Input : labeled source set $\mathcal{D}_l = \{(x_i^l, y_i^l)\}_{i=1}^{N_l}$; unlabeled target set $\mathcal{D}_u = \{x_i^u\}_{i=1}^{N_u}$;
inference encoder G, intervention encoder G', classifier g; number of epochs E; batch size B;
learning rate η; trade-off hyperparameters $\lambda_1, \ldots, \lambda_5$

Output: optimized model $f^* = \{G^*, G'^*, g^*\}$

/* --- Source-domain causal dependency, KL index, and separation --- */

Function CDR_Source$(\mathcal{B}_l, G, G', g)$**:**

  Sample mini-batch $\mathcal{B}_l = \{(x_i^l, y_i^l)\}_{i=1}^{B_l}$ from $\mathcal{D}_l$;

  $u_s \leftarrow G(\mathcal{B}_l)$, $u'_s \leftarrow G'(\mathcal{B}_l)$;

  $\hat{y}_s \leftarrow g(u_s)$, $\hat{y}'_s \leftarrow g(u'_s)$;

  Compute source CDR loss $\mathcal{L}_{\mathbb{P}}$ using Eq. 7;

  Compute variability index $\mathcal{L}_{\mathrm{KL}}^{\mathbb{P}}$ using Eq. 9;

  Compute separation loss $\mathcal{L}_{\mathrm{sep}}^{\mathbb{P}} \leftarrow \kappa - \|u_s - u'_s\|_2^2$;

  return $\mathcal{L}_{\mathbb{P}}, \mathcal{L}_{\mathrm{KL}}^{\mathbb{P}}, \mathcal{L}_{\mathrm{sep}}^{\mathbb{P}}$

/* --- Target-domain CDR with pseudo labels --- */

Function CDR_Target$(\mathcal{B}_u, G, G', g)$**:**

  Sample mini-batch $\mathcal{B}_u = \{x_i^u\}_{i=1}^{B_u}$ from $\mathcal{D}_u$;

  $u_t \leftarrow G(\mathcal{B}_u)$, $u'_t \leftarrow G'(\mathcal{B}_u)$;

  $\hat{y}_t \leftarrow g(u_t)$, $\hat{y}'_t \leftarrow g(u'_t)$;

  Obtain pseudo labels $\tilde{y}_t$ (e.g., $\tilde{y}_t = \arg\max \hat{y}_t$ or confidence-based);

  Compute target CDR loss $\mathcal{L}_{\mathbb{Q}}$ using Eq. 7;

  Compute variability index $\mathcal{L}_{\mathrm{KL}}^{\mathbb{Q}}$ using Eq. 9;

  Compute separation loss $\mathcal{L}_{\mathrm{sep}}^{\mathbb{Q}} \leftarrow \kappa - \|u_t - u'_t\|_2^2$;

  return $\mathcal{L}_{\mathbb{Q}}, \tilde{y}_t, u_t, \mathcal{L}_{\mathrm{KL}}^{\mathbb{Q}}, \mathcal{L}_{\mathrm{sep}}^{\mathbb{Q}}$

/* --- Prototype update and exogeneity loss --- */

Function Update_Prototypes$(\mathcal{B}_l, \tilde{y}_t, u_s, u_t)$**:**

  Update source prototypes $e_{\mathbb{P}}^k$ by averaging $G(x_i^l)$ over $y_i^l = k$ in $\mathcal{B}_l$;

  Update target prototypes $e_{\mathbb{Q}}^k$ by weighted averaging u_t over pseudo labels $\tilde{y}_t = k$;

  Compute exogeneity loss $\mathcal{L}_{\mathrm{e}}$;

  return $\mathcal{L}_{\mathrm{e}}, \{e_{\mathbb{P}}^k\}, \{e_{\mathbb{Q}}^k\}$

/* --- Causal Geometric Manifold Constraint (CGMC) --- */

Function Compute_CGMC$(\mathcal{B}_u, G, \{e_{\mathbb{P}}^k\}, \{e_{\mathbb{Q}}^k\})$**:**

  Generate style-transferred samples $\tilde{x}_i^u = \mathcal{F}^{-1}(\mathcal{A}_i^l \cdot e^{j\mathcal{P}_i^u})$ via Fourier-based style transfer;

  $Z_{\mathbb{P}} \leftarrow G(\tilde{x}_i^u)$, $Z_{\mathbb{Q}} \leftarrow G(x_i^u)$;

  For each known class k and unknown class r:

    Compute covariance matrices and principal directions to obtain manifolds $\mathrm{CG}_{\mathbb{P}}^k, \mathrm{CG}_{\mathbb{P}}^r, \mathrm{CG}_{\mathbb{Q}}^k, \mathrm{CG}_{\mathbb{Q}}^r$;

    Compute geometric correlations $\mathrm{Sim}_{\mathbb{P}}(k, r)$ and $\mathrm{Sim}_{\mathbb{Q}}(k, r)$;

  Compute $\mathcal{L}_{\mathrm{cgmc}}$

  return $\mathcal{L}_{\mathrm{cgmc}}$

/* --- Main training procedure --- */

Initialize parameters of G, G', g;

Compute initial source prototypes $\{e_{\mathbb{P}}^k\}$ on $\mathcal{D}_l$;

for *epoch* $= 1$ **to** E **do**

  for $b = 1$ **to** $\max\{|\mathcal{D}_l|, |\mathcal{D}_u|\}/B$ **do**

    Sample $\mathcal{B}_l$ and $\mathcal{B}_u$ from $\mathcal{D}_l$ and $\mathcal{D}_u$;

    $\mathcal{L}_{\mathbb{P}}, \mathcal{L}_{\mathrm{KL}}^{\mathbb{P}}, \mathcal{L}_{\mathrm{sep}}^{\mathbb{P}} \leftarrow$ CDR_Source$(\mathcal{B}_l, G, G', g)$

    $\mathcal{L}_{\mathbb{Q}}, \tilde{y}_t, u_t, \mathcal{L}_{\mathrm{KL}}^{\mathbb{Q}}, \mathcal{L}_{\mathrm{sep}}^{\mathbb{Q}} \leftarrow$ CDR_Target$(\mathcal{B}_u, G, G', g)$

    $\mathcal{L}_{\mathrm{e}}, \{e_{\mathbb{P}}^k\}, \{e_{\mathbb{Q}}^k\} \leftarrow$ Update_Prototypes$(\mathcal{B}_l, \tilde{y}_t, G(\mathcal{B}_l), u_t)$

    $\mathcal{L}_{\mathrm{cgmc}} \leftarrow$ Compute_CGMC$(\mathcal{B}_u, G, \{e_{\mathbb{P}}^k\}, \{e_{\mathbb{Q}}^k\})$

    Compute total loss:

    $\mathcal{L} = \mathcal{L}_{\mathbb{P}} + \lambda_1 \mathcal{L}_{\mathbb{Q}} + \lambda_2 \mathcal{L}_{\mathrm{KL}} + \lambda_3 \mathcal{L}_{\mathrm{e}} + \lambda_4 \mathcal{L}_{\mathrm{cgmc}} + \lambda_5 \mathcal{L}_{\mathrm{sep}}$;

    Backward pass: compute $\nabla_\theta \mathcal{L}$ with $\theta = \{G, G', g\}$;

    Update parameters: $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$;

  end for

end for

return *optimized model* $f^* = \{G^*, G'^*, g^*\}$

Table 10. Clustering results when Real is taken as the known source domain and Clipart as the unknown target domain, along with performance on the remaining unseen domains.

Methods	Real			Clipart			others		
	All	Old	New	All	Old	New	All	Old	New
RankStats+	34.0	62.4	19.4	24.1	**45.1**	6.2	15.8	**27.0**	6.4
UNO+	44.5	66.1	33.3	21.9	35.6	10.1	16.2	23.2	10.5
ORCA	32.0	49.7	23.9	19.1	31.8	4.3	13.7	19.9	8.6
GCD	47.7	53.8	44.3	22.4	34.4	16.0	18.0	24.1	12.1
SimGCD	61.6	77.2	53.6	23.9	31.5	17.3	19.2	23.6	15.6
SPTNet	63.0	77.7	53.9	24.4	31.8	17.9	21.2	24.5	16.7
RLCD	63.1	77.8	54.1	24.9	32.3	18.5	22.0	25.1	17.0
CDAD-Net	62.9	77.6	53.8	25.8	33.0	18.1	22.2	25.7	16.1
HiLo	63.8	77.6	56.6	27.7	34.6	21.7	19.8	23.6	16.8
FREE	66.1	**78.3**	60.1	29.4	37.1	24.9	23.4	23.9	20.1
CausalGCD	**68.5**	78.0	**63.1**	**32.2**	39.4	**26.1**	**25.8**	24.0	**22.6**

Table 11. Clustering results on additional domains when Real is used as the known source domain and Clipart as the unknown target domain.

Methods	Painting			Quickdraw			Sketch			Infograph		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New
RankStats+	30.0	**50.3**	9.7	2.6	2.3	2.9	17.4	**31.9**	6.8	13.1	**23.6**	6.2
UNO+	31.5	43.3	19.6	2.8	2.1	**3.6**	17.3	26.8	10.2	13.3	20.6	8.5
ORCA	29.3	36.9	9.2	1.3	1.5	1.2	13.7	21.9	8.3	10.3	19.4	6.3
GCD	33.4	40.4	22.2	**3.6**	**5.7**	2.2	19.5	27.7	12.7	15.5	22.7	11.1
SimGCD	39.0	45.9	32.1	0.8	0.5	1.1	21.1	27.3	16.5	15.9	20.8	12.7
SPTNet	40.2	46.4	33.1	0.6	0.4	1.0	22.3	27.9	17.1	16.1	20.1	13.5
RLCD	41.7	47.4	34.7	1.2	0.9	1.3	23.5	28.8	18.5	16.2	20.4	13.1
CDAD-Net	41.0	46.8	33.8	1.0	0.8	1.1	22.6	27.0	17.2	15.8	20.1	11.2
HiLo	40.7	46.3	35.1	1.3	0.4	2.3	21.2	26.9	17.0	15.9	20.6	12.8
FREE	42.3	47.1	37.2	2.1	0.9	3.2	23.1	27.5	18.9	16.5	20.9	13.0
CausalGCD	**44.3**	46.5	**39.5**	2.8	1.1	3.5	**25.6**	29.5	**19.5**	**18.7**	22.5	**14.3**

2.6. Impact of Stronger Backbone Architectures

We also investigate how replacing the feature extractor affects the overall performance of different methods. To this end, we substitute the original backbone with a pretrained CLIP encoder [4] and replicate all experiments using this more expressive and powerful visual representation.

Table 14 shows that adopting a stronger backbone leads to noticeable performance improvements across all compared models. More importantly, with the enhanced representation capacity of CLIP, our CausalGCD framework continues to outperform the best-performing baseline (FREE), further emphasizing the robustness and scalability of our approach. These results suggest that CausalGCD can readily benefit from advances in visual backbones and remains competitive even in high-capacity settings.

2.7. Unknown category number

In the previous experiments, we adopted the common assumption that the number of unknown categories is provided beforehand. This assumption, however, rarely holds in realistic open-world scenarios. To evaluate the practicality of our framework,

Table 12. Clustering performance when Real is used as the known source domain and Infograph as the unknown target domain, together with results on the remaining unseen domains.

Methods	Real			Infograph			others		
	All	Old	New	All	Old	New	All	Old	New
RankStats+	34.2	62.4	19.6	12.5	**21.9**	6.3	18.5	**32.1**	6.4
UNO+	42.8	69.4	29.0	10.9	15.2	8.0	18.2	28.0	9.6
ORCA	29.1	47.7	20.1	8.6	13.7	7.1	13.8	24.8	5.4
GCD	41.9	46.1	39.0	10.9	17.1	8.8	19.0	29.1	11.1
SimGCD	52.7	67.0	44.8	11.6	15.4	9.1	20.8	28.4	14.2
SPTNet	53.4	67.9	45.1	12.1	16.2	8.9	20.9	28.6	14.4
RLCD	53.9	68.3	45.8	12.5	16.7	9.1	21.5	29.4	14.6
CDAD-Net	53.5	65.6	47.2	13.6	17.2	9.8	22.0	29.1	16.2
HiLo	64.2	78.1	57.0	13.7	16.4	11.9	23.0	28.5	18.3
FREE	66.5	80.4	60.3	15.9	17.2	13.8	24.1	29.4	19.5
CausalGCD	**68.3**	**81.7**	**62.4**	**17.7**	16.9	**16.3**	**26.5**	28.8	**22.5**

Table 13. Clustering results on the remaining domains when Real is used as the known source domain and Infograph serves as the unknown target domain.

Methods	Painting			Quickdraw			Sketch			Clipart		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New
RankStats+	29.6	**49.2**	10.0	2.5	1.6	3.4	17.4	**32.2**	6.5	24.4	**45.5**	5.8
UNO+	30.8	44.8	16.8	2.7	2.3	3.1	17.0	27.0	9.7	22.3	37.8	8.7
ORCA	20.0	40.2	8.1	1.6	1.8	1.2	13.2	21.1	8.0	20.5	36.0	4.1
GCD	30.8	45.1	18.4	**3.6**	**4.7**	2.5	18.8	26.4	11.2	22.9	40.0	12.3
SimGCD	35.9	45.6	26.3	2.1	1.7	2.5	20.8	29.3	14.5	24.5	36.9	13.6
SPTNet	36.3	46.1	26.4	2.3	1.8	2.7	21.2	29.7	14.8	24.9	37.2	14.1
RLCD	37.2	46.8	26.9	2.5	1.9	2.9	22.3	30.5	15.3	25.1	37.3	14.7
CDAD-Net	39.3	46.5	29.6	2.5	1.7	**3.4**	21.8	30.5	16.1	25.6	36.5	16.4
HiLo	40.1	46.1	35.8	2.0	2.2	1.5	22.6	29.4	17.6	26.6	36.3	18.1
FREE	44.5	47.2	39.7	2.8	2.9	2.3	24.1	30.2	19.3	28.7	37.1	20.2
CausalGCD	**47.1**	48.5	**41.2**	3.1	3.5	3.0	**26.2**	**32.0**	**21.2**	**30.4**	**38.3**	**22.3**

Table 14. Clustering performance on DomainNet with Real as the known source domain and Painting as the unknown target domain.

Methods	Backbone	Real			Painting		
		All	Old	New	All	Old	New
FREE	DINO	67.7	**78.1**	61.2	45.6	46.1	44.8
CausalGCD	DINO	**69.9**	77.9	**64.1**	**48.0**	**47.3**	**46.4**
FREE	CLIP	78.2	78.3	69.5	51.0	54.2	49.3
CausalGCD	CLIP	**80.3**	**78.4**	**72.0**	**53.1**	**55.7**	**52.8**

we further consider a more challenging setting where the number of novel classes is *not* known and must be inferred automatically. We employ a state-of-the-art offline class-number estimation method [5] to predict the number of novel categories before the discovery stage, and then rerun the complete pipeline using the estimated value.

As reported in Table 15, our CausalGCD model maintains a clear performance advantage even under this more realistic

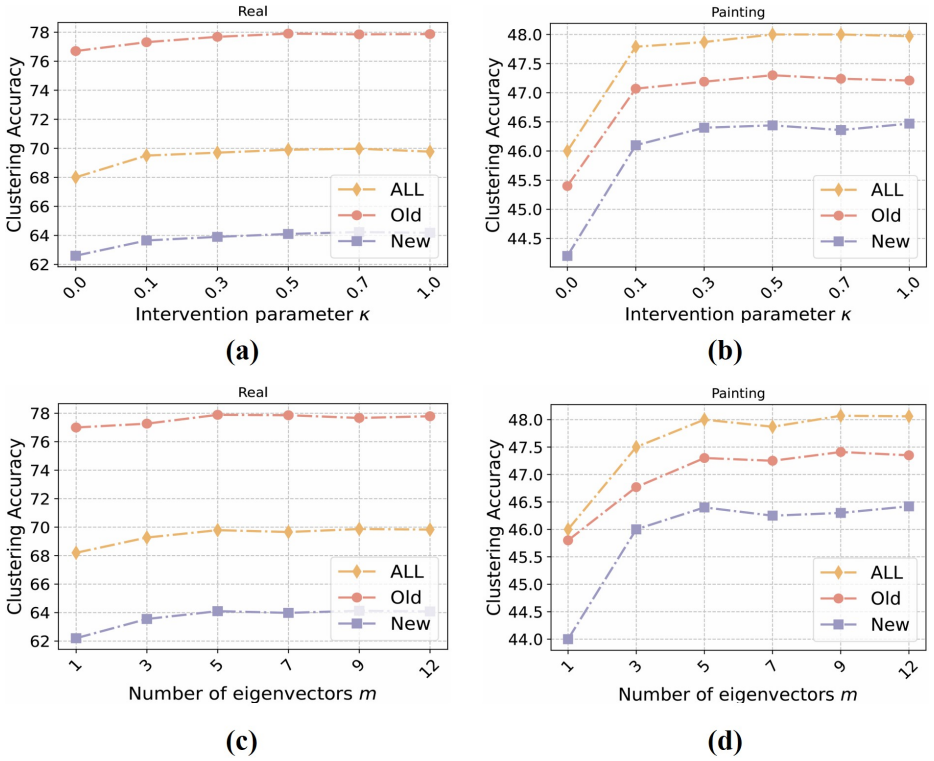

Figure 1. Hyperparameter sensitivity analysis.

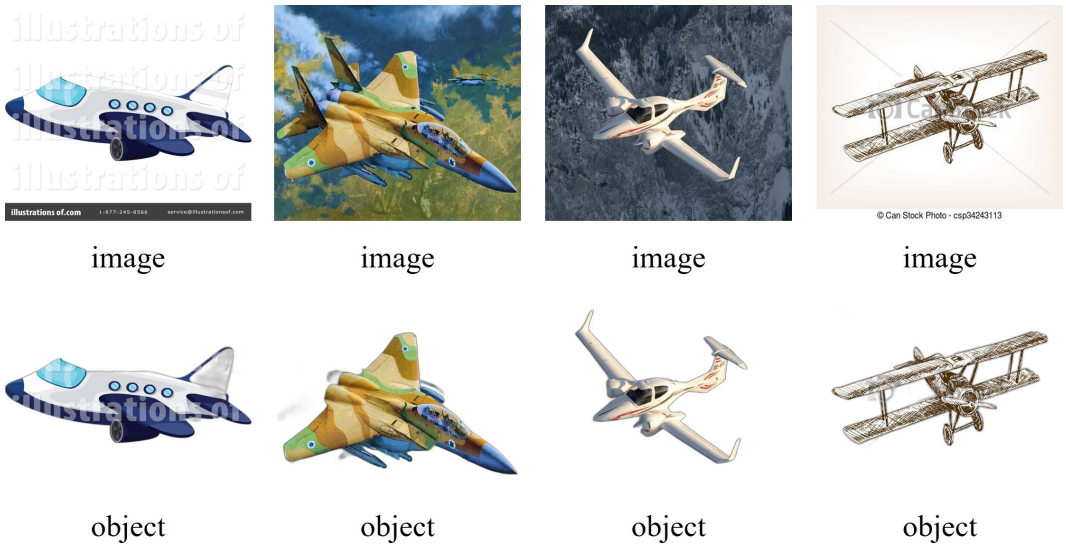

Figure 2. Object-region visualizations extracted using IS-Net [3]. These regions serve as proxies for the underlying causal semantic features.

configuration. Although class-number estimation inevitably introduces additional uncertainty, CausalGCD consistently surpasses the strongest competing method (FREE), demonstrating that our framework remains reliable even when the underlying category structure is only approximately known.

Table 15. Clustering performance using the estimated number of categories.

Methods	$\lvert C^u \rvert$	Original			Corrupted		
		All	Old	New	All	Old	New
FREE	GT. (200)	60.4	58.5	63.2	55.7	**57.1**	53.7
CausalGCD	GT. (200)	**62.2**	**60.1**	**64.9**	**57.8**	56.6	**56.9**
FREE	Est. (257)	59.3	56.1	62.0	54.4	**56.1**	52.3
CausalGCD	Est. (257)	**61.4**	**58.4**	**63.8**	**56.7**	55.7	**55.2**

References

[1] Olivier Catoni. Pac-bayesian supervised classification: the thermodynamics of statistical learning. *arXiv preprint arXiv:0712.0248*, 2007. 2

[2] Pascal Germain, Amaury Habrard, François Laviolette, and Emilie Morvant. A new pac-bayesian perspective on domain adaptation. In *International conference on machine learning*, pages 859–868. PMLR, 2016. 2, 3

[3] Xuebin Qin, Hang Dai, Xiaobin Hu, Deng-Ping Fan, Ling Shao, and Luc Van Gool. Highly accurate dichotomous image segmentation. In *European Conference on Computer Vision*, pages 38–56. Springer, 2022. 11

[4] 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. 9

[5] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7492–7501, 2022. 10