ANFI: Rethinking Neighbor Feature Interaction in Person Re-ID

Xulin Li[1,2], Yan Lu[3], Bin Liu[1,2⋆], Jiaze Li[1,2], Qinhong Yang[1,2], Tao Gong[1,2], Qi Chu[1,2], and Nenghai Yu[1,2]

[1] University of Science and Technology of China, China
[2] Anhui Province Key Laboratory of Digital Security, China
[3] The Chinese University of Hong Kong, China
lxlkw@mail.ustc.edu.cn, yanlu@cuhk.edu.hk, flowice@ustc.edu.cn
{jz_li,qhyang233}@mail.ustc.edu.cn, {qchu,tgong,ynh}@ustc.edu.cn

Abstract. In person re-identification, neighbor-based methods have achieved significant success by interacting with neighbor samples to obtain more robust representations. However, existing methods rely only on affinity relations, causing their success to depend heavily on the reliability of selected neighbors. We find that affinity-only interaction often fails in challenging scenarios due to the inevitable presence of noisy neighbors. To enable effective interactions under noisy neighborhoods, we revisit neighbor-based methods under distinct reliability conditions and propose a novel **Adaptive Neighbor Feature Interaction (ANFI)** method. The core idea of ANFI is to account for negative effects from noisy neighbors, allowing samples to remain distinguishable from false positive neighbors. Unlike existing methods, ANFI models not only affinity relations but also discrepancy relations, and employs sample-wise adaptive weighting for these two types of relations. Given that capturing negative effects from noisy neighbors differs significantly from traditional relation learning, we derive discrepancy relations from a new **neighborhood similarity**, which provides more information than pairwise similarity. In addition, we propose **Noisy Relation Supervision (NRS)** to train ANFI, gradually injecting robustness to noisy relations into the model. Extensive experiments conducted under standard, cross-modal, and cross-domain settings, including comparisons with neighbor-based methods and re-ranking methods, demonstrate the superiority of our method across various neighbor distributions.

Keywords: Person Re-ID · Feature Interaction· Noisy Neighbors

1 Introduction

Person re-identification (Re-ID) aims to retrieve a specific person from a gallery set containing a large number of images captured by different cameras and scenes. Most research focuses on learning discriminative representations from single images [8,12,17,22,27,32]. However, due to the diversity of illumination, occlusion,

⋆ Corresponding author.

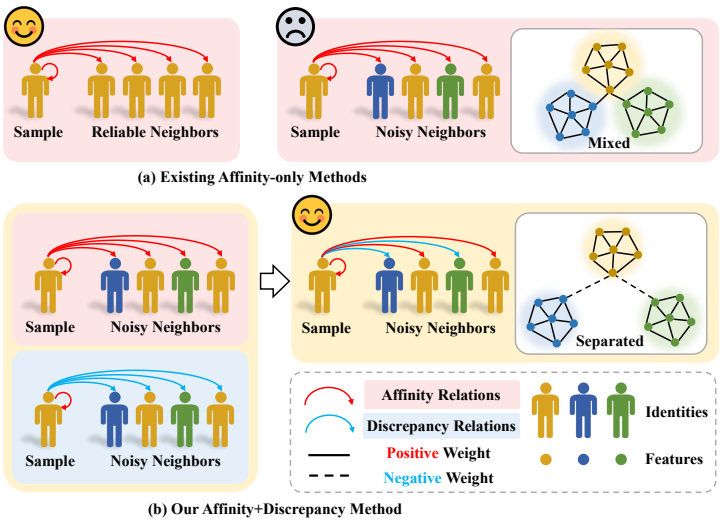

Fig. 1: (a) Existing neighbor-based approaches always aggregate neighbor representations by affinity-only relations. (b) Our approach dynamically incorporates both positive (affinity) and negative (discrepancy) relations, allowing samples to remain distinguishable from false positive neighbors.

viewpoint, and resolution, it is hard to extract sufficiently discriminative features from each image alone.

To overcome this limitation, neighbor-based methods [13,15,16,23,28,31,35] have been developed to enhance image representations that contain not only their own information but also information from multiple related images. In practice, these methods implement feature enhancement as a graph message-passing process. Specifically, each image is seen as a node in a graph, with cross-image affinity represented as the edges of this graph. By applying graph convolution or transformer layers [23], information from each image interacts with its neighbors, leading to richer and more discriminative image representations.

Unfortunately, the aforementioned scheme requires strict conditions to work well. The calculated affinity has to be reliable to ensure that information from neighbors is related and complementary to the current image, rather than conflicting. However, we observe that this condition does not always hold, especially in challenging situations, such as fuzzy query samples, small-scale gallery sets and cross-domain retrieval. In these cases, affinities are not reliable, which makes the existing methods propagate irrelevant features between nodes.

To tackle this issue, we rethink neighbor relation learning under three distinct reliability conditions and conclude our main idea as follows: 1) In the good case of reliable neighbors, affinity relations effectively reduce the distance between related samples. 2) In the bad case of unreliable neighbors, affinity relations incorrectly reduce the distance between unrelated samples. A new type of relation (discrepancy) is required to maintain the distinctions among these samples. 3) In

the intermediate case of noisy neighbors, both affinity and discrepancy relations are necessary.

To achieve our idea, we propose a novel Adaptive Neighbor Feature Interaction (ANFI) model, which is a more comprehensive neighbor relation model that goes beyond affinity-only modeling. As shown in Fig. 1, ANFI dynamically incorporates affinity (positive weight) and discrepancy (negative weight) relations, playing a role complementary to affinity and filtering out noisy propagation when faced with inaccurate neighbors. In ANFI, we employ a new neighborhood similarity for discrepancy modeling, which takes into account neighborhood structural information beyond pairwise similarity, thereby enabling effective discrimination of noisy neighbors. To train this new relational model, we propose a Noisy Relation Supervision (NRS) method that incorporates noise simulation and relational regularization. The noise simulation effectively broadens the model's applicability beyond the low-noise neighbor distribution of the training set, while relational regularization supervises model outputs toward ground-truth relational features.

Moreover, we find that existing Re-ID testing setups may not fully cover diverse neighbor distributions. A large number of simplistic gallery samples often lead to the dominance of low-noise neighbors, which can obscure the limitations of neighbor-based methods. To provide a more comprehensive evaluation, we include testing settings that cover a wider range of neighbor distributions. We compare ANFI with three categories of methods: pure image-based methods, neighbor-based methods, and re-ranking methods. For a fair evaluation, we additionally report results under the same backbone to isolate the influence of backbone feature extraction.

Our main contributions are summarized as follows:

• We investigate the challenge of learning neighbor relations in person Re-ID, particularly in handling complex scenarios with noisy neighbor relations. To address this, we propose a novel Adaptive Neighbor Feature Interaction (ANFI) model, which is the first Re-ID method to capture negative influences of noisy neighbors.

• To achieve ANFI, we design discrepancy relations with negative weights; introduce neighborhood similarity to incorporate structural information into relation modeling; and propose Noisy Relation Supervision (NRS) to enhance effective interaction among relevant features across various neighbor distributions.

• We conduct comprehensive evaluations of neighbor-based methods under various neighbor distributions. Extensive experiments on both standard and cross-domain settings demonstrate the superiority of our method over state-of-the-art methods.

2 Related Work

In this section, we first briefly review mainstream Re-ID methods that focus on learning single-image representations. Then we introduce neighbor-based methods, which model sample interactions to improve representations. Finally, we

review ranking optimization methods that leverage neighbor information to refine ranking lists.

Person Re-ID. The key to person Re-ID is extracting discriminative features. Most existing methods focus on learning better representations from single images. Some methods design stronger feature extraction networks, while others introduce effective Re-ID loss functions to improve representation learning [9, 11, 14, 21, 22, 32]. These methods fully exploit information from single images to learn discriminative features.

Neighbor-based Methods. Due to complex variations in Re-ID scenarios, such as changes in cameras, lighting, viewpoints, and occlusions, single-image feature learning methods often lead to unstable matching and sensitivity to outliers. To address this challenge, neighbor-based methods conduct interactions among neighboring images in the representation space to obtain more robust features. For instance, some methods use graph neural networks, while NFormer uses a transformer to extract additional information from k-nearest neighbors [16, 23, 28, 33, 35]. Furthermore, in visible-infrared Re-ID settings, many methods [13, 15, 31] explore cross-modal pairwise relation learning with graph networks to extract complementary information from neighbors and alleviate modality discrepancies. These methods encourage the model to cluster neighbor representations tightly, but the aggregation operation may still be constrained by false positive neighbors. In contrast, our approach enables the model to adaptively learn both similarity and discrepancy, aiming to reduce the negative impact of false positive neighbors while preserving gains from true positive neighbors.

Ranking Optimization Methods. Ranking optimization is a post-processing technique used to improve the original ranking order at test time. Recently, some approaches [1, 2, 20, 29, 38] focus on exploring neighborhood similarity from the original ranking list itself for re-ranking, with the core idea that similar samples tend to share similar neighbors. These methods operate directly on the initial ranking list and do not involve feature propagation. In contrast, our method leverages representation-level interactions to yield more discriminative features.

3 Revisiting Neighbor-based Methods and Analysis

In this section, we first review neighbor-based methods. Next, we analyze the reasons for their success in scenarios with reliable neighbor distributions. Finally, we propose a solution to deal with noisy neighbors.

3.1 Revisit of Neighbor-based Methods

As shown in Fig. 2, the pipeline of existing neighbor-based Re-ID methods can be summarized as follows:

• **Step 1: Input Feature X.** The backbone network extracts features $X = [x_1, ..., x_n]$ for all input images. Each image is considered a node in the graph, with X as node features.

• **Step 2: Pairwise Similarity** S**.** $S \in \mathbb{R}^{n\times n}$ is derived from cosine similarity [13,16]:

$$S_{ij} = \exp(\cos(\varphi(x_i), \varphi(x_j))/\tau), \tag{1}$$

where φ is a projection function and τ is a temperature parameter. Other methods compute S by Euclidean distance [28,35] ($S_{ij} = \exp(-||x_i - x_j||_2/\tau)$) or dot product [23] ($S_{ij} = \exp(\varphi_q(x_i) \cdot \varphi_k(x_j)/\tau)$).

• **Step 3: Neighbor Matrix** N**.** The k-nearest neighbors of the i-th sample are $\mathcal{N}_i = \{x_j | N_{ij} = 1\}$. The neighbor matrix $N \in \{0,1\}^{n\times n}$ is computed as follows:

$$N = \text{Topk}(S, k), \tag{2}$$

which sets the largest-k values in each row of S to 1, and sets the others to 0. In this definition, the sample itself is also included in the neighbor set $x_i \in \mathcal{N}_i$. k is a critical hyper-parameter that is kept constant for all samples, but varies depending on the dataset and the testing setups.

Some methods further refine N using reciprocal-nearest-neighbor filtering [23, 28,35]: $\hat{N}_{ij} = N_{ij}N_{ji}$ or $(N_{ij} + N_{ji})/2$.

• **Step 4: Affinity Relations** $\hat{A}$**.** $\hat{A} \in \mathbb{R}^{n\times n}$ is calculated by combining the weights from S with the connectivity from N, followed by row normalization:

$$A = S \odot N, \quad \hat{A} = D^{-1}A, \tag{3}$$

where $D \in \mathbb{R}^{n\times n}$ is the diagonal degree matrix for normalization, with $D_{ii} = \sum_j A_{ij}$.

• **Step 5: Affinity Features.** The enhanced features $F = [f_1, ..., f_n]$ are obtained via message passing:

$$F = \hat{A}\varphi(X), \tag{4}$$

where φ is a projection function.Since the matrix $\hat{A}$ inherits the sparsity of N, sample interaction is constrained within a neighborhood, which filters out the influence of numerous irrelevant samples.

• **Step 6: Relation Learning.** The enhanced features F are typically optimized with the identity loss:

$$L_{id} = -\sum\nolimits_i y_i \log(P_i), \tag{5}$$

where y_i is the identity label, $P_i = \text{Cls}(f_i)$ is the classification probability output by the classifier $\text{Cls}(\cdot)$.

3.2 Analysis of Neighbor-based Methods

To further analyze neighbor-based Re-ID methods, we rewrite their core formula as follows:

$$f_i = \sum_{x_j \in \mathcal{N}_i^+} \hat{A}_{ij}\varphi(x_j) + \sum_{x_j \in \mathcal{N}_i^-} \hat{A}_{ij}\varphi(x_j), \quad \sum_{x_j \in \mathcal{N}_i} \hat{A}_{ij} = 1, \tag{6}$$

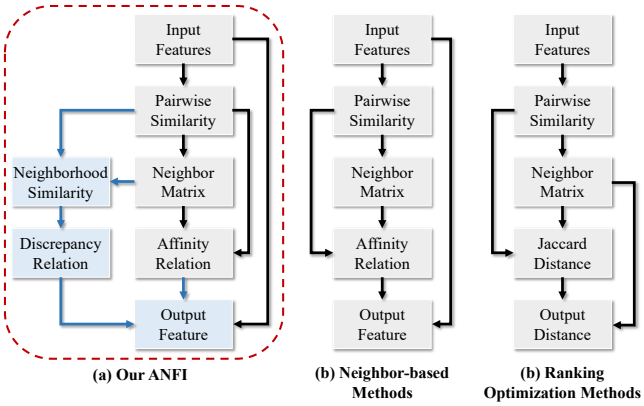

Fig. 2: Pipelines of ANFI and related methods.

where $\hat{A}_{ij} \geq 0$, $\mathcal{N}_i^+ = \{x_j | x_j \in \mathcal{N}_i, y_i = y_j\}$ is the positive neighbor set, $\mathcal{N}_i^- = \{x_j | x_j \in \mathcal{N}_i, y_i \neq y_j\}$ is the negative neighbor set.

• **Fully Reliable Neighbors (Good Case).**

When all neighbors are positive samples (i.e., $|\mathcal{N}_i^-| = 0$), the affinity relations assign positive weights to positive samples, effectively reducing the distance between positive samples.

• **Fully Unreliable Neighbors (Bad Case).**

When all non-self neighbors are negative samples (i.e., $|\mathcal{N}_i^+| = 1$ and $|\mathcal{N}_i^-| = k-1$), the affinity relations assign positive weights to negative samples, mistakenly decreasing the distance between positive and negative samples.

To ensure that the neighbor relations are suitable for this case, we need to assign negative weights to all non-self neighbors. The relational model for the bad case can be modified as follows:

$$f_i^d = \varphi'(x_i) + \sum_{x_j \in \mathcal{N}_i} (-\hat{A}_{ij}^d)\varphi'(x_j), \quad \sum_{x_j \in \mathcal{N}_i} \hat{A}_{ij}^d = 1, \tag{7}$$

where $\hat{A}^d$ is the discrepancy relation matrix. The weight assigned to itself is $(1 - \hat{A}_{ii}^d) > 0$ and the weights assigned to other neighbors are $(-\hat{A}_{ij}^d) < 0$. Eq. (7) can be rewritten as follows:

$$f_i^d = \sum_{j \neq i, x_j \in \mathcal{N}_i} \hat{A}_{ij}^d (\varphi'(x_i) - \varphi'(x_j)). \tag{8}$$

It is clear from Eq. (8) that it captures the discrepancy between samples and their neighbors; thus $\hat{A}^d$ reflects discrepancy relations.

• **Noise Neighbors (Intermediate Case).**

In practice, neighbors contain both positive and negative samples. Therefore, it is reasonable to simultaneously consider both the affinity relations $\hat{A}$ and the discrepancy relations $\hat{A}^d$. For a sample x_i, a lower $\text{nnr}(x_i)$ implies stronger

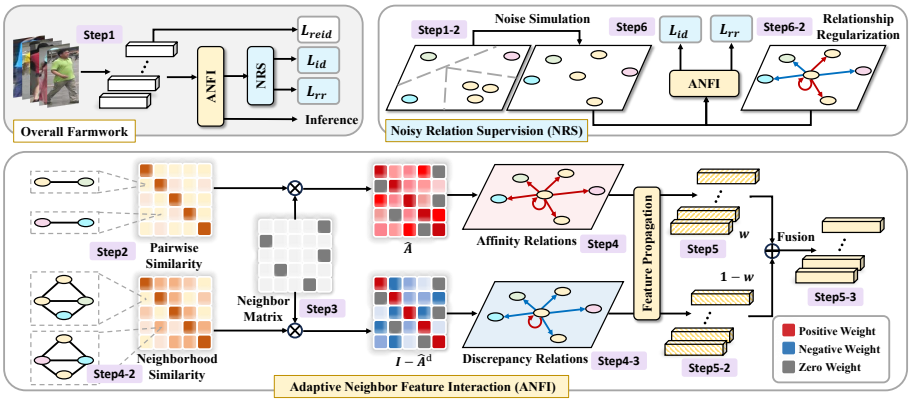

Fig. 3: The framework of the proposed Adaptive Neighbor Feature Interaction (ANFI) model achieves robust representations through interactions among multiple images.

reliance on affinity interaction, whereas a higher $\mathrm{nnr}(x_i)$ calls for stronger discrepancy interaction.

To quantify this trade-off, we define the Noise Neighbor Ratio (NNR) as a sample-level noise indicator:

$$\mathrm{nnr}(x_i) = \frac{|\mathcal{N}_i^-|}{|\mathcal{N}_i|}, \quad \mathrm{nnr}(X) = \frac{1}{n}\sum_i \mathrm{nnr}(x_i). \tag{9}$$

Therefore, $\mathrm{nnr}(x_i)$ specifies the desired mixing trend between f_i and f_i^d, while the final sample-wise mixing is learned by the model.

4 Proposed Method

As shown in Fig. 3, we introduce Adaptive Neighbor Feature Interaction (ANFI) and Noisy Relation Supervision (NRS) at both the model and training levels.

4.1 Adaptive Neighbor Feature Interaction

Based on the previous revisit and analysis, the existing methods establish affinity relations (Step 4, Eq. (3)) and utilize the established relations for feature interaction (Step 5, Eq. (4)). This affinity is applicable only to reliable neighbors, and when faced with noisy neighbors, it is necessary to consider discrepancy relations.

• **Step 4-2: Neighborhood Similarity E'.** Existing methods derive sample relations through pairwise similarity S, which is more suitable for affinity relations rather than for complex discrepancy relations. Therefore, in Step 4, we introduce neighborhood similarity $E' \in \mathbb{R}^{n\times n}$ to explore the differences between

samples by leveraging multiple third-party samples x_k. E' can be calculated as follows:

$$E'_{ij} = \sum_{x_k \in \mathcal{N}_i \cap \mathcal{N}_j} S'_{ik} \cdot S'_{jk} = \sum_k S'_{ik} \cdot S'_{jk} \cdot N_{ik} \cdot N_{jk}, \tag{10}$$

Here, third-party samples are automatically selected as shared neighbors $x_k \in \mathcal{N}_i \cap \mathcal{N}_j$; their number is determined by neighborhood overlap under k, without introducing additional hyper-parameters. where $S'_{ij} = \exp(\cos(\varphi'(x_i), \varphi'(x_j))/\tau)$ is computed with a distinct projection function φ'. E' considers information from the shared neighborhoods $\mathcal{N}_i \cap \mathcal{N}_j$ of two samples, providing neighborhood structure information beyond pairwise similarity. This approach differs from existing reciprocal nearest neighbors [23, 35], which do not consider third-party samples.

• **Step 4-3: Discrepancy Relations $\hat{A}^d$.** After obtaining the neighborhood similarity, we model the discrepancy relations as follows:

$$A^d = E' \odot N, \quad \hat{A}^d = (D^d)^{-1} A^d. \tag{11}$$

where N is the neighbor matrix, D^d is a diagonal matrix for normalization and $D^d_{ii} = \sum_j A^d_{ij}$.

• **Step 5-2: Discrepancy Features F^d.** Based on the analysis in Eq. (7), we use $\hat{A}^d$ for feature interaction:

$$F^d = (I - \hat{A}^d)\varphi'(X). \tag{12}$$

• **Step 5-3: Adaptive Mixing.** Finally, we dynamically calculate the sample-wise importance scores w_i to effectively fuse f_i and f^d_i and obtain the final representation f^{mix}_i:

$$(1 - w_i; w_i) = \text{softmax}(h(f_i); h'(f^d_i)), \tag{13}$$

$$f^{mix}_i = (1 - w_i) \cdot f_i + w_i \cdot f^d_i, \tag{14}$$

where $h(\cdot)$ and $h'(\cdot)$ are two linear layers with an output dimension of 1. For different samples, we obtain distinct values w_i so that f^{mix}_i can adaptively leverage the affinity and discrepancy relations among neighbors.

Compared with existing neighbor-based methods that focus solely on the positive influence of neighbors, ANFI simultaneously considers both affinity and discrepancy among neighbors. ANFI also provides an adaptive weighting strategy that enables our method to apply to diverse neighbor distributions, ranging from fully reliable neighbors to fully unreliable neighbors.

4.2 Noisy Relation Supervision

• **Step 1-2: Noise Simulation.** We observe that the graph constructed during training by existing methods is biased, exhibiting a low Noise Neighbor Ratio (NNR). This phenomenon is attributed to the supervision of the backbone, which leads to high similarity among intra-identity features and low similarity

among inter-identity features in the training set, ultimately resulting in a predominance of reliable neighbor samples. At test time, unseen identities typically come with noisy neighbors, making it difficult for relation models trained on reliable neighbors to generalize.

To address this issue, we add noise to the image features x_i during the training process to simulate a testing environment with a noisy neighbor distribution:

$$x_i := (1 - \alpha_n) \cdot x_i + \alpha_n \cdot \mathrm{sg}(x_j), \tag{15}$$

where $:=$ denotes feature update, $\mathrm{sg}(\cdot)$ denotes the stop-gradient operation, x_j is a randomly sampled feature in the training batch, and α_n is sampled from a uniform distribution:

$$\alpha_n \sim \mathcal{U}(0, \alpha_{max} \cdot (1 - \mathrm{nnr}(X))), \tag{16}$$

where α_{max} represents the upper limit of noise intensity, and $\mathrm{nnr}(X)$ is the batch-level Noise Neighbor Ratio computed from ground-truth identity labels using Eq. (9) before noise simulation. Therefore, cleaner batches (lower $\mathrm{nnr}(X)$) receive stronger perturbation, while noisier batches receive milder perturbation; this adaptation is used only during training. As training progresses, the noise intensity is adaptively adjusted, gradually injecting noise tolerance into the model.

• **Step 6-2: Relationship Regularization.** Furthermore, the commonly used identity loss L_{id} in existing methods only constrains the final output features of the graph to perform effective classification, without directly ensuring the accuracy of each neighbor relationship.

To learn neighbor relationships more accurately, we propose a new relationship regularization. Specifically, we set the weights of negative neighbors in the affinity matrix $\hat{A}$ to zero to obtain the ground truth affinity feature g_i. Similarly, we set the weights of positive neighbors (except the sample itself) in the discrepancy matrix $\hat{A}^d$ to zero to obtain the ground truth discrepancy feature g_i^d. After masking, we re-normalize each row of the two relation matrices so that the remaining weights sum to 1 before computing g_i and g_i^d. Finally, we replace the adaptive weights w_i with $\mathrm{nnr}(x_i)$ for each sample, yielding the final ground truth feature g_i^{mix}:

$$g_i^{mix} = (1 - \mathrm{nnr}(x_i)) \cdot g_i + \mathrm{nnr}(x_i) \cdot g_i^d. \tag{17}$$

This replacement is only used to construct supervision targets; during inference, the sample-wise mixing weights are still predicted by the adaptive mixing module. By matching f_i^{mix} to g_i^{mix} through L_{rr}, this supervision propagates the $\mathrm{nnr}(x_i)$ trend to w_i, encouraging larger w_i under noisier neighbors.

Next, we propose a relationship regularization loss to constrain the learned graph feature to be close to the ground-truth feature:

$$\mathcal{L}_{rr} = \mathbb{E}[\mathrm{D_{KL}}(\mathrm{Cls}(f_i^{mix}) || \mathrm{Cls}(g_i^{mix}))], \tag{18}$$

where $\mathrm{D_{KL}}(\cdot||\cdot)$ denotes the KL divergence, $\mathrm{Cls}(\cdot)$ represents the classifier in Eq. (5). This regularization effectively constrains the relational model to distinguish affinity and discrepancy relations.

The whole model is trained end-to-end, and the overall loss comprises the Re-ID loss [13,17] for training the backbone feature extraction network, as well as the identity loss L_{id} in Eq. (5) and the relationship regularization loss L_{rr} for training our ANFI module:

$$\mathcal{L} = \mathcal{L}_{reid}(x_i) + \mathcal{L}_{id}(f_i^{mix}) + \mathcal{L}_{rr}(f_i^{mix}, g_i^{mix}). \tag{19}$$

5 Experiments[1]

Datasets. We conduct comprehensive experiments to evaluate our method on three standard Re-ID datasets, Market1501 [36], CUHK03 [10], and MSMT17 [24], as well as two visible-infrared cross-modal Re-ID datasets, SYSU-MM01 [25] and RegDB [18]. Detailed descriptions of these datasets are provided in the supplementary materials.

Evaluation Protocol. The experiments follow existing standard evaluation settings [23, 28] and employ two assessment metrics: Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP). Following [23, 28], we eliminate interactions among different query images (single-query setup).

Implementation Details Our method is implemented using the PyTorch framework. We employ ResNet-50 [3] as our backbone network and draw upon improvements from the widely used baseline model AGW [32]. The images are resized to a fixed resolution of 384×192. We adopt random cropping, random horizontal flipping, and random erasing [39] for data augmentation. For cross-modal Re-ID, we also incorporate random channel enhancement [30], and the loss term L_{reid} in Eq. (19) is derived from the baseline provided by CIFT [13]. We use the SGD optimizer to train the model for 120 epochs. The learning rate is set to 0.01 with a warm-up scheme and cosine decay schedule. The batch size is set to 64 with 8 identities per mini-batch. We set the projection function φ and φ' to batch normalization layers. The temperature parameter τ is set to 0.4, while the noise parameter α_{max} is set to 0.4.

5.1 Comparison with SOTA methods

We compare ANFI with representative state-of-the-art (SOTA) methods on Market1501, CUHK03, MSMT17, SYSU-MM01, and RegDB in Tab. 1.

Results on Standard Re-ID Datasets. As shown in the left block of Tab. 1, ANFI achieves the best average mAP and rank-1 accuracy across the three standard Re-ID datasets. Specifically, the mAP of ANFI surpasses that of the single-image representation method Instruct-ReID [5] by 2.2% / 5.3% / 10.5% on Market1501 / CUHK03 / MSMT17, respectively. The gain is larger on mAP than on rank-1, indicating that ANFI improves overall ranking quality rather than only top retrieval.

[1] Hyperparameter analysis, adaptive weight analysis, and more visualizations are provided in the supplementary materials.

Table 1: Comparison of rank-1 (R1) and mAP (%) with representative SOTA methods on Market1501, CUHK03, MSMT17, SYSU-MM01 and RegDB datasets.

Method	Market1501		CUHK03		MSMT17		Method	SYSU-MM01		RegDB	
	mAP	R1	mAP	R1	mAP	R1		mAP	R1	mAP	R1
CAL [19]	89.5	95.5	-	-	64.0	84.2	CAJ [30]	66.9	69.9	78.5	84.9
CLIP-ReID [9]	89.6	95.5	81.6	80.9	73.4	88.7	MPANet [26]	68.2	70.6	80.8	83.3
Instruct-ReID [5]	93.5	96.5	85.4	86.5	72.4	86.9	DEEN [34]	71.8	74.7	84.3	90.3
VersReID [37]	93.2	96.8	-	-	74.2	88.8	PartMix [7]	74.6	77.8	82.4	85.3
CA-Jaccard [1]	94.5	96.2	-	-	74.1	86.2	DNS [6]	74.4	77.3	88.4	93.3
NFormer [23]	93.0	95.7	76.4	79.0	62.2	80.8	SAAI [2]	77.0	75.9	91.8	91.6
GCR [35]	95.1	96.6	89.0	89.7	76.9	86.8	cm-SSFT [15]	54.1	47.7	64.9	64.6
Cheb-GR [28]	93.2	95.2	85.4	83.9	-	-	CIFT [13]	74.8	74.1	91.4	91.2
ANFI (Ours)	**95.7**	**96.8**	**90.7**	**91.3**	**82.9**	**90.3**	ANFI (Ours)	**81.1**	**80.3**	**94.4**	**93.7**

Table 2: Comparison under the same backbone. For fair evaluation, we report mAP (%) for backbone-only features ('Raw') and backbone + gallery-based methods.

Dataset	Backbone	Raw	K-recip. [38]	CA-Jacc. [1]	Cheb-GR [28]	GCR [35]	ANFI
CUHK03	TransReID [4]	75.4	86.4	86.9	85.4	84.0	**88.1**
MSMT17	CLIP-ReID [9]	73.4	84.1	86.0	83.3	83.5	**86.5**
SYSU-MM01	DEEN [34]	71.8	77.2	77.2	75.8	76.3	**79.3**

ANFI also outperforms the neighbor-based methods Cheb-GR [28] and GCR [35] in terms of mAP on these datasets. Compared with the re-ranking method CA-Jaccard [1], ANFI also achieves better mAP.

Results on Cross-modal Re-ID Datasets. As shown in the right block of Tab. 1, ANFI outperforms the neighbor-based methods cm-SSFT [15] and CIFT [13], as well as the re-ranking method SAAI [2]. Specifically, on SYSU-MM01 / RegDB, ANFI improves mAP by 4.1% / 2.6% over SAAI [2]. Compared with CIFT [13], the gains are 6.3% / 3.0%, respectively. Given the large modality gap in cross-modal retrieval, these gains indicate stronger robustness under noisier neighbor distributions.

5.2 Comparison Under the Same Backbone

To isolate the effect of backbone feature extraction and provide a fair apples-to-apples comparison among gallery-based methods, we further report comparisons under the same backbone. For each dataset-backbone pair, we compare K-reciprocal [38], CA-Jaccard [1], Cheb-GR [28], GCR [35], and ANFI (ours) on top of the same backbone features while using gallery information at inference. This protocol also verifies that ANFI can be flexibly plugged into different single-image backbones.

As shown in Tab. 2, ANFI achieves the best mAP on all datasets. Compared with the strongest baseline in each row, ANFI further improves mAP by 1.2%

Table 3: The Δ mAP performance of Re-ID methods on different neighbor distributions. 'n-shot' means that the average number of gallery images per identity is n. 'M→C' and 'M→S' denote cross-domain testing from Market1501 to CUHK03 and from Market1501 to SYSU-MM01, respectively.

Method	Market1501					Cross-Domain	
	Standard	1-shot	2-shot	Hard Query	Poor Model	M→C	M→S
Ranking Optimization Methods							
K-recip. [38]	+4.8	-0.1	+1.6	-1.5	+0.8	+2.8	-2.2
ECN [20]	+4.5	-1.5	+1.4	-2.1	-0.2	+2.2	-3.0
SAAI [2]	+4.0	-0.9	+1.5	-0.8	+0.6	+2.4	-1.9
CA-Jacc. [1]	+4.9	-0.3	+1.0	-1.9	+0.9	+2.7	-2.2
Neighbor-based Methods							
SFT [16]	+1.8	+0.1	+0.3	-0.3	+0.3	+1.9	-0.8
NFormer [23]	+1.5	-0.8	-0.1	-1.8	+0.2	+1.3	-2.5
GCR [35]	+4.7	-1.0	+1.4	-1.4	+0.5	+1.9	-1.6
Cheb-GR [28]	+3.8	-0.7	+1.1	-1.0	+0.5	+1.6	-1.9
ANFI (Ours)	**+5.4**	**+0.9**	**+2.3**	**+0.6**	**+1.1**	**+3.5**	**+0.2**

/ 0.5% / 2.1% on CUHK03 / MSMT17 / SYSU-MM01, respectively. We attribute this to the fact that K-reciprocal [38], CA-Jaccard [1], Cheb-GR [28], and GCR [35] mainly rely on affinity-based neighbor interaction, which is more sensitive to noisy neighbors, while ANFI jointly models affinity and discrepancy relations for more robust gallery-based interaction.

5.3 Comprehensive Testing Benchmark

To facilitate a comprehensive and fair evaluation of neighbor-based methods and ranking optimization methods, we conduct an in-depth investigation under varying neighbor distributions, as shown in Tab. 3. The benchmark settings are as follows:

1) In common Re-ID tests, reliable neighbors dominate, which can obscure the inadequacies of affinity-only relations. Following recent neighbor-based methods [23,28], we conduct **cross-domain** testing and **small-scale gallery** testing to encompass a broader range of neighbor cases, particularly noisy neighbor distributions. For small-scale gallery testing, we sample several gallery subsets, each containing n images per identity, referred to as 'n-shot' testing. Additionally, we conduct tests involving 'Hard Query' and 'Poor Model'. 'Hard Query' denotes the 10% query samples with the lowest Rank-1 accuracy under the baseline model; this subset is selected using the baseline only, avoiding circular evaluation. 'Poor Model' is simulated by prematurely terminating model training (30/120 epochs), thereby representing a weaker baseline.

2) Neighbor-based methods are plug-and-play modules, while ranking optimization methods are training-free. To ensure a fair comparison, we use the same baseline model for all approaches to extract single-image representations and report the performance variations (Δ mAP) relative to this baseline.

Table 4: Ablation study on each component of ANFI on Market1501 under 'Standard', '1-shot', and 'Hard Query' test modes. mAP accuracy (%) is reported. The terms 'Affi' and 'Disc' refer to affinity and discrepancy relations.

i	ANFI		NRS		Standard	1-shot	Hard Query
	Affi.	Disc.	Noise	L_{rr}			
1					90.3	93.5	65.5
2	✓				94.2	92.5	63.1
3	✓	✓			94.8	93.5	65.3
4	✓	✓	✓		95.1	93.8	65.9
5	✓	✓	✓	✓	95.7	94.4	66.1

3) Due to substantial differences in neighbor distributions across testing modes, we adjust neighbor-related hyper-parameters for all methods in each configuration, e.g., k in ANFI, k_1 and k_2 in K-reciprocal [38], and λ in Cheb-GR [28].

Based on the experimental results, the following conclusions can be drawn:
1) When the number of images per identity in the test set is low, such as 1 or 2, a significant number of false positive neighbors may arise. Conducting relation learning in such a noisy environment is extremely challenging, and most existing methods tend to reduce retrieval performance. This reveals the shortcomings of relying solely on affinity relations.
2) Moreover, challenging query samples can lead to significantly noisier neighbors. Because this subset is relatively small, its impact is difficult to reflect in standard tests. Compared with same-task baseline degradation ('Poor Model'), cross-domain transfer shows asymmetric behavior: 'M→S' causes clear performance drops for most methods, while 'M→C' usually still yields gains but with limited margins.
3) Our ANFI method effectively handles noisy relations and achieves the largest gains across all seven settings. ANFI is the only method with positive Δ mAP in all seven settings (+5.4/+0.9/+2.3/+0.6/+1.1/+3.5/+0.2), while other methods show negative Δ mAP in 2–4 settings. SFT [16] remains positive in 5/7 settings, but its gains are smaller than ANFI. This indicates that discrepancy relation modeling complements affinity relations and improves robustness under noisy neighbors.
4) Moreover, our ANFI method remains effective under reliable relations, achieving a 5.4% improvement in mAP that surpasses other methods in the standard test mode on Market1501.

5.4 Ablation Study

To evaluate the contribution of each component in ANFI, we conduct ablation experiments on Market1501 under 'Standard', '1-shot', and 'Hard Query' test modes, and report mAP in Tab. 4.

Effectiveness of ANFI. As shown in the 2^{nd} row, applying affinity relations improves mAP by 3.9%. This indicates that affinity relations among reliable

neighbors can outperform single-image representations. However, under the '1-shot' and 'Hard Query' test modes, noisy neighbors cause mAP drops of 1.0% and 2.4%, respectively. As illustrated in the 3^{rd} row, introducing discrepancy relations enhances robustness across different neighbor distributions. Compared with using only affinity relations, mAP increases by 1.0% and 2.2% under the '1-shot' and 'Hard Query' test modes.

Effectiveness of NRS. As shown in the 4^{th} and 5^{th} rows, we analyze the effects of relationship regularization L_{rr} and noise simulation in NRS. Experimental results indicate that both components facilitate noisy-relation learning and improve retrieval performance, with total gains of 0.9%, 0.9%, and 0.8% under the 'Standard', '1-shot', and 'Hard Query' modes. With NRS, the model reaches the best result, i.e., 95.7% mAP on Market1501. Most importantly, with ANFI and NRS, the model still achieves improvements in challenging scenarios where there is at most one non-self positive sample among neighbors.

These ablation experiments demonstrate that ANFI and NRS are complementary and together provide robust multi-image relational learning.

5.5 Efficiency Analysis

To provide a controlled efficiency comparison, all gallery-based methods are evaluated with the same backbone features, runtime environment, and hardware under a unified PyTorch implementation with shared acceleration strategies, including GPU tensor computation, chunked evaluation to avoid dense full-matrix bottlenecks, and optional Faiss-based neighbor search. All experiments in this section are conducted on a server with dual Intel(R) Xeon(R) Gold 5220R CPUs and NVIDIA GeForce RTX 3090 GPUs, and we report the fastest stable implementation for each method under this unified setting.

As shown in Tab. 5, the additional latency of gallery-based methods remains moderate compared with the shared raw-backbone extraction cost. To reduce small run-to-run fluctuations in feature extraction, the reported total time uses a shared backbone extraction baseline plus the measured method-specific evaluation time. On Market1501, all methods incur only limited extra cost over the backbone baseline. On MSMT17, the relative ranking is similar, while the larger gallery size makes the latency differences more visible. Among the compared methods, ANFI keeps competitive practical runtime while explicitly modeling discrepancy relations. Cheb-GR shows a much larger overhead on MSMT17 because its row-wise adaptive threshold graph retains substantially more effective edges than kNN-style methods on the large-scale gallery, making graph propagation much denser in practice.

Table 5: Practical end-to-end testing time (s) under the standard all-query protocol. 'Extra' denotes additional time over the raw backbone. For each method, we report the fastest stable implementation among the tested variants.

Method	Market1501			MSMT17		
	Accel	Time	Extra	Accel	Time	Extra
Backbone	cuda	24.5	+0.0	cuda	145.7	+0.0
K-reciprocal [38]	cuda-faiss	31.6	+7.1	cuda-faiss	184.4	+38.7
CA-Jaccard [1]	cuda-faiss	28.2	+3.7	cuda-faiss	167.2	+21.5
ECN [20]	cuda	25.0	+0.5	cuda-faiss	147.5	+1.7
SFT [16]	cuda	24.6	+0.1	cuda	148.5	+2.8
AIM [2]	cuda	24.7	+0.2	cuda	150.8	+5.0
NFormer [23]	cuda-sparse	25.2	+0.7	cuda-sparse	148.6	+2.9
GCR [35]	cuda-sparse	25.2	+0.7	cuda-sparse	159.3	+13.6
Cheb-GR [28]	cuda-dense	25.3	+0.8	cuda-sparse	361.8	+216.1
ANFI (Ours)	cuda-sparse	24.8	+0.3	cuda-sparse	148.5	+2.7

6 Conclusion

In this paper, we revisit neighbor-based Re-ID under varying neighbor reliability and show that affinity-only interaction is insufficient in the presence of noisy neighbors. Based on this analysis, we propose Adaptive Neighbor Feature Interaction (ANFI), which jointly models affinity and discrepancy relations and adaptively mixes them for each sample. We further introduce neighborhood similarity for discrepancy construction and Noisy Relation Supervision (NRS) to improve robustness under noisy neighbor distributions during training. Extensive experiments across standard, cross-modal, and cross-domain settings, including comparisons with pure image-based, neighbor-based, and re-ranking methods, validate the effectiveness and robustness of ANFI.

Ethics Statement Person Re-ID can support public-safety and forensic search, but also has dual-use risks such as unauthorized tracking, mass surveillance, privacy invasion, and harms from misidentification. This work uses public benchmarks and does not collect new personal data or deploy a surveillance system. Real-world use should require legal authorization, data minimization, access control, auditing, and bias assessment.

Acknowledgements This work is supported by the National Natural Science Foundation of China (Grant No. 62272430).

References

1. Chen, Y., Fan, Z., Chen, Z., Zhu, Y.: Ca-jaccard: camera-aware jaccard distance for person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 17532–17541 (2024)

2. Fang, X., Yang, Y., Fu, Y.: Visible-infrared person re-identification via semantic alignment and affinity inference. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11270–11279 (2023)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
4. He, S., Luo, H., Wang, P., Wang, F., Li, H., Jiang, W.: Transreid: Transformer-based object re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 15013–15022 (2021)
5. He, W., Deng, Y., Tang, S., Chen, Q., Xie, Q., Wang, Y., Bai, L., Zhu, F., Zhao, R., Ouyang, W., et al.: Instruct-reid: A multi-purpose person re-identification task with instructions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17521–17531 (2024)
6. Jiang, Y., Cheng, X., Yu, H., Liu, X., Chen, H., Zhao, G.: Domain shifting: A generalized solution for heterogeneous cross-modality person re-identification. In: European Conference on Computer Vision. pp. 289–306. Springer (2024)
7. Kim, M., Kim, S., Park, J., Park, S., Sohn, K.: Partmix: Regularization strategy to learn part discovery for visible-infrared person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 18621–18632 (2023)
8. Li, C., Chen, S., Ye, M.: Adaptive high-frequency transformer for diverse wildlife re-identification. In: European Conference on Computer Vision. pp. 296–313. Springer (2024)
9. Li, S., Sun, L., Li, Q.: Clip-reid: exploiting vision-language model for image re-identification without concrete text labels. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 1405–1413 (2023)
10. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 152–159 (2014)
11. Li, X., Lu, Y., Liu, B., Li, J., Liu, Y., Chu, Q., Ye, M., Ouyang, W., Yu, N.: Causal clothes-invariant feature learning for cloth-changing person re-id. IEEE Transactions on Circuits and Systems for Video Technology (2026)
12. Li, X., Lu, Y., Liu, B., Li, J., Yang, Q., Gong, T., Chu, Q., Ye, M., Yu, N.: Towards anytime retrieval: A benchmark for anytime person re-identification. In: Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25. pp. 1467–1475. International Joint Conferences on Artificial Intelligence Organization (2025)
13. Li, X., Lu, Y., Liu, B., Liu, Y., Yin, G., Chu, Q., Huang, J., Zhu, F., Zhao, R., Yu, N.: Counterfactual intervention feature transfer for visible-infrared person re-identification. In: European conference on computer vision. pp. 381–398. Springer (2022)
14. Li, X., Lu, Y., Liu, B., Yang, Q., Chu, Q., Gong, T., Yu, N.: Mfen: Multi-frequency expert network for visible-infrared person re-id. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18471–18480 (2026)
15. Lu, Y., Wu, Y., Liu, B., Zhang, T., Li, B., Chu, Q., Yu, N.: Cross-modality person re-identification with shared-specific feature transfer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13379–13389 (2020)
16. Luo, C., Chen, Y., Wang, N., Zhang, Z.: Spectral feature transformation for person re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4976–4985 (2019)

17. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of tricks and a strong baseline for deep person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 0–0 (2019)
18. Nguyen, D.T., Hong, H.G., Kim, K.W., Park, K.R.: Person recognition system based on a combination of body images from visible light and thermal cameras. Sensors **17**(3), 605 (2017)
19. Rao, Y., Chen, G., Lu, J., Zhou, J.: Counterfactual attention learning for fine-grained visual categorization and re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1025–1034 (2021)
20. Sarfraz, M.S., Schumann, A., Eberle, A., Stiefelhagen, R.: A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 420–429 (2018)
21. Sun, Y., Cheng, C., Zhang, Y., Zhang, C., Zheng, L., Wang, Z., Wei, Y.: Circle loss: A unified perspective of pair similarity optimization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6398–6407 (2020)
22. Sun, Y., Zheng, L., Yang, Y., Tian, Q., Wang, S.: Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: Proceedings of the European conference on computer vision (ECCV). pp. 480–496 (2018)
23. Wang, H., Shen, J., Liu, Y., Gao, Y., Gavves, E.: Nformer: Robust person re-identification with neighbor transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7297–7307 (2022)
24. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 79–88 (2018)
25. Wu, A., Zheng, W.S., Yu, H.X., Gong, S., Lai, J.: Rgb-infrared cross-modality person re-identification. In: Proceedings of the IEEE international conference on computer vision. pp. 5380–5389 (2017)
26. Wu, Q., Dai, P., Chen, J., Lin, C.W., Wu, Y., Huang, F., Zhong, B., Ji, R.: Discover cross-modality nuances for visible-infrared person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4330–4339 (2021)
27. Yang, B., Chen, J., Ye, M.: Towards grand unified representation learning for unsupervised visible-infrared person re-identification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11069–11079 (2023)
28. Yang, J., Li, H., Du, B., Ye, M.: Cheb-gr: Rethinking k-nearest neighbor search in re-ranking for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19261–19270 (2025)
29. Ye, M., Liang, C., Yu, Y., Wang, Z., Leng, Q., Xiao, C., Chen, J., Hu, R.: Person reidentification via ranking aggregation of similarity pulling and dissimilarity pushing. IEEE Transactions on Multimedia **18**(12), 2553–2566 (2016)
30. Ye, M., Ruan, W., Du, B., Shou, M.Z.: Channel augmented joint learning for visible-infrared recognition. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 13567–13576 (2021)
31. Ye, M., Shen, J., J. Crandall, D., Shao, L., Luo, J.: Dynamic dual-attentive aggregation learning for visible-infrared person re-identification. In: European conference on computer vision. pp. 229–247. Springer (2020)

32. Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., Hoi, S.C.: Deep learning for person re-identification: A survey and outlook. IEEE transactions on pattern analysis and machine intelligence **44**(6), 2872–2893 (2021)
33. Zhang, X., Jiang, M., Zheng, Z., Tan, X., Ding, E., Yang, Y.: Understanding image retrieval re-ranking: A graph neural network perspective. arXiv preprint arXiv:2012.07620 (2020)
34. Zhang, Y., Wang, H.: Diverse embedding expansion network and low-light cross-modality benchmark for visible-infrared person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2153–2162 (2023)
35. Zhang, Y., Qian, Q., Wang, H., Liu, C., Chen, W., Wang, F.: Graph convolution based efficient re-ranking for visual retrieval. IEEE Transactions on Multimedia **26**, 1089–1101 (2023)
36. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: Proceedings of the IEEE international conference on computer vision. pp. 1116–1124 (2015)
37. Zheng, W.S., Yan, J., Peng, Y.X.: A versatile framework for multi-scene person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(3), 1362–1380 (2024)
38. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1318–1327 (2017)
39. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 13001–13008 (2020)