Bridging the Sky and Ground: Towards View-Invariant Feature Learning for Aerial-Ground Person Re-Identification

Supplementary Material

8. Overview

This supplementary material is organized as follows:
- Retrieval result visualizations on the CARGO [42] and AG-ReIDv1 [28] datasets across different protocols.
- Performance evaluation of the proposed VIF method on the AG-ReIDv2 [29] dataset under cross-view protocols.
- Cross-domain evaluation for transfer learning from the CARGO dataset to AG-ReIDv1.
- Effectiveness analysis of the Multi-Patch-Level Probability (MPLP) mechanism.
- Hyper-parameter analysis of the PLRM augmentation, including θ_{max}, P_0, α, and the loss balancing hyper-parameters γ and δ.

Note: All references cited in this supplementary material correspond to the original reference list of the main paper.

9. Visualization of Retrieval Results

Figure 5 presents a visualization of retrieval results on the CARGO dataset (under four protocols) and the AG-ReIDv1 dataset (under two protocols). Our method consistently outperforms the baseline across all protocols, demonstrating its effectiveness in addressing extreme viewpoint variations. The baseline struggles with cross-view discrepancies, failing to retrieve the correct target image in Rank1 results across both datasets. In contrast, our approach captures view-invariant features, leading to more accurate identity retrieval. The visualizations further highlight the capability of the proposed VIF method to effectively handle diverse viewpoint variations, making it highly suitable for AG-ReID.

10. Performance Evaluation on AG-ReIDv2 Dataset

AG-ReIDv2 [29] dataset contains 100,502 images of 1,615 identities, captured from three types of cameras: CCTV (approximately 3 meters altitude), wearable device (approximately 1.5 meters), and UAV (15 to 45 meters altitude).

Comparison. We further evaluate our method on the larger-scale AG-ReIDv2 [29] dataset across two evaluation protocols ("A→G", "G→A"). As shown in Table 6, the proposed VIF method outperforms both the baseline and the V2E [29] approach without relying on any auxiliary information, demonstrating strong generalization capability.

Figure 5. Comparison of retrieval visualizations on the CARGO and AG-ReIDv1 datasets across multiple protocols. "P" denotes the protocol. Green and red boxes indicate correct and incorrect matches, respectively. The top five retrieval results are displayed.

Table 6. Comparison on the AG-ReIDv2 dataset under two evaluation protocols ("A→G", "G→A"). Rank1, mAP, and mINP are reported in (%) with the best results highlighted in **bold**.

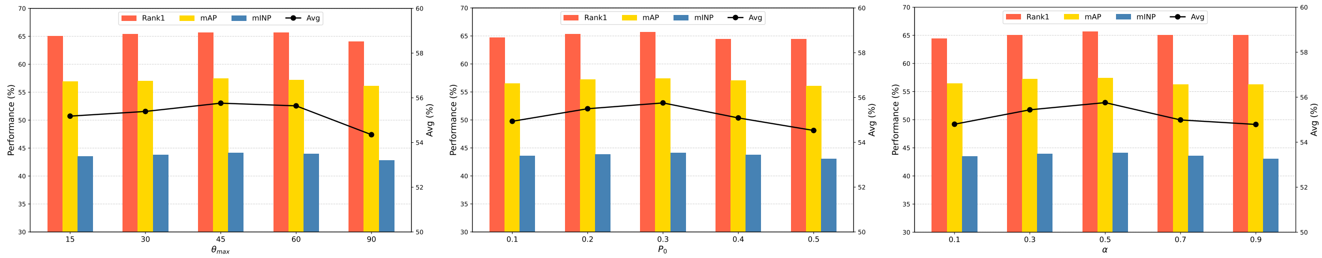
Methods	Protocol 1: A→G			Protocol 2: G→A		
	Rank1	mAP	mINP	Rank1	mAP	mINP
Baseline	85.40	77.03	55.84	84.65	75.90	47.21
V2E [29]	88.77	80.72	-	87.86	78.51	-
VIF (Ours)	**89.29**	**81.58**	**58.71**	**88.20**	**79.11**	**49.92**

Table 7. Cross-domain evaluation (%) for transfer learning from the CARGO dataset to AG-ReIDv1.

Methods	CARGO→AG-ReIDv1					
	Protocol 1: A→G			Protocol 2: G→A		
	Rank1	mAP	mINP	Rank1	mAP	mINP
ViT [7]	1.59	1.95	0.8	3.01	2.31	0.95
VDT [42]	19.33	11.81	1.63	15.38	11.73	3.38
VIF (Ours)	**23.90**	**14.92**	**3.34**	**19.54**	**14.43**	**4.85**

11. Cross-Domain Evaluation

As shown in Table 7, the cross-domain evaluation (training on the synthetic CARGO [42] dataset and testing on the real AG-ReIDv1 [28] dataset) is highly challenging. Neverthe-



(a) (b) (c)

Figure 6. Analysis of PLRM hyper-parameters under "ALL" protocol on the CARGO dataset. (a) θ_{max} is a maximum rotation angle. (b) P_0 is a base probability for each patch. (c) α is a blending factor.

Table 8. Performance of PLRM with and without the Multi-Patch-Level Probability (MPLP) mechanism on the CARGO dataset under "ALL" and "A↔G" protocols. The best results are highlighted in **bold**.

Methods	Protocol 1: ALL			Protocol 4: A↔G		
	Rank1	mAP	mINP	Rank1	mAP	mINP
Baseline	61.54	53.54	39.62	43.13	40.11	28.20
+PLRM w/o MPLP	62.18	53.73	39.88	45.63	41.15	28.92
+PLRM with MPLP	**63.14**	**54.74**	**40.90**	**48.12**	**42.95**	**30.09**

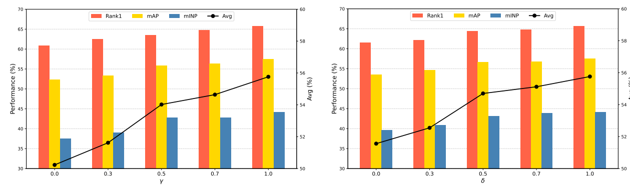


Figure 7. Analysis of balancing hyper-parameters γ and δ under "ALL" protocol on the CARGO dataset.

less, our VIF method outperforms both the baseline (ViT [7]) and VDT [42] (SOTA), further demonstrating its stability and effectiveness in learning view-invariant features.

12. The Effectiveness of the Multi-Patch-Level Probability (MPLP) Mechanism

To assess the impact of the Multi-Patch-Level Probability (MPLP) mechanism in the PLRM augmentation strategy, we conduct experiments on the CARGO dataset, as outlined in Table 8. The results demonstrate the MPLP mechanism's ability to probabilistically prioritize rotational diversity in central patches, which are typically more closely related to identity, thereby enhancing the model's robustness against viewpoint variations.

13. Hyper-parameter Analysis

13.1. Effect of PLRM Hyper-parameters

To better understand the impact of key hyper-parameters on the robustness of the PLRM augmentation, we analyze the effects of θ_{max} (Eq. 4), P_0 (Eq. 3), and α (Eq. 5) under the "ALL" protocol on the CARGO dataset, as shown in Figure 6. First, θ_{max} controls the maximum rotational range. Varying θ_{max} from $15°$ to $90°$, we find the model achieves optimal performance at $\theta_{max} = 45°$, indicating that moderate tilt rotations are beneficial for the AG-ReID task. Next, we investigate the effect of P_0 the base probability for patch transformation, by testing values from 0.1 to 0.5. The best results are achieved at $P_0 = 0.3$, suggesting that a balanced base probability promotes effective augmentation. Finally, the blending factor α governs the contribution of the original and rotated patches. After testing values ranging from 0.1 to 0.9, we find that $\alpha = 0.5$ yields the best performance, striking an ideal balance between preserving structural integrity and introducing rotational diversity.

13.2. Effect of Balancing Hyper-parameters

We evaluate the impact of the balancing hyper-parameters γ and δ (Eq. 13) on the CARGO dataset under the "ALL" protocol, as shown in Figure 7. The results indicate that the triplet and VIAL objectives contribute complementarily to enhancing discriminative and view-invariant feature learning.