To Reviewer #5 **Q1:** Explain the poor results of MoCo in Tab. 4. **A1:** MoCo alone underperforms because it treats each **instance** as a single class, while the core of re-ID tasks is to encode and model **intra-/inter-class** variations. MoCo is good at unsupervised pre-training but its resulting networks need finetuning with (pseudo) class labels.

Q2: "Src. class+tgt. cluster (w/ self-paced)" v.s. "Ours (full)" in Tab. 5. **A2:** The difference is whether using un-clustered outliers. Reasons for the drop: 1) There are many un-clustered outliers ($>$ half of all samples), especially in early epochs. 2) Outliers serve as difficult samples and excluding them over-simplifies the training task. 3) The baseline doesn't update outliers in the memory, making them unsuitable to be used in pseudo classes in the later epochs.

To Reviewer #6 **Q3:** Hard to scale up? **A3:** Caching a 2048d instance needs ~ 0.05M. Our method can cache 10,000,000+ instances in 500G CPU memory. If caching in 11G GPU memory, 200,000+ instances can be easily stored.

Q4: Explain the cluster reliability criterion better. **A4:** The intuition is to measure the stability of clusters by hierarchical structures, *i.e.*, a reliable cluster should be consistent in clusters at multiple levels. It leads to evident performance gains, *i.e.*, >2% mAP gains on two tasks in Tab. 5 ("Ours w/o self-paced $\mathcal{R}_{\text{comp}}\&\mathcal{R}_{\text{indep}}$" v.s. "Ours (full)").

Q5: Relations to [13, 45]. **A5:** We discussed the differences from [13, 45] on L3-9 of supplementary material and we will further discuss their relations to our work following your advice.

To Reviewer #8 **Q6:** DukeMTMC is not available. **A6:** We added experiments on MSMT17 as suggested. For the source-domain performance on Market (Tab. 3), our method can boost the mAP by +6.3% by training with unsupervised MSMT. For the unsupervised performance (Tab. 4), we reached 19.1% mAP, outperforming 11.2% mAP of SOTA [42].

Q7: Lack of theoretical grounding. **A7:** Indeed, the effectiveness is mainly demonstrated via ablation studies in both main text and supplementary, which show significant improvements. We will look into more theories in future studies.

Q8: Difference to memory usage in MoCo [13]. **A8:** Other than centroids, we for the first time treat clusters and instances as equal classes. Our self-paced strategy dynamically determines confident clusters and un-clustered instances.

Q9: Relation to [A, B]. **A9:** We tested HDBSCAN [A] to replace our reliability criterion and observed $0.9\%/4.3\%$ mAP drops on unsupervised Market/MSMT tasks. We will further discuss earlier works and improve our method.

Q10: Hyper-parameter sensitivity and choice of clustering algorithms. **A10:** We discussed hyper-parameters in Sec. E of Appendix. We adopted DBSCAN to fairly compare with [9, 50, 47, 51] in Tab. 2. We also tested Agglomerative Clustering algorithm on unsupervised Market: 74.9% mAP by "Ours (full)" v.s. 70.4% mAP by "Ours w/o self-paced".

To Reviewer #9 **Q11:** Joint learning is not new [57, 58]. The gain is natural. **A11:** We use unified training of source classes, target clusters and target outliers, which is totally different from [57, 58]. They use multi-task learning and treat source and target class *separately* (Appendix L10-20). Naive cross-domain training would hurt the performance [10].

Q12: The form of contrastive learning is not new. **A12:** We **never** claimed that the form of contrastive learning is our novelty. We focused on exploiting all available information by jointly distinguishing different kinds of prototypes with a novel hybrid memory. We discussed the differences from previous contrastive learning methods on L92-98 (main paper) and L3-9 (Appendix). Previous methods (*e.g.*, MoCo) fail in Tab. 4. See **A1** for reasons.

Q13: The assumption of disjoint label sets is unrealistic. **A13:** Actually quite common in real-world cases. One collect annotations from city A and generalize the models to other cities. Face recognition datasets have similar phenomenon.

Q14: Why simultaneous class- and instance-level loss work? **A14:** MoCo *alone* not working on re-ID tasks doesn't imply that the proposed joint class+cluster+instance training would fail. Cluster outliers are crucial to the training (see **A2**), and treating them as single-instance classes boosts the performance significantly, given the ablation study in Tab. 5: using source class-level + only target instance-level losses ("Src. class+tgt. instance") totally fails, similar to MoCo; using source class-level + only target cluster-level losses ("Src. class+tgt. cluster (w/ self-paced)") shows inferior result.

Q15: Lack of ablation studies. **A15:** 1) "Src. class + tgt. cluster (w/o self-paced)" discards both self-paced strategy (cluster reliable criterion) and un-clustered instances from training. "Ours w/o self-paced $\mathcal{R}_{\text{comp}}\&\mathcal{R}_{\text{indep}}$" only removes self-paced strategy. 2) All the combinations of losses have been investigated in Tab. 5, *i.e.*, "Src. class", "Src. class + tgt. instance" and "Src. class + tgt. cluster". "tgt. cluster + tgt. instance" is the same as "Ours w/o source-domain data" in Tab. 4. 3) Same, as described on L79-80 of Appendix. 4) The learnable classifiers in the source domain don't match the semantic meaning of target-domain centroids and thus cause inferior performance (L142-144).

Q16: Reliability criterion is tricky and incremental. **A16:** It is meaningless to evaluate $\mathcal{R}_{\text{comp}}, \mathcal{R}_{\text{indep}}$ independently, as they complement each other and leads to over 2% mAP gain. Please see also **A4** for intuition.

Q17: Positive sample for un-clustered outlier $\boldsymbol{f}_k$. **A17:** It is $\boldsymbol{v}_k$ (L139-140) cached in the hybrid memory (Eq. (4)).

Q18: Compare to softmax/triplet loss. **A18:** Duke$\to$Market (mAP): 25.0% by cross-entropy loss, 30.1% by cross-entropy+triplet loss, 74.2% by unified contrastive+triplet loss, which are all lower than those reported (76.7%). As both cross-entropy and unified contrastive loss are variants of softmax loss, the key to success is our well-designed hybrid memory, which provides **continuous** learning targets for **dynamically** changing clusters and un-clustered instances.

Q19: The Temperature τ is sensitive. **A19:** All methods using temperature softmax function (e.g. [57, 58]) have similar effects on τ. See also Tab. 1 of [57, 58]. We set $\tau = 0.05$ following [57, 58] and achieve the best performance using the same $\tau = 0.05$ for 8 UDA tasks (Tab. 2) and 2 unsupervised tasks (Tab. 4), showing the robustness of τ=fixed 0.05.

Q20: Evaluate the clusters. **A20:** At the last epoch of Duke$\to$Market, F1 & NMI scores are: 0.82 & 0.94 (Ours full), 0.79 & 0.92 (Ours w/o self-paced), 0.73 & 0.90 (Src. class + tgt. cluster (w/ self-paced)). We will show the curves.