Thank you for the authors' detailed response. However, several of my concerns remain unaddressed:

(1) On the Motivation and Role of Optimal Similarity and Multi-Granularity Alignment.

I still have doubts about the motivation and theoretical clarity behind the use of optimal similarity and multi-granularity alignment. Based on my understanding, the "optimal similarity" component corresponds to an optimal transport plan computed using cosine similarity as the cost, while the "multi-granularity alignment" applies entropy regularization to both the cost matrix and the transport plan.

Despite the authors' efforts to clarify, the rationale for integrating these specific elements into the ROSDA framework remains insufficiently justified. For instance, Eq. (1) suggests the use of entropy-regularized OT, which makes the transport plan γ denser, while the multi-granularity alignment tends to make it more sharp. This raises a concern: is this entropy-induced density conceptually at odds with the objectives of multi-granularity alignment?

Furthermore, from a transfer learning perspective, it is still unclear how the proposed components contribute to mitigating negative transfer, especially in the context of open-set domain adaptation. The current rebuttal does not convincingly explain what makes this approach particularly suited for OSDA. Therefore, my concerns regarding the core motivation and theoretical justification remain unresolved.

(2) On the Use of SVD to Estimate the Number of Unknown Classes,

While I agree that SVD is a plausible geometric heuristic to estimate the number of unknown classes, the explanation provided through Lemma 3.2 only illustrates that the singular values of the cosine similarity matrix M are more concentrated than those of Z. This observation alone does not offer a sufficient theoretical foundation for using SVD as a means of estimating unknown class counts.

I encourage the authors to better structure their justification and provide a more rigorous explanation for why SVD is suitable in this context, potentially by connecting the spectral behavior of similarity matrices with the clustering structure of unknown categories.

Q1: **On the Motivation and Role of Optimal Similarity and Multi-Granularity Alignment._**
A1: Thanks for your detailed comments.

1. We sincerely apologize for the sign mistake in Eq. (1). The correct formulation is:

Lemma 3.1 The optimal similarity between $\boldsymbol{Z}^t$ and $\mathbf{P}$ is as follows:

$$\boldsymbol{\gamma}^* = \arg\max_{\boldsymbol{\gamma}\in\Pi(\mu,\nu)} \sum_{i,j} \boldsymbol{\gamma}_{ij} C\_ij^{cosine} + \eta H(\boldsymbol{\gamma})$$

where $\boldsymbol{C}^{\text{cosine}}$ denotes the cosine similarity between $\boldsymbol{Z}^t$ and $\mathbf{P}$, and $\boldsymbol{\gamma}^*$ is the optimal coupling matrix. The value of $\boldsymbol{\gamma}_{ij}^*$ represents the similarity of $\boldsymbol{z}_i^t$ and $\mathbf{p}_j$. η is a regularization coefficient, and $H(\boldsymbol{\gamma})$ denotes the entropy of $\boldsymbol{\gamma}$.

which is consistent with the “maximum-entropy regularization” in Sinkhorn OT [1]. **High entropy helps prevent sparse solutions and overfitting to local structures**

2. The motivations and complementary roles of the two components.

- **Optimal Similarity** assigns pseudo-labels to target common samples and distinguishes unknown samples via confidence scores, which is more robust than cosine similarity in OSDA.
- **Multi-Granularity Alignment** refines the feature space at global and local levels and does not directly construct the cost matrix or transport plan.

These components function at different stages:

- (a) The **high-entropy transport plan** in optimal similarity encourages soft matching and avoids overconfident assignments, reducing noise from unknown samples.
- (b) The **low-entropy global alignment** forms compact clusters, improving separability.
- (c) The **self-supervised cross-entropy local alignment** further aligns known-class samples, boosting their classification accuracy.

(a) supports (c) by providing accurate pseudo-labels, while (b) enhances both (a) and (c) by clarifying cluster structure. **They are complementary, not conflicting**.

Stage	Objective	EntropyLevel	Role
Optimal Similarity	Assign pseudo-labels to common samples; detect	High($\eta > 0$)	Mitigates noise and reduces class confusion

Stage	Objective	EntropyLevel	Role
	unknowns		
Global Alignment	Align over all feature distributions	Low	Forms clear cluster structures
Local Alignment	Refine class-wise alignment of common classes	None	Improve classification accuracy of known classes, enhancing known/unknown discrimination

3. Negative transfer in OSDA arises mainly from pseudo-label noise, domain shift, and label shift caused by unknown classes contamination. LASER addresses these challenges as follows:

Source of Negative Transfer	Conventional DA Approach	Additional Challenge in OSDA	Mitigation in LASER
Domain Shift	Global distribution matching	Unknown classes distort global statistics	**Global alignment** with soft indicators aligns common features while isolating unknowns
Label Shift/ Class Confusion	Class-wise alignment	Unknown classes overlap with known classes	**Optimal similarity** generates soft pseudo-labels; **local alignment** filters low-confidence samples
Pseudo-Label Noise	Iterative pseudo-label refinement	Unknown-class contamination amplifies noise	**High-entropy transport plan** avoids overconfident assignments, improving robustness
Unknown-Class Structure	Typically ignored (single “unknown”)	Internal structure of unknown classes is unused	**Singular value–based cluster estimation**+K-means for fine-grained unknown classification

In summary, optimal similarity provides robust pseudo-labels resilient to unknown-class noise,

while multi-granularity alignment enhances global and local feature separability. Their entropy settings serve distinct roles and jointly mitigate negative transfer for effective open-set adaptation.

Q2: ***On the Use of SVD to Estimate the Number of Unknown Classes***
A2: Thank you for your follow-up comment. We agree that simply stating “the singular values are more concentrated” is insufficient. Below we provide a concise justification linking spectral properties of the similarity matrix to the number of unknown classes.

1. **Separability and Spectral Rank**
   Let the feature matrix of unknown samples be $\boldsymbol{Z} \in \mathbb{R}^{N\times d}$. If there are k unknown classes and features are approximately linearly separable, the row space decomposes into k nearly orthogonal subspaces. The spectral rank

$$rank_{\epsilon_T}(\boldsymbol{Z}) \triangleq \min\left\{ r : \sum_{i>r} \sigma_i^2 \le \epsilon_T \sum_{i\ge 1} \sigma_i^2 \right\}$$

   equals the number of dominant directions (Eckart–Young–Mirsky theorem), with $\epsilon_T = 0.15$ (85% energy) offering robustness to noise [2–4].
2. **Spectral Properties of Cosine-Similarity Matrix**
   For $\boldsymbol{M} = \boldsymbol{Z}\boldsymbol{Z}^\top$, under high intra-class similarity and low inter-class similarity:

- within-class variance ≪ between-class distance (class centers nearly orthogonal)
- noise energy ≤ between-class energy

$$\boldsymbol{M} = \sum_{i=1}^{k} \lambda_i u_i u_i^\top + E, \qquad \|E\|_2 \le \delta$$

   where N_c and σ_c^2 are the sample count and within-class energy of class c, respectively, and δ bounds the noise. The first k eigenvalues λ_i are well separated from the remainder, creating a spectral gap. Thus, the number of dominant eigenvalues equals the number of unknown classes k.

3. **Connection to Graph-Cut Theory: A Spectral-Clustering View.**
   Regard $\boldsymbol{M}$ as the weighted adjacency matrix of a graph. The subspace spanned

by the top k eigenvectors of $\boldsymbol{M}$ is equivalent to the solution space of the normalized cut [5]. Also, a sharp drop between the k-th and $(k+1)$-th eigenvalue is sufficient for spectral clustering to recover the true number of clusters[6].

A response combining numerical and geometric insights is provided at here.

4. **Theoretical Guarantee.**
   Based on the above analysis, we summarize a new lemma as follows:

Lemma: Consider a similarity matrix $M \in \mathbb{R}^{n\times n}$ formed by samples from k disjoint classes.

- Ideal case.
  If intra-class similarity is constant and inter-class similarity is zero, M is block-diagonal:

$$M = \begin{bmatrix} a_1\mathbf{1}_{n_1}\mathbf{1}_{n_1}^\top & 0 & \cdots & 0 \\ 0 & a_2\mathbf{1}_{n_2}\mathbf{1}_{n_2}^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_k\mathbf{1}_{n_k}\mathbf{1}_{n_k}^\top \end{bmatrix}$$

  with $a_c \in (0,1]$. Each block is rank-1, so the total rank is k; the top k singular values are non-zero and correspond to the k classes.
- Noisy Case.
  If $M = M_0 + E$, where M_0 is the ideal block-diagonal matrix above and $\|E\|_2 \le \delta$, then by Weyl's inequality [7]

$$|\sigma_i(M) - \sigma_i(M_0)| \le \delta, \forall i$$

  so the top k singular values remain separated, reliably indicating the true class number even under perturbations.

[1] Sinkhorn Distances: Lightspeed Computation of Optimal Transport. NeurIPS 2013.
[2] The approximation of one matrix by another of lower rank. Psychometrika, 1936.
[3] Symmetric gauge functions and unitarily invariant norms. The Quarterly Journal of Mathematics, 1960.

[4] High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press, 2018.
[5] Normalized Cuts and Image Segmentation. IEEE TPAMI, 2000.
[6] A Tutorial on Spectral Clustering. Statistics and Computing, 2007.
[7] The asymptotic distribution law for the eigenvalues of linear partial differential equations. Matrix Analysis, 1985.