A PROOF OF CLAIM 3.1 AND THEOREM 3.1

A.1 Claim 3.1 and Proof

CLAIM 3.1. *Suppose G satisfies a Lipschitz condition; there exists a global threshold $\rho \in (0,1)$ and scale of models' learning status τ_i such that the inner set I is consistency robust,* i.e., $R_{\mathcal{B}}(G) = 0$. *More specifically,*

$$r \le (2\max\{\tau_i\}\rho - 1)\frac{2}{L},$$
$$\forall i \in [C].$$

Proof. Let $G(x)_{[k]}$ denote the predicted probability of the model on class k. Integrating the dynamic thresholds, we say that there exists a constant γ such that $G(x)_{[j]} \in [\tau_j\rho - \gamma, 1 - \tau_i\rho]$, where $j \neq \arg\max(G(x))$. Suppose I is defined by model's learning state, denote

$$I \triangleq \{x : \max(G(x)) \ge \tau_i\rho \ s.t. \ i = \arg\max(G(x))\}.$$

If $\exists x, x' \in I, x' \in \mathcal{B}(x) \cap I$ s.t. $G(x) \neq G(x')$. Specifically, define

$$\arg\max G(x) = i, \ \arg\max G(x') = j,$$

where $i \neq j$. Along with the inequality we know that

$$|G(x)_{[i]} - G(x')_{[j]}| \ge 2\tau_i\rho - \gamma - 1.$$

By the definition of Lipschitz constant, we have:

$$\begin{aligned} L\left\|x - x'\right\| &\ge \left\|G(x) - G\left(x'\right)\right\| \\ &\ge \left|G(x)_{[i]} + G\left(x\right)_{[j]} - G(x')_{[j]} - G\left(x'\right)_{[i]}\right| \\ &\ge \left|G(x)_{[i]} - G\left(x'\right)_{[j]}\right| + \left|G(x)_{[j]} - G\left(x'\right)_{[i]}\right| \\ &\ge 4(\tau_i + \tau_j)\rho - 2 - 2\gamma \\ &\ge 4(\tau_i + \tau_j)\rho - 2. \end{aligned} \tag{13}$$

As a result,

$$L\left\|x - x'\right\| \ge 4\max(\tau_i)\rho - 2,$$
$$\forall i \in [C].$$

Since by definition of $x' \in \mathcal{B}$, we have $\|x - x'\| \le r$. Combining Eq. 13 and the Lipschitz constant $L > 0$, we know that this forms a contradiction with $L\|x - x'\| \ge Lr$. Thus, $\forall x \in I, x' \in \mathcal{B}(x) \cap I$, the model predictions are consistent, *i.e.,* $R_{\mathcal{B}}(G) = 0$.

A.2 Proof Sketch for Theorem 3.1

THEOREM 3.1. *Suppose Assumption 3.1 and Claim 3.1 hold and I, O satisfies (q,μ)-constant expansion. Then the expected error of model G is bounded,*

$$\epsilon_{\mathcal{D}_T}(G) \le 4\max(q,\mu)\kappa + \mu(1+\kappa).$$

To prove the Theorem 3.1, we introduces some concepts and notations following [1]: (i) the robust set of G, $RS(G)$; (ii) the minority robust set of G on U, M.

For a given model G, define the robust set to be the set for which G is robust under input transformations:

$$RS(G) := \{x|G(x) = G(x'), \forall x' \in \mathcal{B}(x)\}.$$

Let $A_{ik} \triangleq RS(G) \cap U_i \cap \{x|G(x) = k\}$ $s.t.$ $i,k \in [C]$, where U_i denote the conditional distribution of U. Towards define the minority robust set M on U, we consider the majority class label of G:

$$y_i^{\text{Maj}} \triangleq \arg\max_{k\in[C]} \mathbb{P}_U[A_{ik}].$$

Thus, we denote

$$M \triangleq \bigcup_{k\in[C]\backslash\{y_i^{\text{Maj}}\}} A_{ik}$$

be the minority robust set of G. In addition, let

$$\widetilde{M} \triangleq \bigcup_{i\in[C]} \left(U_i \cap \{x|G(x) \neq y_i^{\text{Maj}}\}\right)$$

be the minority set of G.

By the Lemma A.1 in [1], under the (q,μ)-constant expansion, we have

$$\mathbb{P}_U[M] \le 2\max(q,\mu),$$
$$\mathbb{P}_U[\widetilde{M}] \le 2\max(q,\mu) + \mu.$$

LEMMA A.1 (UPPER BOUND ON THE INNER SET I). *Suppose the condition of Claim 3.1 holds, then*

$$\epsilon_I(G) \le \mathbb{P}_I[M] + R_{\mathcal{B}}(G).$$

Proof. Based on the definition of the minority robust set M, we know that $I \triangleq M \cup \{x : G(x) \neq G(x'), x \in I, \text{ and } x' \in \mathcal{B}(x) \cap O\}$. Therefore, we can write:

$$\begin{aligned} \epsilon_I(G) &= \mathbb{P}_I[G(x) \neq G^*(x)] \\ &= \mathbb{P}_I[M \cap (G(x) \neq G^*(x))] \\ &\quad + \mathbb{P}_I[(G(x) \neq G(x')) \cap (G(x) \neq G^*(x))] \\ &\le \mathbb{P}_I[M] + \mathbb{P}_I[\overline{RS(G)}] \\ &\le \mathbb{P}_I[M] + R_{\mathcal{B}}(G). \end{aligned} \tag{14}$$

LEMMA A.2 (UPPER BOUND ON THE OUTLIER SET O). *Let $O = \mathcal{D}_T \backslash I$, then*

$$\epsilon_O(G) \le \mathbb{P}_O[M] + \mathbb{P}_O[\widetilde{M}] + R_{\mathcal{B}}(G).$$

Proof. By the definition of the outlier set O, we note that $\{x : G(x) \neq G^*(x), \text{ and } x \in O\} \subseteq M \cup \widetilde{M} \cup (\overline{RS(G)} \backslash \widetilde{M})$. Thus, we obtain

$$\epsilon_O(G) \le \mathbb{P}_O[M] + \mathbb{P}_O[\widetilde{M}] + R_{\mathcal{B}}. \tag{15}$$

Based on the above results, we can now apply Lemma A.1 and Lemma A.2 to bound the target error $\epsilon_{\mathcal{D}_T}(G)$. Under the conditions of Theorem 3.1, we have:

$$\begin{aligned} \epsilon_{\mathcal{D}_T}(G) &= \mathbb{P}_{\mathcal{D}_T}[I]\epsilon_I(G) + \mathbb{P}_{\mathcal{D}_T}[O]\epsilon_O(G) \\ &\le \mathbb{P}_{\mathcal{D}_T}[I]\ (\mathbb{P}_I[M] + R_{\mathcal{B}}(G)) \\ &\quad + \mathbb{P}_{\mathcal{D}_T}[O](\mathbb{P}_O[M] + \mathbb{P}_O[\widetilde{M}] + R_{\mathcal{B}}(G)) \\ &\text{(Lemma A.1 and Lemma A.2)} \\ &\le \mathbb{P}_{\mathcal{D}_T}[I]\ (\kappa\mathbb{P}_U[M] + R_{\mathcal{B}}(G)) \\ &\quad + \mathbb{P}_{\mathcal{D}_T}[O](\kappa\mathbb{P}_U[M] + \kappa\mathbb{P}_U[\widetilde{M}] + R_{\mathcal{B}}(G)) \\ &\text{(Assumption 3.1)} \\ &\le \kappa\mathbb{P}_U[M] + R_{\mathcal{B}}(G) + \kappa\mathbb{P}_{\mathcal{D}_T}[O]\mathbb{P}_U[\widetilde{M}] \\ &\le 4\max(q,\mu)\kappa + \mu(1+\kappa). \end{aligned} \tag{16}$$

Table 7: Ablation study on Office-Home.

$\mathcal{L}_{seq}$	$\mathcal{L}_{lr}$	$\mathcal{L}_{alr}$	$\mathcal{L}_{air}$	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg.
✓	✓			**59.3**	79.3	82.1	68.9	79.8	79.5	67.2	57.4	83.1	72.1	58.5	85.4	72.7
✓		✓		56.7	79.4	83.0	70.2	79.7	81.2	69.3	56.3	82.9	74.1	60.5	87.9	73.4
✓			✓	57.4	79.2	84.3	73.5	81.1	81.8	73.0	56.0	83.9	77.6	61.8	89.2	74.9
✓		✓	✓	58.5	**79.8**	**85.5**	**74.8**	**82.5**	**83.1**	**73.8**	**58.4**	**85.0**	**78.2**	**63.3**	**89.6**	**76.1**

Table 8: The SND and average accuracy (%) in different losses.

$\mathcal{L}_{seq}$	$\mathcal{L}_{lr}$	$\mathcal{L}_{alr}$	$\mathcal{L}_{air}$	Office-31		Office-Home	
				SND↑	Avg.	SND↑	Avg.
✓	✓			4.4501	89.9	3.7515	72.7
✓		✓		4.4897	90.3	4.0533	73.4
✓			✓	4.5095	91.0	**4.1571**	74.9
✓		✓	✓	**4.5174**	**91.5**	4.1380	**76.1**

Table 9: Unsupervised hyperparameter selection of β.

β	A→D	A→W	D→W	W→D	D→A	W→A	SND ↑	Avg.
0	92.6	93.1	98.8	98.7	78.8	**79.0**	4.4628	90.2
0.25	93.0	94.7	98.9	99.8	79.1	**79.0**	4.4175	90.7
1	95	91.8	98.5	**100.0**	78.0	76.0	4.5086	90.8
2	**96.4**	**95.1**	**99.0**	**100.0**	**80.0**	78.2	**4.5174**	**91.5**

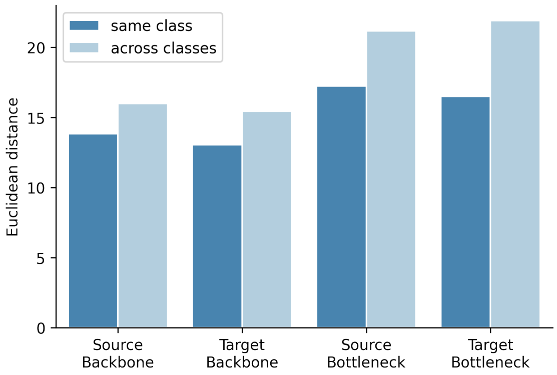

Figure 4: Histogram of the Euclidean distance within the same class and across classes on Office-Home.

B ADDITIONAL ANALYSIS

B.1 Ablation studies on Office-31 and Office-Home

We show extended ablation studies with our CtO due to lack of space in the main paper, but the main results are the same. Table 7 shows the ablation analysis on Office-Home. It is worth noting that every part of our approach helps improve performance, except for Ar→Cl task.

Intuitively, a well-adapted target classifier should have strong intra-class compactness and class-seen sparsity in the feature space. Under unsupervised scenario, Soft Neighborhood Density (SND) [25] can consider the implicit local neighborhood density in the target domain. In general, a larger SND represents more compact clustering. Therefore, we additionally validate the SND and accuracy under different losses. In our approach, Adaptive Input-consistency Regularization (AIR) is employed to promote the participation of outlier samples during training via flexible thresholds, resulting in a higher SND. However, it should be noted that relying solely on AIR cannot ensure the accuracy of pseudo-labels due to confirmation bias. Additionally, while misalignment does improve SND scores, it does not achieve better performance. We also observe that the combination of ALR and AIR yields improved accuracy and competent SND. This indicates that the extended property enhances local neighborhood information, facilitating the propagation of structural information among subpopulations.

B.2 Analysis of the decay factor β

In Table 9, we show the accuracy of various decay factor β values on Office-31. Following the unsupervised hyperparameter selection strategy of AaD [35], we used SND to select β. Notably, for a fair comparison, we set the hyperparameter β the same as AaD in the comparison experiments (Section 5). On hard transfer tasks (e.g., D→A and W→A), a smaller β can achieve the best or second-best performance. For these tasks, the target samples of the source model are scattered, meaning there are numerous outlier samples. The lower β prompted inter-sample sparseness, which prevented the outlier samples from collapsing into a limited set of categories. For easy transfer tasks, the target features tend to cluster quickly and form inter-class boundaries. In this case, a larger β allows the model to focus more on consistency loss, which prompts the clustering process. In addition, combined with the results from Table 8 and Table 9, it indicates that SND can effectively select the optimal β.

C ANALYSIS OF DIFFERENT SIMILARITY MEASURES

To check the effect of different metric functions on target feature clustering, we compared the Cosine similarity and the Euclidean distance on Office-Home for feature similarity. Fig. 4 shows the average Euclidean distance for all tasks on Office-Home. A smaller value indicates more similarity. It can be seen that the Euclidean distance also indicates the presence of rich semantic information in the high-dimensional features of the source model backbone network. However, compared to Fig. 2, the differences in similarity based on Euclidean distance are insignificant. As is well known, the Euclidean distance reflects absolute differences in values, while the cosine distance reflects relative differences in direction. Therefore, Cosine similarity maintains "1 for identical, 0 for orthogonal, -1 for opposite" in high-dimensional space. Euclidean distance, in contrast, is influenced by dimensions, and its numerical space is not unstable.

Layer		Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg.
Layer 4 (source)	same class	0.355	0.464	0.378	0.305	0.386	0.341	0.322	0.314	0.363	0.301	0.305	0.422	0.355
	across classes	0.189	0.135	0.106	0.152	0.125	0.122	0.159	0.201	0.126	0.126	0.163	0.115	0.143
Layer 4 (target)	same class	0.298	0.429	0.373	0.322	0.429	0.373	0.322	0.298	0.373	0.322	0.298	0.429	0.356
	across classes	0.121	0.119	0.102	0.120	0.119	0.102	0.120	0.121	0.102	0.120	0.121	0.119	0.115
Bottleneck (source)	same class	0.278	0.461	0.407	0.257	0.362	0.323	0.266	0.218	0.357	0.354	0.327	0.510	0.343
	across classes	0.054	0.026	0.002	0.029	0.014	0.015	0.022	0.070	0.003	0.022	0.102	0.028	0.032
Bottleneck (target)	same class	0.370	0.549	0.550	0.367	0.481	0.484	0.384	0.306	0.507	0.414	0.332	0.536	0.440
	across classes	0.060	0.014	0.009	0.031	0.012	0.033	0.007	0.061	0.004	0.001	0.040	0.002	0.023

Table 10: Cosine similarity within the same class and across classes in each task on Office-Home.

Layer		Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg.
Layer 4 (source)	same class	14.848	12.155	12.918	13.849	11.914	12.650	16.183	17.240	14.453	14.183	13.989	11.680	13.839
	across classes	16.674	15.556	15.548	15.449	14.317	14.597	18.152	18.730	16.924	15.978	15.410	14.548	15.990
Layer 4 (target)	same class	12.800	12.961	12.275	14.183	12.961	12.275	14.183	12.800	12.275	14.183	12.800	12.961	13.055
	across classes	14.422	16.174	14.750	16.396	16.174	14.750	16.396	14.422	14.750	16.396	14.422	16.174	15.435
Bottleneck (source)	same class	19.041	15.302	16.117	18.614	16.453	17.260	18.656	20.286	16.868	16.799	17.428	14.102	17.244
	across classes	21.975	20.965	21.160	21.529	20.717	20.926	21.778	22.301	21.182	21.033	20.275	20.274	21.176
Bottleneck (target)	same class	19.813	13.601	14.845	16.137	13.223	14.362	18.703	22.772	16.631	16.223	18.673	12.934	16.493
	across classes	24.252	20.307	22.235	20.172	18.489	19.838	23.969	26.605	23.761	21.410	22.489	19.170	21.891

Table 11: Euclidean distance within the same class and across classes in each task on Office-Home.

Particularly in the case of distribution shifts, the variance of the sample fluctuations is too large, leading to poor Euclidean distance performance.

We also shows feature similarities among samples within the same class and across classes in each transfer task. As shown in Table 10 and Table 11, the differences between the Euclidean distance within the same class and that across classes are not clear if the distributions are significantly different (e.g., Ar→Cl, Pr→Cl, Rw→Cl tasks). Weak inter-category discrimination exacerbates spurious clustering, which biases the adaptation process.

Through the experiment, we notice two things: 1) the source model contains sufficient inductive biases; and 2) under domain shift conditions, there is a strong correlation between the metric method and target feature clustering. In future work, one possible direction is to study how different metrics affect the performance of target clustering.