DOI: https://doi.org/10.62985/j.huit_ojs.vol26.no2E.404

ORITRIP: ENHANCING PERSON RE-IDENTIFICATION WITH ORIENTATION-AWARE TRIPLET LOSS

Trinh Quoc Nguyen[1,2,*], Syahid Al Irfan[1], Oky Dicky Ardiansyah Prima[1], Tista Pal[1], Nguyen Gia Minh Thao[3,*], Dung Thi My Nguyen[4]

[1]*Iwate Prefectural University, Takizawa, Japan*

[2]*CyberCore Co., Ltd., Morioka, Japan*

[3]*Interdisciplinary Faculty of Science and Engineering, Shimane University, Matsue, Japan*

[4]*Hung Vuong University of Ho Chi Minh City, Ho Chi Minh City, Vietnam*

[*]Email: *g236v201@s.iwate-pu.ac.jp, nguyenthao@ecs.shimane-u.ac.jp*

Received: 14 December 2025; Revised: 3 April 2026; Accepted: 14 April 2026

ABSTRACT

Person re-identification faces real challenges when the same person appears in different poses and camera angles across multiple views. While standard triplet loss functions are effective for learning features that distinguish between individuals, they struggle with capturing spatial orientation details. This creates problems when the system needs to match the same person who appears at different angles or in different poses. This study presents an orientation-aware triplet loss function that incorporates human body orientation information during training. The approach builds on conventional triplet loss by adding orientation constraints that capture the geometric relationships between different poses and viewpoints of the same person. This orientation-aware mechanism can maintain consistent feature representations across varying body orientations while keeping the ability to distinguish between different individuals. Thorough evaluations on three popular person re-identification datasets showed performance gains when integrating the orientation-aware triplet loss with state-of-the-art algorithms. The results demonstrated better matching accuracy compared to methods using traditional triplet loss. In addition, the orientation-aware approach tackles key limitations in existing person re-identification systems by modeling the geometric properties of human subjects during feature learning. This research contributes to advancing person re-identification technology by providing a more effective loss function that better captures the inherent spatial characteristics of human appearance across different orientations.

Keywords: Human body orientation, metric learning, person re-identification, triplet loss, viewpoint variation.

1. INTRODUCTION

Person re-identification (Re-ID) is defined as the process of determining if an image of a person corresponds to the same identity in images from non-overlapping camera views [1]. It plays an important role in applications such as video surveillance, public safety, and access control [2], [3]. Despite progress in recent years, person Re-ID remains difficult because the same person can appear in very different ways depending on pose, body orientation, lighting, and background, as depicted in Fig. 1.

A common approach in person Re-ID is to learn discriminative features using metric learning. Among the most widely used methods is the triplet loss [4], which trains a model to bring features of the same person closer together while pushing features of different people farther apart. This approach works well when visual differences are consistent, but it struggles when body orientation changes significantly affect appearance. For example, a back-facing image and a side-facing image of the same individual may appear less similar than images of two different people in the same human body orientation if they are wearing similar clothes (Fig. 1). This limitation highlights the importance of orientation in person Re-ID. Most existing methods treat human body orientation as a side factor rather than as part of the core loss function [5], [6]. As a result, learned embeddings often fail to stay consistent across changes in body orientation.

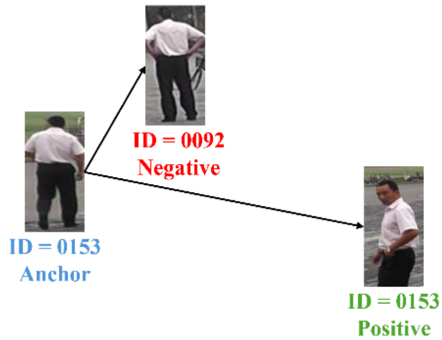

Fig. 1. Back-facing and side-facing images of the same individual with greater visual dissimilarity than images of two different individuals captured in the same body orientation wearing similar clothing.

In this paper, we propose an orientation-aware triplet loss (OriTrip), an orientation-aware triplet loss that incorporates body orientation directly into the sampling process. Instead of selecting positive and negative samples based solely on feature distance, our method also uses orientation information. The anchor is an image of a person with a given orientation. The positive is an image of the same person but with a different orientation. The negative is an image of a different person with a similar orientation. This design encourages the model to learn features that remain stable across orientations while still separating different identities.

The main novelty and contributions of this study are summarized as follows:

a. OriTrip, an orientation-aware triplet loss that integrates body orientation into the loss function, addresses a key limitation of conventional triplet loss under viewpoint changes.

b. A sampling strategy that explicitly pairs positives with different orientations and negatives with similar orientations, enforcing orientation-invariant embeddings while maintaining strong identity discriminability.

c. Comprehensive evaluation showing consistent performance gains across three popular benchmark datasets with multiple state-of-the-art person Re-ID algorithms.

2. RELATED WORK

2.1. Person Re-Identification Methods

Early research in person Re-ID relied on handcrafted descriptors such as color histograms, texture patterns, or local key-points [7], [8]. These features captured appearance but were sensitive to lighting and viewpoint changes [1]. With the growth of deep learning, convolutional neural networks (CNNs) became a more effective approach [9], [10]. CNN-based models learn global appearance representations that are more resilient to variation in clothing and environment [11], [12].

More recent methods extend this idea by focusing on local features, such as human body parts or fine-grained regions, to handle occlusion and pose differences [11], [13]. Transformer-based models have also been introduced, making use of self-attention to capture long-range dependencies and structural relationships in person images [14], [15]. These advances have improved accuracy, yet orientation and pose variation remain difficult to model directly.

2.2. Metric Learning Approaches

As a key technique in Re-ID, metric learning focuses on structuring the embedding space so that images of the same person are clustered together, and images of different people are clearly separated [16]. Contrastive loss [17], triplet loss [4], quadruplet loss [18], and circle loss [19] are common approaches. Among them, the triplet loss is widely used because it encourages both inter-class separation and intra-class compactness. It selects an anchor image, a positive image from the same identity, and a negative image from a different identity. The loss function then enforces that the anchor is closer to the positive than to the negative by a margin. Despite its success, standard triplet loss depends heavily on sample selection. It struggles when positive pairs have large orientation differences, or when negatives share similar orientations with the anchor. In these cases, embeddings may collapse or fail to generalize across camera views.

2.3. Pose and Orientation in Re-ID

To address pose and orientation variations, some works integrate pose estimation into the Re-ID pipeline [5], [6]. Pose estimation models such as OpenPose [20] or HRNet [21] provide body key-points that can guide feature alignment. Other methods divide the person into body parts and learn part-based features that are less sensitive to occlusion.

Orientation-specific approaches are less common. Some research uses orientation classifiers to assign each image a coarse orientation label, such as front, back, or side [22]. Others rely on metadata, when available, to indicate the view angle [23], [24]. These orientation cues are usually applied as auxiliary supervision or used to normalize features. However, they are not typically integrated into the loss function itself.

2.4. Gap in Literature

While metric learning and pose-based alignment have advanced the field, to the best of our knowledge, no prior work has explicitly incorporated human body orientation into the design of the triplet loss. Most methods treat orientation as a side problem rather than as part of the sampling strategy. This gap leaves embeddings vulnerable to orientation variation, which is one of the most common sources of mismatch in practical deployments.

OriTrip addresses this issue by making orientation an explicit factor in triplet construction. It pairs positives with different orientations and negatives with similar orientations to the anchor, which forces the model to learn features that stay consistent across body orientation while still separating different identities.

3. METHODOLOGY

This section describes the design of the proposed OriTrip and how it is integrated into standard person re-identification pipelines. The key idea is to preserve the overall algorithmic flow of existing methods while replacing the conventional triplet loss with the OriTrip.

Most person re-identification frameworks follow a standard structure: input images are passed through a backbone network to extract features, which are then used for metric learning with classification or triplet-based objectives [10], [25]. Our approach does not alter this pipeline. Instead, the OriTrip directly replaces the triplet loss function during training by taking into account the body orientation information, as illustrated in Fig. 2. This design choice ensures that the OriTrip can be adopted by any existing person re-identification method without requiring changes to the backbone, data preprocessing, or optimization strategy.

3.1. Human Body Orientation Estimation

To incorporate orientation information, we estimate body orientation for each training image. We adopt the High-Resolution Network (HRNet) [21] trained on the COCO-MEBOW dataset [26], a pose estimation model that provides fine-grained joint locations across the human body. We compute orientation as a continuous degree value (0°-360°), not discrete classes like Front, Back, or Side. This uses an auxiliary prediction module (HRNet) with no manual annotations. Figure 3 shows examples of computed degrees. This results in an orientation label or continuous orientation vector for each image. The orientation data is then stored alongside the image embeddings for use during triplet sampling.

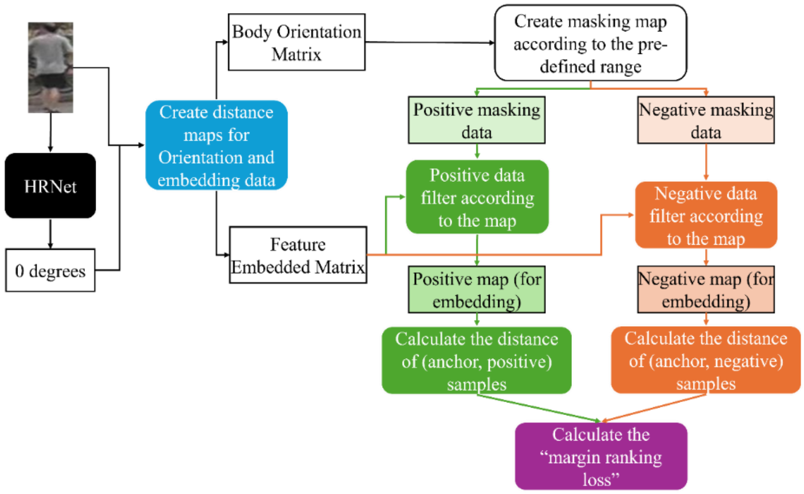

Fig. 2. Overall pipeline of proposed OriTrip.

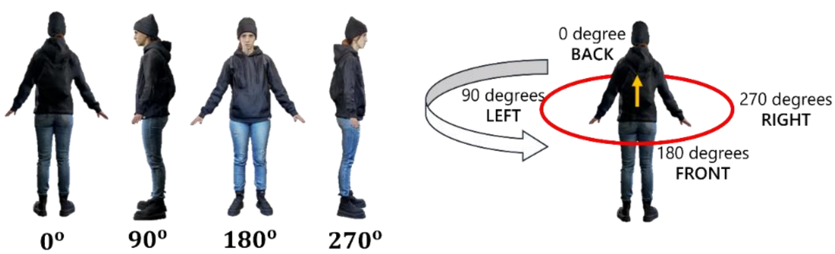

Fig. 3. Body orientation degree related to human pose.

3.2. Orientation-aware Triplet Loss

The standard triplet loss selects an anchor, a positive, and a negative sample:

- Anchor (A): an image of a target person.
- Positive (P): another image of the same person.
- Negative (N): an image of a different person.

Conventional sampling selects positive samples that are visually similar to the reference and negative samples that are visually dissimilar. It determines this by comparing embeddings. However, this does not account for the body orientation of the subject. As a result, models struggle when the same person appears in different poses or camera views.

The OriTrip modifies the sampling process by adding orientation constraints, as shown in Table 1:

- Anchor: a person's image with a given orientation.

- Positive: the same person's image but with a different orientation from the anchor image. It was calculated as the highest distance of orientation from the anchor one. This encourages the model to learn orientation-invariant features.
- Negative: a different person's image but with a similar orientation to the anchor image. It was calculated as the closest orientation distance from the anchor one. This increases the difficulty of the discrimination task and strengthens body orientation-aware feature learning.

Table 1. Comparison between original triplet loss and oritrip

Aspect	Original Triplet Loss	OriTrip (Ours)
Anchor selection	Image of a person (no orientation considered).	Image of a person with a given orientation.
Positive selection	Another image of the same person (any orientation, usually closest in feature space).	Same person image but with a different orientation (largest orientation difference from anchor).
Negative selection	Image of a different person (any orientation, usually the hardest negative in feature space).	Different person with a similar orientation to the anchor (smallest orientation difference).
Sampling strategy	Based solely on embedding distances.	Embedding distances and orientation constraints (orientation-aware sampling).
Feature learning objective	Enforce high cohesion within classes and low coupling between classes in the embedding space.	Enforce orientation-invariant embeddings while maintaining inter-class separability.
Loss function	$Loss = \max\left(0, \left[d_x\left(x_a, x_{p_a}\right) - d_x\left(x_a, x_{n_a}\right) + m\right]\right).$	Same structural form, but positives and negatives are constrained by orientation.
Limitation	Struggles when the same identity appears in very different orientations.	Better against orientation changes, and embeddings remain consistent across viewpoints.

In the OriTrip, we extend the triplet loss by adding orientation as an explicit selection criterion. Each training batch contains samples of: $\{(x_i, y_i, \theta_i)\}$, where x_i denotes the embedding vector of the sample i, y_i presents the identity label, and θ_i is the orientation. In practice, these triples can be stored as three separate lists: embeddings, labels, and orientations.

To determine positive and negative candidates, we first compare the orientation of an anchor with other samples in the batch. We used a circular metric to measure the orientation difference between the two samples of i and j:

$$d_\theta(i,j) = min\left(|\theta_i - \theta_j|, 360° - |\theta_i - \theta_j|\right), \tag{1}$$

where θ_i and θ_j are the orientations of the samples i and j, respectively. The formula keeps the angular difference within the range of $[0°, 180°]$. In practice, to compare the orientation of an anchor a with other samples in the batch, we compute $d_\theta(i,j)$ as $d_\theta(a,j)$ for each sample j. A smaller value of $d_\theta(a,j)$ indicates a higher similarity in orientation between the anchor and the sample. From there, the candidate sets for an anchor are defined as follows:

$$P(a) = \{p \mid y_p = y_a \wedge d_\theta(a,p) \geq minP\}, \tag{2}$$

$$N(a) = \{n \mid y_n \neq y_a \wedge minN \leq d_\theta(a,n) \leq maxN\}. \tag{3}$$

Here, $P(a)$ contains positive candidates (same-identity samples) where $minP$ is a threshold ensuring sufficiently different in orientation from the anchor, and $N(a)$ contains negative candidates (different-identity) samples where $[minN, maxN]$ defines an orientation range that keeps negative samples relatively similar in orientation to the anchor. The condition of $y_p = y_a$ ensures that positives belong to the same identity, while $y_n \neq y_a$ guarantees that negatives are from different identities. From these sets, the hardest positive and negative are selected based on the embedding distance $d_x(x_i, x_j)$:

$$Loss_a = \max\left(0, \left[d_x\left(x_a, x_{p_a}\right) - d_x\left(x_a, x_{n_a}\right) + m\right]\right). \tag{4}$$

All parameters are defined as follows: θ_i is orientation in degrees; $minP$, $minN$, $maxN$are thresholds for orientation differences (set to 45°, 0°, 30° in experiments); m is the fixed margin (0.3); d_x is Euclidean distance; y is identity label. Orientation similarity modulates triplet selection, not the margin m. Positives require large d_θ for invariance; negatives small d_θ for harder discrimination.

The loss formulation remains structurally similar to the original triplet loss, but the sampling strategy produces more informative triplets by exploiting orientation metadata. This ensures the network learns both identity-preserving and orientation features.

3.3. Integration with Existing Frameworks

Since the proposed OriTrip only modifies the loss function and sampling procedure, it can be directly integrated into existing frameworks without altering their architecture or training pipelines. To thoroughly evaluate generality, the OriTrip was tested with three state-of-the-art person re-identification methods:

a. CLIP-ReID [27]: a vision-language-based ReID framework.

b. SOLIDER-ReID [25]: a strong global-local feature learning model.

c. FastReID [28]: a widely used high-performance ReID framework.

In all cases, the OriTrip replaced the triplet loss function used in the baseline implementations with additional orientation data. The number of epochs was set longer for convergence. No other components of the algorithms were changed.

3.4. Training Setup

During training, each mini-batch is processed as follows:

(i) Before training, HRNet is used to prepare orientation data as metadata in the dataset.

(ii) During training per batch, images are passed through the backbone network to generate feature embeddings along with orientation as the metadata.

(iii) The triplet sampler selects anchor-positive-negative triplets using both embeddings and orientation constraints.

(iv) The OriTrip computes the loss for the selected triplets.

(v) The optimizer updates the network parameters.

This simple replacement ensures that the rest of the system, including backbones, batch construction, and optimization hyperparameters, remains identical to the baseline. The integration replaces the standard triplet loss directly in existing Re-ID frameworks with the only changes in triplet selection and loss computation, as illustrated in Fig. 2. For implementation, orientations are precomputed offline, and during each epoch, the sampler filters candidates in O(N) time per batch (where N is the batch size).

The proposed OriTrip incurs negligible additional computational cost: orientation estimation using HRNet is performed only once offline before training, producing lightweight metadata. During training, the orientation-aware sampling requires only simple circular distance calculations (Eq. (1)) and threshold-based filtering, adding minimal overhead. In our experiments, training epochs with OriTrip took less than 3% longer than the original baselines on the same hardware, ensuring comparable overall training times. Orientation estimation does not run during inference, as the model relies on learned features alone.

4. EXPERIMENTAL RESULTS

The introduced OriTrip was evaluated on three widely used person re-identification datasets: Market-1501 [29], CUHK03 [30], and MSMT17 [31]. Each dataset presents distinct challenges. Market-1501 contains images from multiple cameras with varying illumination. CUHK03 includes both detected and manually labeled bounding boxes, which test the validity of detection errors. MSMT17 is the largest and most diverse, with wide variations in lighting, backgrounds, and camera viewpoints.

4.1. Experimental Setup

The proposed OriTrip was integrated into three state-of-the-art Re-ID frameworks: CLIP-ReID [27], SOLIDER-ReID [25], Fast-ReID [28]. In each case, OriTrip directly replaced the standard triplet loss without modifying the rest of the training pipeline. This design choice allowed us to measure the direct effect of orientation-aware sampling on existing architectures.

For body orientation estimation, we employed HRNet [21], which produces reliable pose and orientation information across diverse camera views. Orientation categories were discretized into degrees, based on the estimated skeletal configuration. Performance was measured using mean Average Precision (mAP), Rank-1, Rank-5, and Rank-10 accuracies, the four most common evaluation metrics for person Re-ID. mAP measures retrieval precision across all ranks, while Rank-k accuracy represents the probability that a correct match appears within the top-k retrieved results. All models were trained and tested under the same settings as reported in their original works, except that OriTrip replaced the triplet loss component. Re-ranking was also applied during evaluation for all methods to ensure fair comparison.

4.2. Results on Market-1501

On Market-1501 [29], the OriTrip gave steady improvements across all three baselines, as presented in Table 2. For CLIP-ReID, mAP increased from 93.62% to 93.73% (+0.12% relative improvement), with Rank-1 rising from 95.04% to 95.24% (+0.21%). SOLIDER-ReID saw mAP gains from 95.62% to 95.70% (+0.08%) and Rank-1 from 96.97% to 97.09% (+0.12%). Fast-ReID improved mAP from 95.27% to 95.43% (+0.17%) and Rank-1 from 96.76% to 96.93% (+0.18%). While absolute gains were modest (average mAP +0.12% across baselines), the consistent enhancements in mAP and ranking metrics demonstrate OriTrip's reliability and added value from its orientation-aware mechanism, even with strong baseline models.

Table 2. Results on market1501

Method	Settings	mAP	R-1	R-5	R-10
CLIP-ReID	Original	93.62	95.04	97.74	98.25
	OriTrip (Ours)	93.73	95.24	98.63	99.20
SOLIDER-REID	Original	95.62	96.97	98.60	99.02
	OriTrip (Ours)	95.70	97.09	98.75	99.34
Fast-ReID	Original	95.27	96.76	98.22	98.81
	OriTrip (Ours)	95.43	96.93	98.57	98.99

Bold numbers indicate improved performance with the proposed OriTrip.

4.3. Results on CUHK03

The CUHK03 in [30] posed a different challenge, as many samples include bounding box noise. Even under these conditions, the OriTrip showed improvements, as depicted in Table 3. For CLIP-ReID, mAP increased from 82.45% to 82.64% (+0.19%) and Rank-1 from 82.93% to 83.36% (+0.43%), with larger gains at Rank-5 and Rank-10 (both +0.93%). SOLIDER-

ReID showed the largest improvement, with mAP rising from 92.15% to 93.36% (+1.21%) and Rank-1 from 91.86% to 92.57% (+0.71%). Fast-ReID also improved, with mAP increasing from 88.67% to 89.01% (+0.34%) and Rank-1 from 86.50% to 87.14% (+0.64%). Overall, the consistent gains across models indicate that OriTrip handles bounding box noise effectively by preserving positive pair consistency under spatial misalignment.

Table 3. Results on CUHK03

Method	Settings	mAP	R-1	R-5	R-10
CLIP-ReID	Original	82.45	82.93	91.21	94.64
	OriTrip (Ours)	82.64	83.36	92.14	95.57
SOLIDER-REID	Original	92.15	91.86	94.16	96.11
	OriTrip (Ours)	93.36	92.57	95.71	97.86
Fast-ReID	Original	88.67	86.50	92.14	95.93
	OriTrip (Ours)	89.01	87.14	92.64	95.86

Bold numbers indicate improved performance with the proposed OriTrip.

4.4. Results on MSMT17

The MSMT17 in [31] contained the most variation in orientation and background clutter. On this dataset, the OriTrip gave stable improvements across all three methods. On this dataset, OriTrip produced consistent improvements across all three baselines, as shown in Table 4. For Fast-ReID, mAP increased from 76.86% to 77.23% (+0.37%), with Rank-1 rising from 87.87% to 88.03% (+0.16%). CLIP-ReID also improved, with small gains in mAP (+0.01%) and Rank-1 (+0.14%). For SOLIDER-ReID, mAP increased from 86.60% to 87.10% (+0.50%) and Rank-1 from 91.80% to 92.30% (+0.50%). Although the absolute gains are modest, the consistent improvements across all methods indicate that OriTrip adapts well to large-scale datasets with frequent orientation changes.

TABLE 4. RESULTS ON MSMT17

Method	Settings	mAP	R-1	R-5	R-10
CLIP-ReID	Original	86.73	91.20	94.58	95.21
	OriTrip (Ours)	86.74	91.34	94.63	95.33
SOLIDER-REID	Original	86.60	91.80	99.57	96.20
	OriTrip (Ours)	87.10	92.30	95.70	96.20
Fast-ReID	Original	76.86	87.87	91.92	93.13
	OriTrip (Ours)	77.23	88.03	92.17	93.46

Bold numbers indicate improved performance with the proposed OriTrip.

5. CONCLUSIONS

In this research, we introduced the OriTrip, an orientation-aware triplet loss for person re-identification. By using orientation in triplet sampling, the OriTrip keeps embeddings of the same person more consistent across human orientation while separating different identities. It works as a drop-in replacement for the standard triplet loss in CLIP-ReID, SOLIDER-ReID, and Fast-ReID without changes to the backbone or training pipeline. Moreover, numerous tests on Market-1501, CUHK03, and MSMT17 showed steady gains in mAP and Rank-k accuracy, with the largest improvements on datasets with strong orientation and viewpoint changes.

Despite the above advantages, the OriTrip has some limitations as follows. Its performance depends on accurate orientation estimation, and errors in pose detection can reduce the benefit of orientation-aware sampling. In this study, the human pose estimation model was trained on the COCO-MEBOW dataset, which is not specifically designed for the person Re-ID task. This may limit its ability to capture subtle pose differences. Training the HRNet model on datasets specified for Re-ID could further improve performance.

In future work, the OriTrip can be extended to other metric learning losses such as contrastive loss, circle loss, or quadruplet loss to explore further performance gains. Applying the OriTrip to video-based person ReID and multi-modal ReID could use temporal and cross-modal consistency for improved orientation validity. Furthermore, exploring continuous orientation representations instead of discrete bins could allow the model to capture finer pose variations. Finally, integrating the OriTrip into real-world surveillance or authentication systems would provide an opportunity to evaluate its effectiveness in dynamic environments with complex viewpoints and occlusions.

Acknowledgment: This research was supported by a Grant-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (Grant number: 25K15160). The authors also thank CyberCore Co., Ltd. for providing the necessary training environments.

REFERENCES

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person Re-identification: Past, Present and Future," arXiv:1610.02984, 2016.

[2] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao and S. C. H. Hoi, "Deep Learning for Person Re-Identification: A Survey and Outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 6, pp. 2872-2893, 1 June 2022, doi: https://doi.org/10.1109/TPAMI.2021.3054775.

[3] J. Liu, Z.-J. Zha, H. Xie, Z. Xiong, and Y. Zhang, "CA^{3}Net: Contextual-Attentional Attribute-Appearance Network for Person Re-Identification," in *Proc. 26th ACM Int. Conf. Multimedia*, Seoul, Republic of Korea, 2018, pp. 737-745, doi: https://doi.org/10.1145/3240508.3240585.

[4] Y. Yuan, W. Chen, Y. Yang and Z. Wang, "In Defense of the Triplet Loss Again: Learning Robust Person Re-Identification with Fast Approximated Triplet Loss and Label Distillation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA, 2020, pp. 1454-1463, doi: https://doi.org/10.1109/CVPRW50498.2020.00185.

[5] J. Miao, Y. Wu and Y. Yang, "Identifying Visible Parts via Pose Estimation for Occluded Person Re-Identification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4624-4634, Sept. 2022, doi: https://doi.org/10.1109/TNNLS.2021.3059515.

[6] Y. Zhai, X. Han, W. Ma, X. Gou and G. Xiao, "PGMANet:Pose-Guided Mixed Attention Network for Occluded Person Re-Identification," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Shenzhen, China, 2021, pp. 1-8, doi: https://doi.org/10.1109/IJCNN52387.2021.9534442.

[7] D. Chen, Zejian Yuan, G. Hua, N. Zheng and J. Wang, "Similarity learning on an explicit polynomial kernel feature map for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, 2015, pp. 1565-1573, doi: https://doi.org/10.1109/CVPR.2015.7298764.

[8] Y. Zhang, B. Li, H. Lu, A. Irie and X. Ruan, "Sample-specific SVM learning for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 1278-1287, doi: https://doi.org/10.1109/CVPR.2016.143.

[9] T. Q. Nguyen, O. D. A. Prima, and K. Hotta, "CORE-ReID: Comprehensive optimization and refinement through ensemble fusion in domain adaptation for person re-identification," *Software*, vol. 3, no. 2, pp. 227–249, 2024, doi: https://doi.org/10.3390/software3020012.

[10] T. Q. Nguyen, O. D. A. Prima, S. A. Irfan, H. D. Purnomo, and R. Tanone, "CORE-ReID V2: Advancing the domain adaptation for object re-identification with optimized training and ensemble fusion," *AI Sens*, vol. 1, no. 4, pp. 432–455, 2025, doi: https://doi.org/10.3390/aisens1010004.

[11] V. Somers, C. D. Vleeschouwer and A. Alahi, "Body Part-Based Representation Learning for Occluded Person Re-Identification," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, 2023, pp. 1613-1623, doi: https://doi.org/10.1109/WACV56688.2023.00166.

[12] A. Ghosh, K. Shanmugalingam and W. -Y. Lin, "Relation Preserving Triplet Mining for Stabilising the Triplet Loss in Re-identification Systems," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, 2023, pp. 4829-4838, doi: https://doi.org/10.1109/WACV56688.2023.00482.

[13] M. Kim, M. Cho, H. Lee, S. Cho and S. Lee, "Occluded Person Re-Identification Via Relational Adaptive Feature Correction Learning," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Singapore, Singapore, 2022, pp. 2719-2723, doi: https://doi.org/10.1109/ICASSP43922.2022.9746734.

[14] S. He, H. Luo, P. Wang, F. Wang, H. Li and W. Jiang, "TransReID: Transformer-based object re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Montreal, QC, Canada, 2021, pp. 14993-15002, doi: https://doi.org/10.1109/ICCV48922.2021.01474.

[15] M. Sun, J. Ren and Y. Zhang, "EdgeVPR: Transformer-Based Real-Time Video Person Re-Identification at the Edge," in *Proc. IEEE 44th Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Jersey City, NJ, USA, 2024, pp. 13-24, doi: https://doi.org/10.1109/ICDCS60910.2024.00011.

[16] D. Yi, Z. Lei, S. Liao and S. Z. Li, "Deep Metric Learning for Person Re-identification," in *Proc. 22nd Int. Conf. Pattern Recognit. (ICPR)*, Stockholm, Sweden, 2014, pp. 34-39, doi: https://doi.org/10.1109/ICPR.2014.16.

[17] G. Zhang, J. Li, and Z. Ye, "Unsupervised Joint Contrastive Learning for Aerial Person Re-Identification and Remote Sensing Image Classification," *Remote Sens*, vol. 16, no. 2, 2024, doi: https://doi.org/10.3390/rs16020422.

[18] W. Chen, X. Chen, J. Zhang and K. Huang, "Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 1320-1329, doi: https://doi.org/10.1109/CVPR.2017.145.

[19] Y. Sun, C. Cheng, Y. Zhang, C. Zhang, L. Zheng, and Z. Wang, "Circle Loss: A Unified Perspective of Pair Similarity Optimization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, 2020, pp. 6397-6406, doi: https://doi.org/10.1109/CVPR42600.2020.00643.

[20] Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 172-186, 1 Jan. 2021, doi: https://doi.org/10.1109/TPAMI.2019.2929257.

[21] K. Sun, B. Xiao, D. Liu and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*

(CVPR), Long Beach, CA, USA, 2019, pp. 5686-5696, doi: https://doi.org/10.1109/CVPR.2019.00584.

[22] L. Ma, H. Liu, L. Hu, C. Wang, and Q. Sun, "Orientation Driven Bag of Appearances for Person Re-identification," ArXiv:1605.02464, 2016.

[23] S. Andavar and P. Elango, "Person re-identification based on pose angle estimation and multi-feature extraction," *IET Biometrics*, vol. 7, no. 4, pp. 365-370, 2018, doi: https://doi.org/10.1049/iet-bmt.2016.0198.

[24] X. Sun and L. Zheng, "Dissecting person re-identification from the viewpoint of viewpoint," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 608-617, doi: https://doi.org/10.1109/CVPR.2019.00070.

[25] W. Chen, X. Xu, J. Jia, H. Luo, Y. Wang, and F. Wang, "Beyond appearance: A semantic controllable self-supervised learning framework for human-centric visual tasks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Vancouver, BC, Canada, 2023, pp. 15050-15061, doi: https://doi.org/10.1109/CVPR52729.2023.01445.

[26] C. Wu, Y. Chen, J. Luo, C.-C. Su, A. Dawane, and B. Hanzra, "MEBOW: Monocular Estimation of Body Orientation in the Wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, 2020, pp. 3448-3458, doi: https://doi.org/10.1109/CVPR42600.2020.00351.

[27] S. Li, L. Sun, and Q. Li, "CLIP-ReID: Exploiting Vision-Language Model for Image Re-identification without Concrete Text Labels," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 1405-1413, doi: https://doi.org/10.1609/aaai.v37i1.25225.

[28] L. He, X. Liao, W. Liu, X. Liu, P. Cheng, and T. Mei, "FastReID: A Pytorch Toolbox for General Instance Re-identification," in *Proc. 31st ACM Int. Conf. Multimedia*, New York, NY, USA, 2023, pp. 9664–9667, doi: https://doi.org/10.1145/3581783.3613460.

[29] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang and Q. Tian, "Scalable Person Re-identification: A Benchmark," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, 2015, pp. 1116-1124, doi: https://doi.org/10.1109/ICCV.2015.133.

[30] W. Li, R. Zhao, T. Xiao and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, 2014, pp. 152-159, doi: https://doi.org/10.1109/CVPR.2014.27.

[31] L. Wei, S. Zhang, W. Gao and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 79-88, doi: https://doi.org/10.1109/CVPR.2018.00016.