CLEAR: Class-Aware Ensemble and Curriculum Refinement for Multi-Source Domain Adaptation

Anonymous submission

Abstract

We address multi-source domain adaptation (MSDA), where labeled data from multiple source domains assists adaptation to an unlabeled target domain. Naïve aggregation of source predictions amplifies noise in target pseudo-labels and yields unstable alignment. Existing methods rely on adversarial training or confidence filtering but remain sensitive to noisy pseudo-labels and require careful tuning. To overcome these challenges, we propose CLEAR (Class-consistent Ensemble And Refinement), a unified self-refinement framework that integrates class-aware pseudo-label aggregation with curriculum-guided optimal transport alignment to achieve robust and stable adaptation. The ensemble phase leverages class-aware optimal transport distances to generate reliable pseudo-labels, even under large domain shifts. The refinement phase then strengthens intra-class couplings and suppresses inter-class ones. It also integrates augmentation consistency and progressively tightens coupling constraints across refinement cycles, leading to more stable adaptation. Extensive experiments on Office-31, Office-Home, and DomainNet demonstrate that CLEAR consistently outperforms state-of-the-art MSDA approaches across diverse shift conditions.

Introduction

Deep learning has enabled state-of-the-art performance in a wide range of tasks, including image recognition (Krizhevsky et al., 2012), language understanding (Devlin et al., 2019), and audio processing (Purwins et al., 2019). These advances have been made possible by supervised learning frameworks that rely on access to large-scale, manually labeled datasets. However, in real-world scenarios, models often encounter new environments where labeled data is expensive to obtain. Unsupervised Domain Adaptation (UDA) addresses the lack of labeled data in new environments by transferring knowledge from a labeled source domain to an unlabeled target domain. Early UDA methods, such as Domain-Adversarial Neural Networks (DANN) (Ganin et al., 2016), focus on aligning feature distributions between a single source and target domain. However, these methods struggle in real-world settings, where data comes from multiple diverse sources. Leveraging such complementary sources is essential for effective target adaptation. This has led to the development of Multi-Source Domain Adaptation (MSDA) methods (Peng et al., 2019), which aim to aggregate knowledge from several labeled domains to achieve better generalization on the target domain.

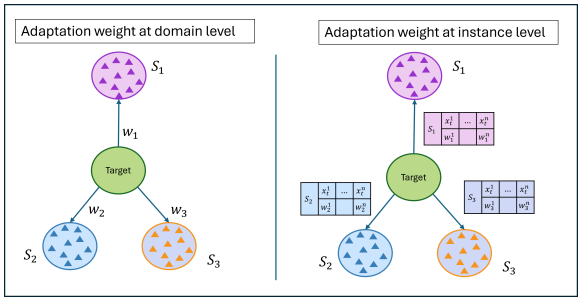

Figure 1: Comparison of domain-level vs. instance-level adaptation. Left: Traditional MSDA assigns a fixed weight w_i per source S_i. Right: Instance-level adaptation assigns sample-specific weights w_i^k for each target sample x_t^k, enabling finer-grained knowledge transfer.

To make the most of multiple source domains, many methods have been proposed to combine either the domains themselves or the models trained on them for adapting to a target domain(Wilson, Doppa, and Cook 2023). One common approach is to give more weight to source domains that are more similar to the target, allowing a shared adaptation model to focus on useful knowledge(Wen, Greiner, and Schuurmans 2020; Turrisi et al. 2022). Another popular strategy is to train a separate model for each source domain and then combine their predictions using weighted averaging during inference(Venkat et al. 2020). Although these methods have shown good performance, most of them assign a single importance score to each source domain for all target samples. This can cause useful predictions from smaller or less-relevant domains to be ignored, even if they are accurate for some target examples. Therefore, it is important to move beyond domain-level weighting and consider assigning weights at the sample level, so that each target prediction can benefit from the most relevant sources. To capture the most relevant and semantically aligned knowledge from each source domain, we propose CLEAR, Class-consistent Ensemble and Refinement, a class-aware adaptation framework that performs instance-level aggregation of source predictions using semantic-guided optimal transport.

The method begins by training domain-specialized subnetworks and constructing a class similarity-aware cost matrix derived from predefined semantic embeddings. For each target sample, an ensemble strategy aggregates predictions from the source subnetworks, where the aggregation weights are determined by prediction confidence and output diversity. Subnetworks that produce high-confidence and semantically consistent predictions are assigned greater influence in the ensemble.The aggregation network not only preserves these strong predictions but also generates more reliable pseudo-labels than any individual subnetwork. These high-confidence pseudo-labels are then used to refine the domain-specialized subnetworks by enhancing their alignment with the target domain. Specifically, the pseudo-labels are used to enforce class-aware marginal alignment between source features and the features learned by the aggregation network. This forms a cycle of mutual refinement: the subnetworks contribute increasingly confident and informative pseudo-labels, while the aggregation network progressively enhances the discriminative capacity of each subnetwork. Our main contributions can be summarized as follows.

- We introduce an instance-level, class-aware ensemble strategy that assigns adaptive weights to source predictions based on class-level semantic similarity, allowing fine-grained and context-sensitive aggregation for each target sample.
- We design a dynamic semantic cost matrix that integrates both predefined semantic priors and learned classifier structures, serving as a regularizer to guide the alignment between source subnetworks and the aggregation network.
- Extensive experiments on Office-31, Office-Home, and DomainNet demonstrate that CLEAR achieves competitive performance close to state-of-the-art methods, while requiring significantly fewer training steps and reduced computation time.

Related Works

Domain Adaptation

Domain adaptation aims to transfer knowledge from a labeled source domain to an unlabeled target domain in order to address domain shift. A prominent setting within this field is unsupervised domain adaptation (UDA), where no target labels are available during training. Many UDA methods are grounded in the theoretical framework of (Ben-David et al. 2010), which emphasizes minimizing the discrepancy distance between source and target domains. This discrepancy is often reduced using statistical techniques such as MMD (Long et al., 2018) and CORAL (Sun and Saenko, 2016), which align global distribution statistics. Adversarial approaches like DANN (Ganin et al. 2016) and MCD(Saito et al. 2018)promote domain-invariant feature learning through adversarial training. More recent efforts, such as the method by Zhou et al. (2023), focus on fine-grained feature alignment. Despite their effectiveness in single-source settings, these methods often fail to generalize well in multi-source domain adaptation (MSDA), where diverse and conflicting source distributions pose additional challenges.

Multiple Source Domain Adaptation

MSDA addresses the challenge of aggregating knowledge from diverse source domains to an unlabeled target. It faces the twin problem of source-target diversity and source models domain shifts. To better aggregate knowledge from multiple sources, (Shui et al. 2021), attempt to emphasize source domains that are similar to target domain. But due to this domain level weightage, some domains that are useful for a particular class or target are discarded. Few methods like (Lin et al. 2021) enforce prediction consistency which enhances robustness but suppress domain specific knowledge that is unique to each source. Teacher–student frameworks in MSDA typically distill knowledge from source-trained teachers to a target-trained student (Amosy and Chechik 2020). However, the student does not refine the teachers making the transfer one-sided and limiting adaptation when the target domain offers new transferable cues. (Li et al. 2022) use attention based weights for source domains at classifier level to dynamically align to target domain. However these methods add extra burden of training and inference time.

Semi Supervised Learning

An important component of MSDA is self-supervised learning. Early methods in semi-supervised learning relied on pseudo-labels, effectively converting the problem into a supervised learning task. Some approaches, such as (**?**), introduced consistency-based strategies to enforce prediction invariance under data augmentations. FixMatch (**?**) and FreeMatch ((Wang et al. 2022)) apply dynamic thresholding to manage the trade-off between prediction confidence and training progression. However, these methods overlook the incorporation of semantic information during consistency training, which is crucial for achieving stable and robust adaptation in MSDA.

The Proposed Method

We consider the multi-source domain adaptation (MSDA) setting, where we are given m labeled source domains $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_m\}$ and an unlabeled target domain $\mathcal{T}$. Each source domain $\mathcal{S}_i$ contains samples (x_s^i, y_s^i) drawn from a domain-specific distribution $P_i(x, y)$, while the target domain samples x_t follow a different distribution $P_T(x)$. The total number of semantic categories is denoted by K, and all domains share the same label space $\mathcal{Y} = \{1, 2, \ldots, C\}$. To efficiently learn from multiple domains, we adopt a shared feature extractor $f_c(\cdot)$ to capture domain-invariant representations that can generalize across all domains. On top of this shared backbone, we attach domain-specific residual modules $\phi_i(\cdot)$ for each source domain to capture domain-specific variations. Thus, the feature extractor for the i-th source domain is defined as $f_i = \phi_i \circ f_c$, where $\circ$ denotes the composition of the function. Consequently, each source domain has its own classifier head h_i for category prediction. For

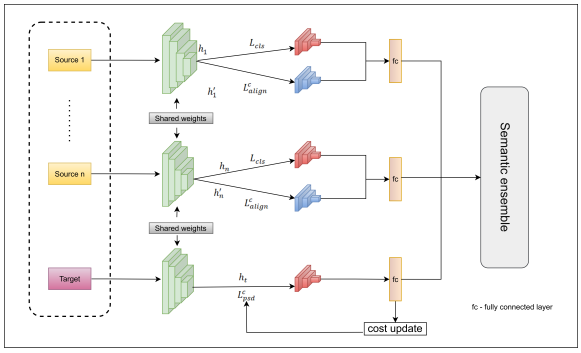

Figure 2: Overview of the proposed method. Each source-specific network is trained using classification and alignment losses, while the target network learns from pseudo-labels. A semantic ensemble aggregates predictions from all sources. The cost matrix is dynamically updated during training to guide alignment of source and target features

the target domain, we deploy a aggregator network consisting of the shared backbone f_c and an ensemble classifier h_e, which aggregates transferable knowledge from all source domains. During training, the domain-ensemble network receives pseudo-labels generated from the source domain subnetworks, refined through instance-level aggregation and class aware alignment.

This architecture allows for two complementary pathways:

1. Source domain subnetworks $\mathcal{S}_i = \{f_i, h_i\}$ preserve the unique transferable knowledge of each source domain.
2. Aggregator network $\mathcal{E} = \{f_c, h_e\}$ which integrates high-confidence information from all sources for better target generalization.

The overall model as shown in Figure 2, includes three major components: (i) source-specific subnetworks that maintain strong supervised performance on their respective domains, (ii) a domain aggregator network that refines target representations, and (iii) dual-level optimal transport modules—Margin-Aware Optimal Transport (MAOT) for stable feature alignment and Dynamic OTMatch for class-distribution alignment. These components work in a cycle-refinement manner, ensuring robust domain alignment and adaptive pseudo-label refinement.

.

Semantic Optimal Transport Ensemble

The main challenge in multi-source domain adaptation (MSDA) lies in effectively aggregating knowledge from diverse source domains to achieve good generalization on the target domain. Many existing methods rely on domain-level weighting, which fails to account for how individual target samples interact with each source domain. In reality, different source domains may specialize in different subsets of classes, and global weighting schemes often overlook such fine-grained class-specific contributions—leading to suboptimal adaptation. Instance-level weighting of source predictions addresses this by offering sample-specific adaptability. However, traditional approaches typically compute these weights based on prediction consistency using cross-entropy, which enforces agreement between source networks or among neighboring samples. This reliance on cross-entropy overlooks the geometric structure of the probability simplex. Since cross-entropy tends to push predictions toward the simplex vertices, it implicitly encourages sharp, overconfident outputs—even during early training stages when models may be uncertain. As a result, the instance-level weights derived for each target sample can become unstable, especially under large domain shifts. The models may struggle to produce reliable pseudo-labels during early adaptation, making the resulting weights fluctuate unpredictably. Furthermore, it is unrealistic to expect confident, sharp predictions on target samples using models trained solely on source domains. Altogether, these factors make traditional instance-based weighting strategies prone to instability and unreliability during training. .To address this, we propose a semantic-aware aggregation strategy in which the weight assigned to each source domain for a given target instance is determined by the Wasserstein distance between that source's prediction and the mean prediction across all sources. This transport-based measure respects the geometry of the probability simplex and allows semantically related class distributions to align at a lower cost, leading to more flexible and stable weighting during training.

Inspired by the fact that, beyond prediction consistency and confidence, a useful source model should also exhibit semantic consistency and provide diverse predictions without collapsing to a single class, we propose a semantic optimal transport-based aggregation mechanism. This approach is geometry-aware and accounts for semantic relationships among classes by incorporating a cost matrix derived from word-level semantic embeddings. For each target sample x_t, we compute a weighted combination of source predictions, where the weight assigned to each source model is determined by a combination of its semantic alignment with other sources, its predictive confidence, and the entropy of its output. Specifically, for each source model i, we define a score as

$$\begin{aligned} score_i(x_t) = & -\alpha \cdot OT\left(p_i(x_t), \bar{p}(x_t)\right) \\ & + \beta \cdot \max_c p_i^c(x_t) \\ & + \gamma \cdot \textstyle\sum_c p_i^c(x_t) \log \frac{1}{p_i^c(x_t)} \end{aligned} \quad (1)$$

where $p_i(x_t) \in \Delta^K$ is the softmax prediction of the i-th source model, $\bar{p}(x_t) = \frac{1}{N} \sum_{j=1}^{N} p_j(x_t)$ is the average prediction across all m sources, and $OT(p_i, \bar{p})$ is the optimal transport distance computed using a class-aware cost matrix that reflects semantic similarity between labels. The term $\max_c p_i^c(x_t)$ captures the confidence of the prediction, while the negative entropy $-\sum_c p_i^c(x_t) \log p_i^c(x_t)$ quantifies its diversity.

Geometry-aware optimal transport is used to measure the semantic discrepancy between source and target predictions. The Wasserstein distance is computed using a cost matrix derived from pre-trained word-level embeddings, which

capture semantic relationships between class labels. This allows the model to tolerate higher cross-entropy values when the predicted class distributions are semantically related. In contrast to traditional consistency losses that favor sharp predictions, our formulation does not penalize soft predictions if the probability mass is spread meaningfully across semantically similar classes. As training progresses, the model is expected to become more confident in its predictions since the objective remains classification. However, early in training, such confidence may not reflect semantic correctness. Our formulation accommodates this through a curriculum learning perspective—parameters such as α, β, and γ be adjusted progressively based on model maturity, allowing the ensemble to transition from soft to sharper predictions in a controlled manner. Diversity is also an essential component of our weighting strategy. Since the final prediction is an aggregation of multiple source models, the model's ability to discriminate between classes depends on the diversity of its predictions. Incorporating entropy as a diversity measure encourages each source model to retain class-wise discriminative power, especially in the early training stages when overconfidence is unreliable. Although optimal transport, confidence, and diversity may have conflicting objectives, curriculum scheduling of their relative weights allows the ensemble to strike a dynamic balance: early training favors diversity and semantic tolerance, while later training prioritizes confident and precise predictions. The final weights over sources are obtained via a softmax with temperature τ:

$$w_i(x_t) = \frac{\exp\left(\frac{score_i(x_t)}{\tau}\right)}{\sum_{j=1}^{N} \exp\left(\frac{score_j(x_t)}{\tau}\right)},$$

and the refined pseudo-label for x_t is computed as a convex combination of the source predictions using these weights. This strategy adaptively emphasizes sources that offer confident, semantically aligned, and informative predictions, resulting in more stable and reliable pseudo-labels for adaptation. Finally, the semantic OT ensemble pseudo-label for target sample x_t is obtained as a weighted average of the source predictions:

$$p^*_{x_t} = \sum_{i=1}^{m} w_{x_t,i}\, p^i_{x_t}. \tag{2}$$

This formulation preserves semantically consistent relationships between classes, yielding more reliable pseudo-labels and improving the aggregation of transferable knowledge across multiple sources.

Semantic regularizer for Margin Alignment

We introduce the concept of mutual refinement between source domain subnetworks and a unified aggregator network to enable bidirectional transfer of knowledge. The aggregator network is trained using an instance-level aggregation strategy that adaptively combines predictions from all source subnetworks. In turn, it guides the source subnetworks to adapt to the target domain through margin-aware feature alignment. In prior work, margin-aware alignment is conducted between features from each source network (on source samples) and features from the aggregator network (on target samples), using predictions from their respective classifiers. This alignment is performed in the class probability space, where the discrepancy in predictions serves as a proxy to align features adversarially. However, such alignment overlooks class-level semantics and may enforce alignment between unrelated categories. To address this, we propose a semantic-aware regularizer that penalizes alignment between semantically distant classes and allows flexibility among semantically related ones. The regularizer is built using a dynamic cost matrix that combines the structural information captured by the classifier heads with external word-level semantics. This cost matrix ensures that the alignment process is not only domain-consistent but also semantically guided, allowing more meaningful and discriminative adaptation. Formally, for each source domain i, the alignment loss is defined as:

$$\begin{aligned}\mathcal{L}^{(i)}_{align} = & \underbrace{CE\left(D_i(f_s(x_s)),\ y_s\right)}_{sourceclassification} \\ & + \underbrace{CE\left(1-\sigma(D_i(f_t(x_t))),\ \hat{y}_t\right)}_{targetadversarialalignment} \\ & + \underbrace{\alpha \sum_{k=1}^{K} \mathcal{C}_{k,\hat{y}_t} \cdot P_k}_{semanticregularizer} \quad . \end{aligned} \tag{3}$$

Here, $f_s(x_s)$ and $f_t(x_t)$ denote the features of the source and target samples respectively; D_i is the domain discriminator for source i; y_s is the ground-truth label of the source sample; $\hat{y}_t$ is the pseudo-label for the target sample x_t; $P \in R^K$ is the predicted class distribution for x_t; $\mathcal{C} \in R^{K\times K}$ is the semantic cost matrix; and α is a regularization weight.

The semantic cost matrix $\mathcal{C}$ is dynamically computed as:

$$\mathcal{C}_{ij} = 1 - \left\langle \beta\, \hat{w}_i + (1-\beta)\, \phi_i,\ \beta\, \hat{w}_j + (1-\beta)\, \phi_j \right\rangle, \tag{4}$$

where $\hat{w_i}$ is the normalized weight vector of the classifier head for class i, ϕ_i is the normalized pretrained word embedding for class i, and $\beta \in [0,1]$ balances the influence of model-learned structure and external semantic priors.

This formulation ensures that alignment respects class semantics: it penalizes alignment between semantically distant class predictions while allowing smoother alignment for semantically related ones. The regularizer acts as a semantic filter over adversarial learning, reducing misalignment noise and promoting meaningful consistency between label spaces.

Curriculum-Aware Semantic label refinement

As the margin-aware feature refinement relies on pseudo-labels produced by the aggregation network, we incorporate a curriculum-aware confidence thresholding strategy to balance exploration and exploitation. This helps the model focus on easy samples in the early stages and gradually include harder ones, accelerating convergence. Specifically,

we adopt a dynamic threshold schedule that increases over time according to:

$$Threshold(t) = 0.7 + (0.95 - 0.7) \cdot \frac{t}{T}, \tag{5}$$

where t is the current training step and T is the total number of training steps.

In addition to thresholding, we apply an augmentation consistency loss between the original and augmented target samples. Unlike strict consistency objectives, our method is semantically aware: the aggregation network is not penalized if the predictions from the two views differ but remain close in the semantic space defined by the cost matrix $\mathcal{C}$. We formalize this with a cost-regularized pseudo-label loss:

$$\mathcal{L}_{psd}^{\mathcal{C}} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{1}_{conf_b \geq \tau(t)} \sum_{i=1}^{K} \sum_{j=1}^{K} \mathcal{C}_{ij} \cdot p_b^{(i)} \cdot q_b^{(j)}, \tag{6}$$

where p_b and q_b are the predicted distributions for the original and augmented target views of the b-th sample, and $\tau(t)$ is the dynamic threshold defined above. This formulation ensures semantic consistency while maintaining flexibility across augmentations.

Total training procedure

Thus, the total objective for training both the source-specific networks and the domain-ensemble network integrates supervised source classification, semantic pseudo-label refinement, and alignment with semantic regularization. The supervised classification loss is denoted as:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \mathcal{L}_{psd}^{\mathcal{C}} + \beta \sum_{i=1}^{m} \mathcal{L}_{align}^{(i,\mathcal{C})}. \tag{7}$$

where h_s^i is the classifier for the i-th source domain and h_t is the domain-aggregator network.

With the remaining terms defined earlier, the overall training objective becomes:

$$\mathcal{L}_{total} = \mathcal{L}_{cls} + \mathcal{L}_{psd}^{\mathcal{C}} + \beta \sum_{i=1}^{m} \mathcal{L}_{align}^{(i,\mathcal{C})}, \tag{8}$$

where β is a hyperparameter that controls the influence of the alignment loss. This formulation enables end-to-end training of the entire framework. Notably, the semantic pseudo-labels used for refining both the source-specific and domain-ensemble networks are dynamically updated in each iteration based on confidence and semantic consistency.

Experiments

Implementation Details

We evaluate our method on three widely-used multi-source domain adaptation benchmarks: Office-31, Office-Home, and DomainNet. Office-31 (Saenko et al. 2010)contains 4,110 images across 31 categories from three domains: Amazon (A), Webcam (W), and DSLR (D). Office-Home(Venkateswara et al. 2017) includes 15,500 images spanning 65 categories from four domains: Art (Ar), Clipart (Cl), Product (Pr), and Real-World (Rw). DomainNet(Peng et al. 2019) is a large-scale benchmark with approximately 0.6 million images covering 345 categories across six diverse domains: Clipart (Clp), Infograph (Inf), Painting (Pnt), Quickdraw (Qdr), Real (Rel), and Sketch (Skt). In our MSDA setup, we follow the standard protocol where one domain is selected as the target, and the remaining domains are treated as sources. We evaluate our approach against several representative state-of-the-art methods in multi-source domain adaptation,including DANN(Ganin et al. 2016) , MCD(Saito et al. 2018), DCA(Saito et al. 2018), PTMDA(Ren et al. 2022), DRT(Li et al. 2021), WADN(Shui et al. 2021), SImpAI(Venkat et al. 2020), SSG(Yuan et al. 2022), MRF-MSDA(Xu et al. 2022) BDT(Kundu et al. 2022) and CSR(Zhou et al. 2024). Following established evaluation protocols, we report results under three settings: (1) Single Best (SB), which reflects the highest accuracy achieved by any individual source domain model; (2) Source Combine (SC), where a single model is trained using the merged data from all source domains; and (3) Multi-Source (MS), which corresponds to methods explicitly designed for multi-source domain adaptation. In our experiments, we adopt ResNet-50 pre-trained on ImageNet as the backbone for the Office-31 and Office-Home datasets,. For DomainNet we used ResNet-101 due to its larger scale and complexity. To enhance training, we apply RandAugment (Cubuk et al. 2020) as the data augmentation strategy for target samples.

Experimentation Results

We performed experiments for Office31, Office-Home and DomainNet in the corresponding tables.

Table 1: Classification accuracy(%) on Office-31 dataset.

Category	Method	$\rightarrow A$	$\rightarrow D$	$\rightarrow W$	Avg
Single Best	DANN	68.2	99.4	96.8	88.1
	MCD	69.7	**100.0**	98.5	89.4
Source Combine	DANN	67.6	99.7	98.1	88.5
	MCD	68.5	99.4	99.3	89.0
Multi-Source	MFSAN	72.7	99.5	98.5	90.2
	SImpAI	70.6	99.2	97.4	89.0
	SSG	71.3	**100.0**	95.5	90.3
	DCA	55.1	99.6	98.9	91.2
	PTMDA	75.4	**100.0**	99.6	91.7
	CSR	78.6	**100.0**	**99.6**	92.7
	Ours	**78.9**	**100.0**	99.3	**92.7**

Ablation Study

To assess the effectiveness of individual components in our CLEAR framework, we perform an ablation study on the Office-Home and Office-31 datasets. The table summarizes the results across target domains under various configurations. We start with the baseline setup, which uses instance-level ensemble pseudo-labels (ot) without incorporating any target-specific refinement losses. Adding EMD-based transport for pseudo-label weighting (ot-emd) provides slight improvements in the Product domain (from 85.02% to 85.25%)

Table 2: Classification accuracy(%) on Office-Home

Category	Method	Ar	Pr	Cl	Rw	Avg
Single Best	DANN	67.9	80.4	55.9	75.8	70.0
	MCD	69.1	79.6	52.2	75.1	69.0
Source Combine	DANN	68.4	79.5	59.1	82.7	72.4
	MCD	67.8	79.2	59.9	80.9	71.9
Multi-Source	MFSAN	72.1	80.3	62.0	81.8	74.1
	DCA	72.1	80.5	63.6	81.4	74.4
	SImpAI	70.8	80.2	56.3	81.5	72.2
	WADN	73.4	86.3	70.2	87.3	79.4
	BDT	72.6	85.9	67.4	83.6	77.4
	CSR	**76.7**	**86.8**	**71.4**	85.5	**80.1**
	Ours	74.8	85.8	70.9	**86.7**	79.5

Table 3: Classification accuracy(%) on DomainNet dataset.

	c						
Method	Clp	Inf	Pnt	Qdr	Rel	St	Avg
SImpAI	66.4	26.5	56.6	18.9	68.0	55.5	48.6
SSG	68.7	24.8	55.7	18.4	66.8	56.3	48.8
DRT+ST	71.0	31.6	61.0	12.3	71.4	60.7	51.3
MRF-MSDA	63.9	28.7	56.3	16.8	67.1	54.3	47.9
PTMDA	66.0	28.5	58.4	13.0	63.0	54.1	47.2
CSR	**73**	**28.1**	58.8	**26.0**	71.1	**60.7**	**52.9**
Ours	71.4	27.6	56	23.4	**71.4**	59.6	51.5

and a minor drop in the Rw domain (from 84.86% to 84.40%). These results suggest that precise transport plans can enhance model weighting, although in some cases, such as Rw, EMD may introduce sensitivity to local distribution mismatches in soft predictions.

To evaluate the impact of target-side refinement, we introduce our OT-based pseudo-label loss (tar-emd), which leads to noticeable performance gains in the Product and Amazon domains, reaching 85.36% and 78.72%, respectively. These results reinforce our hypothesis that geometry-aware cost functions enable more meaningful pseudo-label refinement compared to standard confidence-based thresholds. We further examine the effect of semantic alignment by incorporating an OT-based alignment loss (align). Relative to the base OT configuration, this addition improves performance in the Product, Realworld, and Webcam domains, demonstrating its effectiveness in promoting class-level consistency across domains. In particular, the Rea;World domain achieves 85.55%, the highest score recorded in this setup.

Finally, we assess the Sinkhorn-regularized variant, which achieves competitive performance in the Product and Rw domains, matching or slightly outperforming the EMD-based counterpart. However, due to the absence of results for the Amazon and Clipart domains in this setting, a complete comparison is not possible. Nevertheless, Sinkhorn demonstrates promise as a scalable alternative to EMD, particularly in scenarios where computing exact transport is resource-intensive.

Overall, these findings confirm that each component—OT-based pseudo-label refinement, semantic alignment loss, and the choice of transport metric—makes a distinct contribution to overall performance. Their integration results in consistent improvements over the remaining baseline methods, offering a more principled and computationally efficient approach to modeling class-level semantic structure.

Table 4: Ablation of OT-based refinement. We start from OT ensemble weighting, then add OT-based alignment loss, and finally include OT-based pseudo-label refinement.

Domain	**OT Only**	**+ Align**	**+ Pseudo**
Product	85.2	85.1	**85.8**
RealWorld	84.8	84.8	**86.7**
Art	71.6	73.6	**74.8**
Amazon	77.8	77.9	**78.9**
DSLR	100	100	100.00
Webcam	99.3	**99.3**	**99.3**

Conclusion

In this paper, we introduced the CLEAR framework for multi-source domain adaptation, leveraging optimal transport to compute geometry-aware instance-level aggregation weights. Unlike cross-entropy-based consistency, which enforces sharp predictions throughout training, our method allows prediction probability mass to spread across semantically related classes, capturing richer inter-class relationships. In the second stage, we construct a cost matrix using predefined word embeddings and classifier heads from the aggregator network, which guides augmentation consistency regularization. Mutual refinement between the source-specific subnetworks and the aggregator network is achieved via conditional feature alignment, dynamically steered by the evolving cost matrix. This class-aware alignment brings target samples of semantically similar classes closer in the feature space. Notably, both source weight estimation and cost matrix computation scale with the number of classes, offering computational advantages over clustering-based methods, which scale with the number of samples. Our approach achieves performance comparable to state-of-the-art methods while converging more rapidly. Empirically, we find that incorporating class-level semantic information enhances both convergence speed and training stability. In future work, we aim to further explore adaptive cost matrix learning and extend CLEAR to more challenging open-set and universal domain adaptation settings.

References

Amosy, O.; and Chechik, G. 2020. Teacher-Student Consistency For Multi-Source Domain Adaptation. *arXiv preprint arXiv:2010.10054*.

Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Vaughan, J. W. 2010. A theory of learning from different domains. *Machine learning*, 79(1): 151–175.

Cubuk, E. D.; Zoph, B.; Shlens, J.; and Le, Q. V. 2020. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Con-*

ference on Computer Vision and Pattern Recognition Workshops, 702–703.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1): 2096–2030.

Kundu, J. N.; Kulkarni, A. R.; Bhambri, S.; Mehta, D.; Kulkarni, S. A.; Jampani, V.; and Radhakrishnan, V. B. 2022. Balancing discriminability and transferability for source-free domain adaptation. In *ICML*, 11710–11728.

Li, K.; Lu, J.; Zuo, H.; and Zhang, G. 2022. Dynamic Classifier Alignment for Unsupervised Multi-Source Domain Adaptation. *IEEE Transactions on Knowledge and Data Engineering*, 35: 4727–4740.

Li, Y.; Yuan, L.; Chen, Y.; Wang, P.; and Vasconcelos, N. 2021. Dynamic transfer for multi-source domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10998–11007.

Lin, Z.; Tang, H.; Gong, M.; Zhang, K.; and Tao, D. 2021. DAC-Net: Domain Attention Consistency for Multi-Source Domain Adaptation. In *Proceedings of the British Machine Vision Conference (BMVC)*.

Peng, X.; Bai, Q.; Xia, X.; Huang, Z.; Saenko, K.; and Wang, B. 2019. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1406–1415.

Ren, C.-X.; Liu, Y.-H.; Zhang, X.-W.; and Huang, K.-K. 2022. Multi-source unsupervised domain adaptation via pseudo target domain. *IEEE Transactions on Image Processing*, 31: 2122–2135.

Saenko, K.; Kulis, B.; Fritz, M.; and Darrell, T. 2010. Adapting visual category models to new domains. In *European Conference on Computer Vision*, 213–226. Springer.

Saito, K.; Watanabe, K.; Ushiku, Y.; and Harada, T. 2018. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3723–3732.

Shui, C.; Li, Z.; Li, J.; Gagné, C.; Ling, C. X.; and Wang, B. 2021. Aggregating from multiple target-shifted sources. In *International Conference on Machine Learning*, 9638–9648. PMLR.

Turrisi, R.; Flamary, R.; Rakotomamonjy, A.; and Pontil, M. 2022. Multi-source domain adaptation via weighted joint distributions optimal transport. In *Uncertainty in Artificial Intelligence*, 1970–1980.

Venkat, N.; Kundu, J. N.; Singh, D.; Revanur, A.; Venkatesh, R.; and Babu, R. V. 2020. Your classifier can secretly suffice multi-source domain adaptation. In *Advances in Neural Information Processing Systems*, volume 33, 4647–4659.

Venkateswara, H.; Eusebio, J.; Chakraborty, S.; and Panchanathan, S. 2017. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5018–5027.

Wang, Y.; Chen, H.; Heng, Q.; Hou, W.; Savvides, M.; Shinozaki, T.; Raj, B.; Wu, Z.; and Wang, J. 2022. FreeMatch: Self-adaptive thresholding for semi-supervised learning. *arXiv preprint arXiv:2205.07246*.

Wen, J.; Greiner, R.; and Schuurmans, D. 2020. Domain aggregation networks for multi-source domain adaptation. In *International Conference on Machine Learning*, 10214–10224. PMLR.

Wilson, G.; Doppa, J. R.; and Cook, D. J. 2023. Calda: Improving multi-source time series domain adaptation with contrastive adversarial learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–14.

Xu, Y.; Kan, M.; Shan, S.; and Chen, X. 2022. Mutual learning of joint and separate domain alignments for multi-source domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1890–1899.

Yuan, J.; Hou, F.; Du, Y.; Shi, Z.; Geng, X.; Fan, J.; and Rui, Y. 2022. Self-supervised graph neural network for multi-source domain adaptation. In *ACM MM*, 3907–3916.

Zhou, C.; Wang, Z.; Du, B.; and Luo, Y. 2024. Cycle Self-Refinement for Multi-Source Domain Adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI.