Object-Generalized Re-Identification: A Step Towards Universal Instance Perception

Supplementary Material

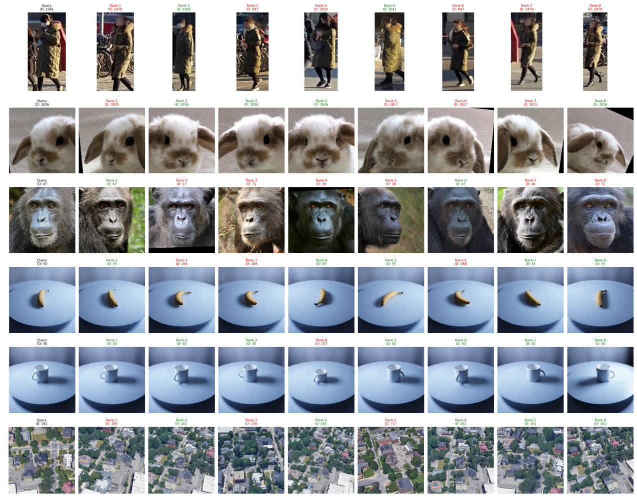
I. Evaluation Protocol

We evaluate object-generalized ReID performance across nine datasets covering diverse object categories and substantial cross-domain variations. For conventional categories such as *person* and *vehicle*, we strictly follow the original query–gallery splits provided by MSMT17 [46] and UAV-VeID [43]. For single-object datasets, such as LeopardID2022 [14] and ELPephants [18], we follow the protocol defined in [20], applying a strict leave-one-out cross-validation strategy on the test set. For geo-localization, we adopt the official split of University-1652 [62], where satellite images serve as queries and UAV images constitute the gallery. For multi-category datasets such as CUTE [19] and Wildlife71 [17], where no unified standard test split is available, we evaluate on the full dataset using leave-one-out evaluation. For PetFace [40] and GZGC [36], we similarly use all available test identities and apply the same evaluation protocol.

In traditional person ReID, evaluation focuses on cross-camera retrieval. However, in this paper, a key characteristic of our testing protocol is that MSMT17 [46] is the only dataset that provides explicit camera annotations, while University-1652 [62] is inherently a cross-view geo-localization task. All other datasets lack camera information. Consequently, for all datasets except MSMT17, we conservatively treat every image as if it were captured under a distinct camera condition. This absence of camera labels may cause images from the same physical viewpoint to be incorrectly treated as different views, leading to inflated CMC scores. To mitigate this issue and provide a more reliable measure of performance on challenging matches, we incorporate mINP (mean Inverse Negative Penalty) [50] as a core evaluation metric, as it explicitly assesses accuracy against the most difficult negative samples.

II. Qualitative Visualization Analysis

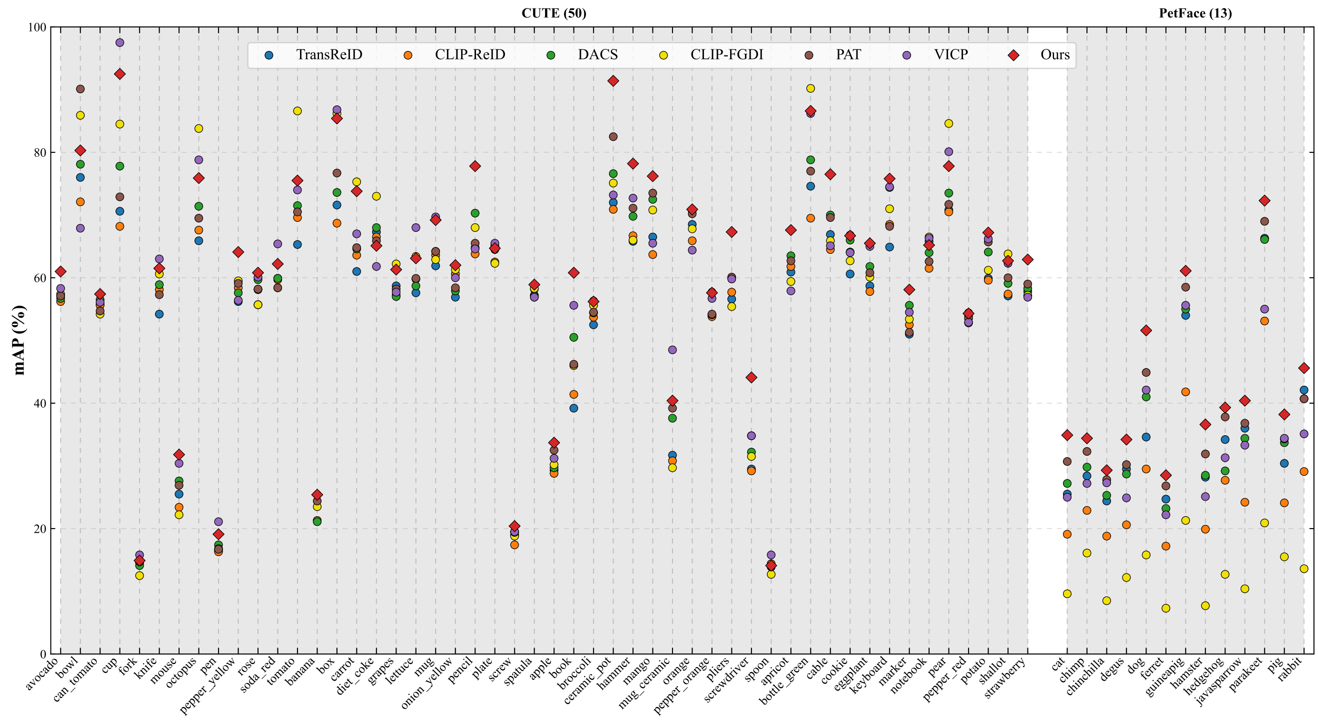
To further illustrate the characteristics and challenges of Object-Generalized ReID, we provide retrieval visualizations across a broad spectrum of object domains, including persons [46], unseen animal species [40], common objects [19], and buildings for geo-localization [62]. Compared with person ReID [51], where different identities typically exhibit substantial appearance variation, the objects considered here present far subtler inter-identity differences. Individuals of the same animal species or items of the same model often share highly similar textures, colors, and shapes, making fine-grained discrimination inherently challenging. The difficulty is further amplified by the fact that many categories in evaluation are entirely unseen during training. Consequently, the model must recognize identity cues for object types it has never encountered, without relying on class-specific priors. This necessitates learning category-agnostic, fundamentally discriminative representations that remain stable across disparate visual domains, from human appearance, to pet facial morphology, to consumer products, to architectural structures, underscoring both the difficulty and the broad applicability of the proposed Object-Generalized ReID task.



Supplementary Figure 1. Visualization of retrieval results across diverse object categories. Each row shows one query image (left) followed by its top retrieved results. Green indicates correct matches (same identity), while red denotes incorrect matches.

III. Category-Level Performance Analysis

To further examine model behavior under multi-category conditions, we conduct per-category evaluations on the CUTE and PetFace datasets. As shown in Fig. 2, we benchmark against several strong DG-ReID and supervised methods. Across both datasets, a consistent pattern emerges: existing methods exhibit substantial performance fluctuations across categories, often excelling on a subset while deteriorating notably on others. This instability indicates their dependence on category-specific appearance cues and dataset biases, which hinders reliable generalization when object semantics vary. In contrast, our method delivers uniformly strong performance across all object types, demon-



Supplementary Figure 2. Per-category evaluation in multi-category settings. Results are reported on two multi-object datasets: CUTE [19], which includes 50 object categories, and PetFace [40], which contains 13 animal categories.

Supplementary Table 1. Comparison between within-category DG ReID and our multi-object DG setting. Blue columns report results from prior DG-ReID methods under the standard single-object (person) scenario, where †indicates results trained on Market-1501 only and the remaining are trained on three person datasets (Market-1501, CUHK03, CUHK-SYSU), following the original papers. Gray columns report our reproduced results under the multi-object generalized ReID setting, where the source domain is composed of five heterogeneous object datasets.

Method	Time	within-category		cross-category	
		mAP	R1	mAP	R1
PAT †[34]	ICCV 2023	18.2	42.8	14.1	35.3
BAU [8]	NeurIPS 2024	24.3	50.9	2.7	9.1
ReNorm [35]	ECCV 2024	25.6	55.6	6.8	24.1
DACS [48]	AAAI 2024	20.3	44.2	12.5	31.8
CLIP-FGDI [58]	TIFS 2025	31.1	59.4	15.8	38.9

strating that its learned representations transfer robustly irrespective of category semantics. This category-wise stability aligns with the multi-category aggregate results reported in the main text, where the performance gap further widens. These findings confirm that our approach scales effectively with category diversity and is well-suited for open-world re-identification scenarios.

IV. Revisiting DG-ReID

We compare existing Domain Generalized ReID methods under two settings: (1) the conventional within-category DG setup restricted to the person class, and (2) our proposed mixed multi-object DG setting composed of five heterogeneous object domains. As shown in Table 1, prior DG-ReID methods perform well when both training and testing remain within the same semantic category.

However, when the source domain is expanded to include multiple semantically diverse object types, all methods suffer a dramatic performance degradation. Although the evaluation target domain remains the same, the accuracy of every model drops sharply, with mAP decreases exceeding 20% in some cases. This indicates that multi-object training introduces substantial semantic interference. Existing DG-ReID approaches are designed for domain shifts within a single semantic class and thus rely heavily on category-consistent cues that are relatively stable across person datasets. When exposed to heterogeneous objects such as vehicles, animals, vessels, and consumer items, these methods learn entangled representations that lose discriminative identity information.

This highlights a fundamental limitation of current DG-ReID methods: they fail to generalize across categories. Even when evaluated on person domain, training on multi-object sources offers little benefit and often leads to degradation. These findings validate the necessity of the proposed Object-Generalized ReID and motivate the development of models capable of learning category-agnostic identity cues rather than depending on category-specific priors.