A Re-Ranking Method Using K-Nearest Weighted Fusion for Person Re-Identification*

Quang-Huy Che[1,2], Le-Chuong Nguyen[1,2], Gia-Nghia Tran[1,2], Dinh-Duy Phan[1,2] and Vinh-Tiep Nguyen[1,2]

[1]*University of Information Technology, Ho Chi Minh City, Vietnam*

[2]*Vietnam National University, Ho Chi Minh City, Vietnam*

huycq@uit.edu.vn, 21520655@gm.uit.edu.vn,{nghiatg, duypd, tiepnv}@uit.edu.vn

Keywords: Re-Ranking, Person Re-Identification, Multi-View Fusion, K-Nearest Weighted Fusion.

Abstract: In person re-identification, re-ranking is a crucial step to enhance the overall accuracy by refining the initial ranking of retrieved results. Previous studies have mainly focused on features from single-view images, which can cause view bias and issues like pose variation, viewpoint changes, and occlusions. Using multi-view features to present a person can help reduce view bias. In this work, we present an efficient re-ranking method that generates multi-view features by aggregating neighbors' features using *K-nearest Weighted Fusion* (*KWF*) method. Specifically, we hypothesize that features extracted from re-identification models are highly similar when representing the same identity. Thus, we select **K** neighboring features in an unsupervised manner to generate multi-view features. Additionally, this study explores the weight selection strategies during feature aggregation, allowing us to identify an effective strategy. Our re-ranking approach does not require model fine-tuning or extra annotations, making it applicable to large-scale datasets. We evaluate our method on the person re-identification datasets Market1501, MSMT17, and Occluded-DukeMTMC. The results show that our method significantly improves Rank@1 and mAP when re-ranking the top **M** candidates from the initial ranking results. Specifically, compared to the initial results, our re-ranking method achieves improvements of **9.8%/22.0%** in Rank@1 on the challenging datasets: MSMT17 and Occluded-DukeMTMC, respectively. Furthermore, our approach demonstrates substantial enhancements in computational efficiency compared to other re-ranking methods.

1 INTRODUCTION

Person re-identification (ReID) (Luo et al., 2019b; Wieczorek et al., 2021; Ni et al., 2023; Chen et al., 2023; Somers et al., 2023; Li et al., 2023) is a computer vision task that involves recognizing and matching a person across multiple images or video frames from different cameras. The goal is to reidentify a person despite variations in pose, lighting, camera views, and occlusion. In a typical ReID system, given a query image, the system retrieves and ranks candidate images from a gallery based on their similarity to the query. Initial rankings are based on features from deep learning models and distance metrics. Re-ranking refines this list to improve re-identification accuracy, aiming to give higher ranks to relevant images.

Re-ranking is a powerful technique widely em-

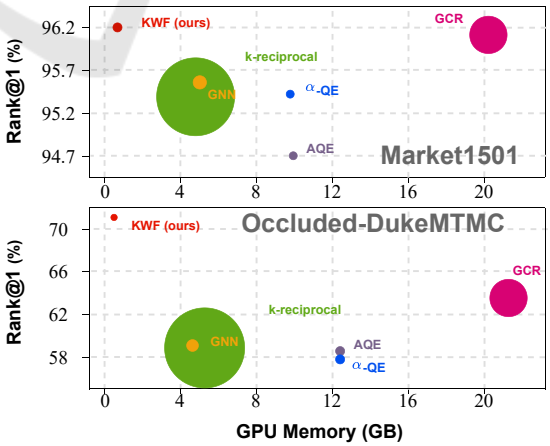


Figure 1: Comparing the computational cost and Rank@1 performance of various reranking methods on the Market1501 and Occluded-DukeMTMC datasets. The y-axis shows Rank@1, the x-axis represents GPU memory usage, and the circle size indicates evaluation time, with larger circles representing longer evaluation times.

* First two authors contribute equally and the fifth is the corresponding author.

Che, Q.-H., Nguyen, L.-C., Tran, G.-N., Phan, D.-D. and Nguyen, V.-T.
A Re-Ranking Method Using K-Nearest Weighted Fusion for Person Re-Identification.
DOI: 10.5220/0013176100003905

In *Proceedings of the 14th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2025)*, pages 79-90
ISBN: 978-989-758-730-6; ISSN: 2184-4313

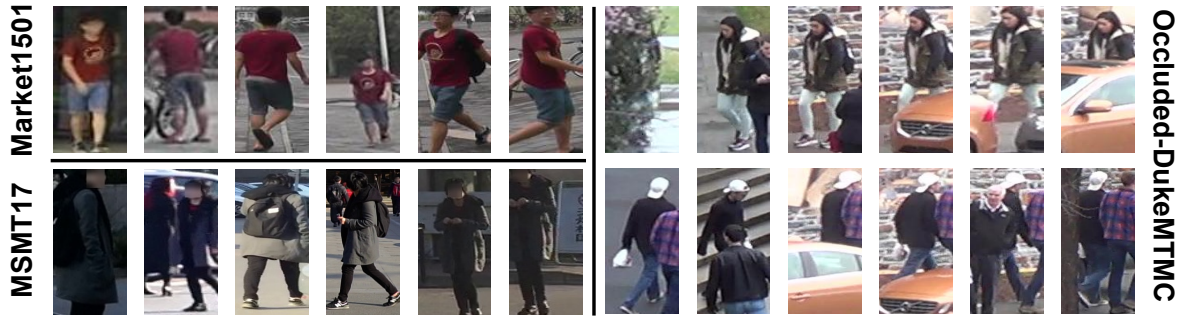


Figure 2: Person re-identification datasets often include eight images of the same person taken from different viewpoints and cameras. The Market1501 and MSMT17 datasets are known for their diverse viewpoints, lighting conditions, and backgrounds. On the other hand, the Occluded-DukeMTMC dataset poses extra challenges with its complex environments and frequent partial occlusions, complicating the re-identification process.

ployed in various tasks, including re-identification (Zhong et al., 2017; Yu et al., 2017; Liu et al., 2018; Wu et al., 2018; Liu et al., 2020; Zhang et al., 2020). A notable advantage of many re-ranking techniques is their ability to be deployed without needing additional training samples, allowing them to be applied to any initial ranking results. Traditionally, common methods for re-ranking include recalculating image similarities by integrating additional information and using advanced similarity metrics (Hirzer et al., 2012; Liao et al., 2015; Yan et al., 2015; Hai Phan, 2022), feature similarity-based methods (Radenović et al., 2018; Chum et al., 2007), neighbor similarity (Bai and Bai, 2016; Zhong et al., 2017; Sarfraz et al., 2018a; Shen et al., 2012; Qin et al., 2011), and graph-based approaches (Zhang et al., 2020; Zhang et al., 2023). However, previous methods still rely on single-view features directly extracted from the feature extraction model to represent gallery features. Relying only on single-view features can result in a lack of the necessary information to fully and accurately represent pedestrians, especially in complex situations where images are captured from different cameras.

The study (Xu et al., 2021; Che et al., 2025) emphasized the limitation of view bias in single-view features when re-identifying pedestrians from different camera viewpoints. Since a pedestrian can be captured by multiple cameras, two images of the same person may not have the same details, as one image might lack information present in the other, as illustrated in Figure 2. Generating features that capture information from multiple viewpoints is an effective solution for mitigating view bias. In this paper, we approach transforming single-view features to multi-view features. We draw several conclusions, including: (1) single-view features can be effectively transformed into multi-view features, (2) selecting single-view features can be done effectively in an unsupervised manner if the correct number of **K** is chosen, without the need for model fine-tuning, and (3) generating single-view features by selecting appropriate weights allows the generation of multi-view features that better capture the information from single-view features.

In this paper, we propose generating multi-view features to represent all images in the gallery set during the re-ranking stage, thereby addressing the issue of view bias. To achieve this, we generate multi-view features using the *K-nearest Weighted Fusion (KWF)* method, which generates multi-view features from K-nearest neighbor features. Our proposed unsupervised neighbor feature selection method does not require fine-tuning of pre-trained models. The contributions of this paper are summarized as follows:

- We propose two-stage hierarchical person re-identification: the first stage involves ranking based on single-view features, followed by a second stage re-ranking using multi-view features.
- We propose the *KWF* method, a multi-view feature representation used during the re-ranking stage. This representation prevents the view bias problem in person re-identification.
- We study the effectiveness of weight selection strategies in multi-view feature generation for *KWF*, including *Uniform, Inverse Distance Power* and *Exponential Decay*.
- We perform extensive experiments comparing our proposed methods with other re-ranking approaches on datasets like Market-1501, MSMT17, and Occluded-DukeMTMC.

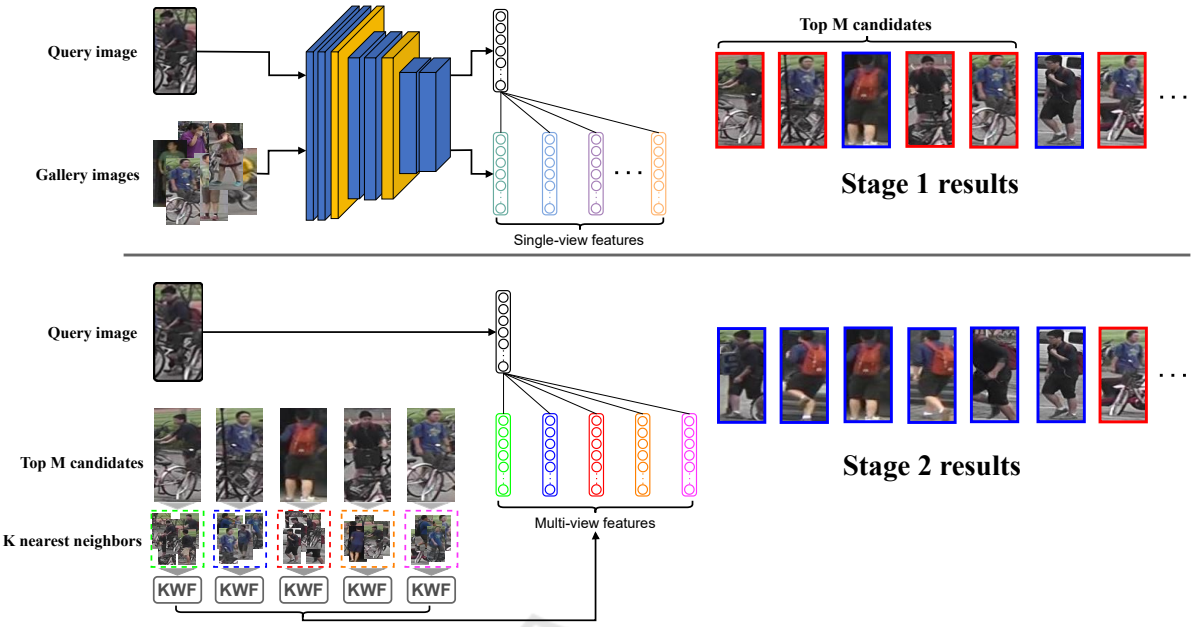


Figure 3: Our person identification pipeline consists of two stages. In the first stage, gallery images are ranked according to the cosine distance between the query and gallery images. In second stage, the top 5 candidates from the first stage are re-ranked using multi-view features.

2 RELATED WORK

2.1 Re-Ranking Approach

Image retrieval in general, and person re-identification in particular, involves searching for images from large databases based on a query image. Re-ranking techniques are crucial in refining the initial search results to enhance retrieval accuracy. Among these, feature similarity-based methods (Chum et al., 2007; Radenović et al., 2018; Arandjelović and Zisserman, 2012) leverage the nearest samples with similar features to enrich the query features by aggregating the features of neighbors. Additionally, neighbor similarity-based methods (Bai and Bai, 2016; Zhong et al., 2017; Qin et al., 2011) rely on the number of shared neighbors between images, using k-reciprocal nearest neighbors to efficiently exploit the relationships among images. Furthermore, graph-based re-ranking methods (Zhang et al., 2023; Zhang et al., 2020) capture the topological structure of the data and refine the learned features, yielding promising results. While some re-ranking methods do not require extra annotations, others require human interaction or label supervision (Bai et al., 2010; Liu et al., 2013; Leng et al., 2015). Most re-ranking methods focus on directly exploiting relationships between the initial features. However, the effectiveness of enhancing multi-view features in this context remains underexplored. Therefore, in this study, we propose a method that employs multi-view features during re-ranking without requiring model fine-tuning or extra annotations.

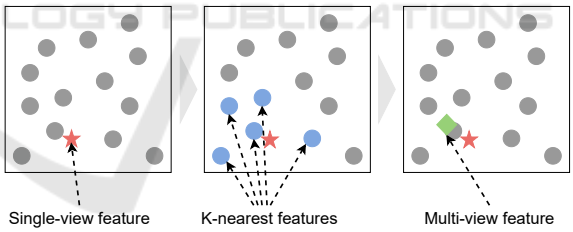


Figure 4: The process of generating multi-view features from single-view features: First, the single-view feature calculates the distance to all features in the gallery to find the **K** nearest neighbors. Then, the *KWF* method aggregates the neighboring features to generate a multi-view feature.

2.2 Multi-View Feature Representation

Previous studies (Wieczorek et al., 2021; Xu et al., 2021) have explored feature fusion in re-identification tasks. In (Wieczorek et al., 2021), the authors suggested aggregating features of the same class during the query process, which proved effective in both accuracy and query time. However, this approach requires additional labels to select representative features for images of the same class. Therefore, this method requires class-level labeling for images when

applied in real-world scenarios, highlighting a limitation of this approach. Yinsong et al. (Xu et al., 2021) proposed feature aggregation using a Graph Neural Network (GNN). While promising, GNN-based methods have drawbacks, such as hardware-dependent computation costs and significant computational and storage demands for large graphs. This study introduces a weighted aggregation method for single-view features based on neighboring features. Our method selects neighboring features unsupervisedly, ensuring efficiency by leveraging the feature extraction capabilities of pre-trained re-identification models.

3 METHOD

In this section, we propose a two-stage hierarchical approach for the person re-identification task. In the first stage, images in the gallery are ranked based on their pairwise distance to a given query. To compute this pairwise distance, we use the cosine distance of features extracted by a re-identification model, where the features representing the gallery images are referred to as single-view features. In the second stage, the single-view features extracted from the first stage are transformed using the *KWF* method. This method aggregates the neighboring features of the single-view features to generate multi-view features. To effectively represent multi-view features, we propose efficient weight selection strategies that allow the multi-view features to capture information from the single-view features more effectively. Figure 3 illustrates the two-stage re-identification process.

3.1 Two-Stage Hierarchical Person Re-Identification

To query an image q within a gallery set of N images $G = \{g_i\}_{i=1...N}$, the distance between q and each gallery image $g_i \in G$ is computed using their feature vectors. To capture essential visual information, we use a pretrained $\mathcal{F}(\cdot)$ to extract representation features. By calculating the distances between $\mathcal{F}(q)$ and each $\mathcal{F}(g_i)$, an initial ranked list $\mathcal{R}(q,G) = \{g_i^0\}_{i=1...N}$ is generated, where $d\big(\mathcal{F}(q),\mathcal{F}(g_i^0)\big) < d\big(\mathcal{F}(q),\mathcal{F}(g_{i+1}^0)\big)$. Our goal is to re-rank $\mathcal{R}(q,G)$ so that more positive samples rank higher in the list, thereby improving the performance of person re-identification.

We propose a two-stage hierarchical person re-identification process as follows:

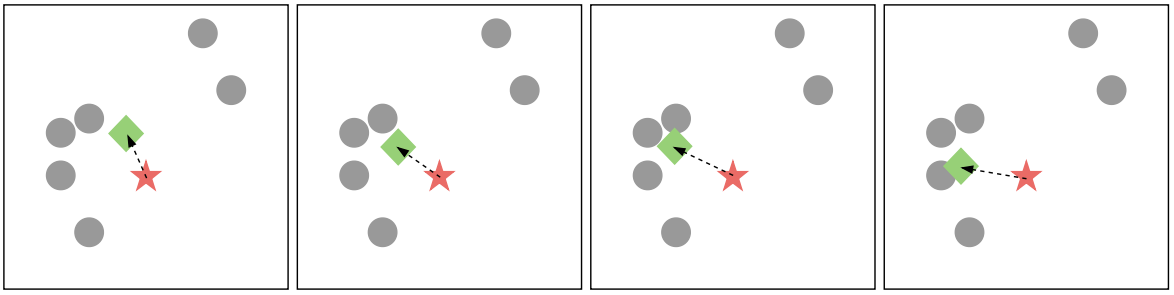
- **Stage 1:** Person re-identification ranking involves ranking gallery images based on the pairwise cosine distances between the query and gallery images in the feature space of a pre-trained feature extraction model.
- **Stage 2:** We perform re-ranking of the top **M** candidates from Stage 1 results by computing feature distances based on multi-view features for the gallery images. Multi-view features is generated by *K-nearest Weighted Fusion* proposed method.

Overall, our approach compares person images through two hierarchical stages (as depicted in Figure 3): single-view feature ranking stage and multi-view feature re-ranking stage.

3.2 K-Nearest Weighted Fusion for Multi-View Features

3.2.1 Generate Multi-View Features from K-Nearest Neighbors

Instead of representing each image using single-view features, we propose using multi-view features to represent the images in the top **M** candidates from the initial ranked list. To generate multi-view features from single-view features, based on the methodology presented in (Che et al., 2025), we propose the K-Nearest Weighted Fusion (*KWF*) method, which aggregates the K-nearest neighboring features for each single-view feature. Our KWF method effectively combines neighboring features to mitigate view bias, providing a more accurate representation of individuals. Specifically, given the results from Stage 1 - $\mathcal{R}(q,G)$, we generate a list of the images in the top **M** candidates. This list of features is denoted as $\mathcal{N}(q,\mathbf{M}) = \{g_j^0\}_{j=1...M}$, where each feature of an image in $\mathcal{N}(q,\mathbf{M})$ is single-view feature. With each single-view feature, K neighboring features are selected and aggregated into a multi-view feature through the proposed *KWF* method. This process is visualized in Figure 4. In our proposed *KWF* method, the nearest neighboring features are selected unsupervised. This selection method may result in neighboring feature lists that include images from different classes with the single-view feature, potentially leading to multi-view features containing inaccurate information. However, theoretically, the feature extraction model $\mathcal{F}(\cdot)$ is trained such that images from the same class have high similarity, and vice versa. Thus, we expect that in the nearest neighbor feature list, the number of images from the same class as the single-view feature's image outnumber those from different classes, allowing positive features to outweigh and suppress negative ones. By recalculating the distance between $\mathcal{F}(q)$ and each multi-view



(a) Uniform. (b) Inverse Distance Power ($p = 1$). (c) Inverse Distance Power ($p = 2$). (d) Exponential Decay.

Figure 5: The process involves transforming single-view features into multi-view features using the *KWF* method. In each figure, the red star represents the single-view feature, gray circles represent the nearest neighboring features (with $\mathbf{K} = 6$), and the green quadrilateral represents the generated multi-view feature.

Table 1: Comparison of our proposed method with various approaches on the Market-1501, MSMT17, and Occluded-DukeMTMC datasets. The results are reported in terms of Rank@1 and mAP. The best results are marked in **bold**, and the second-best results are indicated with underlining.

		Market1501		MSMT17		Occluded-DukeMTMC	
		Rank@1	**mAP**	**Rank@1**	**mAP**	**Rank@1**	**mAP**
BoT Baseline		94.5	85.9	74.1	50.2	48.7	42.6
AQE (Chum et al., 2007)		94.7	91.2	69.8	55.7	58.3	55.7
α-QE (Radenović et al., 2018)		95.4	92.0	70.2	55.9	57.9	55.6
k-reciprocal (Zhong et al., 2017)		95.4	94.2	79.8	**66.8**	58.5	60.3
ECN (Sarfraz et al., 2018b)		95.8	93.2	-	-	-	-
LBR (Luo et al., 2019a)		95.7	91.3	-	-	-	-
GNN (Zhang et al., 2020)		95.8	94.6	-	-	59.3	61.5
GCR (Zhang et al., 2023)		96.1	**94.7**	-	-	63.6	**64.3**
Ours	Uniform	95.4	90.6	82.4	60.0	65.2	54.8
	Inverse Distance Power ($p = 2$)	**96.2**	87.3	**83.9**	56.1	**70.7**	52.2
	Exponential Decay	96.1	90.8	83.2	60.3	66.9	55.3

feature $KWF\left(\mathcal{F}(g_j)\right)$ with $g_j \in \mathcal{N}(q,\mathbf{M})$, we obtain $\mathcal{N}^*(q,\mathbf{M}) = \{g_j^*\}_{j=1\ldots M}$. By re-ranking based on multi-view features, $\mathcal{N}^*(q,\mathbf{M})$ have more positive samples ranked higher compared to $\mathcal{N}(q,\mathbf{M})$. To ensure cross-view matching settings, which is crucial in the re-identification task, the *KWF* feature aggregation method excludes images in the K-nearest neighbors list with the same camera ID as the query image. It is important to note that all features used in Stage 2 are derived from Stage 1 without the need for re-extraction.

3.2.2 Weight Selection Strategy

When generating multi-view features, we aggregate single-view features using the *KWF* method. In the study by (Wieczorek et al., 2021), the aggregated feature of an image is computed by averaging the single-view features. This approach works well when the images used for aggregation belong to the same class as the single-view feature. However, in our unsupervised selection method, treating the contributions of neighboring features equally may lead to suboptimal results. Specifically, features with high similarity (mainly from the same class) and those with low similarity (possibly from different classes) contribute equally to the multi-view feature. Therefore, applying weights during the aggregation process becomes a crucial step. In this study, we thoroughly explore different weight selection strategies for aggregation, as illustrated in Figure 5. Each weight selection strategy generates distinct multi-view features, resulting in unique aggregated representations. The visualized results demonstrate the impact of these weight selection strategies. With $f = \mathcal{F}(I)$ as the feature extracted by the model for any given image I, the multi-view feature $f^{(mv)}$ can be generated following:

$$f^{(\mathrm{mv})} = \sum_{k=1}^{K} w_k \cdot f_k^{(\mathrm{nn})} \tag{1}$$

- $f^{(mv)}$: Multi-view feature representing the image I
- $f_k^{(\mathrm{nn})}$: The k-th neighboring feature in the list of

the **K** nearest neighbors of f

- w_k: The aggregation weight corresponding to $f_k^{(\mathrm{nn})}$

We are suggesting various methods for weight selection strategies to enhance effectiveness:

Uniform Weighting: This method assigns equal weights to all **K** nearest neighbor features, regardless of their distance from the single-view feature. Each neighbor has the same influence on generating the multi-view feature:

$$w_k = \frac{1}{K} \tag{2}$$

Inverse Distance Power Weighting: In this method, weights are assigned inversely proportional to the p-th power of the distance between each neighboring feature $f_k^{(nn)}$ and the single-view feature f. The parameter p controls the rate at which weights decrease as the distance increases. A larger p emphasizes closer features more strongly, while a smaller p allows for a more balanced contribution from distant features. The weights are calculated as:

$$w_k = \frac{1/d^p\left(f, f_k^{(nn)}\right)}{\sum_{j=1}^{K} 1/d^p\left(f, f_j^{(nn)}\right)} \tag{3}$$

After extensive experimentation, we selected $p = 2$ as our default value.

Exponential Decay Weighting: This method employs an exponential function to assign weights based on distance. The weights decrease exponentially as the distance increases, ensuring that only the closest neighbors significantly influence the multi-view feature:

$$w_k = \frac{e^{-d\left(f, f_k^{(nn)}\right)}}{\sum_{j=1}^{K} e^{-d\left(f, f_j^{(nn)}\right)}} \tag{4}$$

4 RESULTS

4.1 Datasets and Settings

Our main results are trained and evaluated on three-person re-identification datasets:

- **Market-1501** (Zheng et al., 2015)**:** The Market-1501 dataset, gathered at Tsinghua University, features 32,668 images of 1,501 individuals captured by six cameras positioned outside a supermarket. The dataset is split into 12,936 training images from 751 individuals and 19,732 testing images from 750 individuals, with 3,368 of the test images designated as the query set.
- **MSMT17** (Wei et al., 2018)**:** MSMT17 is an extensive dataset for multi-scene and multi-time person re-identification. It comprises 180 hours of video from 12 outdoor and 3 indoor cameras across 12 different time slots. The dataset includes 4,101 annotated identities and 126,441 bounding boxes, providing a robust foundation for re-identification studies.
- **Occluded-DukeMTMC** (Miao et al., 2019)**:** This dataset contains 15,618 training images, 17,661 gallery images, and 2,210 occluded query images. Our experiments on the Occluded-DukeMTMC dataset demonstrate the efficacy of our method in handling occluded person re-identification. Importantly, our approach does not require manual cropping during preprocessing.

We utilize the BoT (Luo et al., 2019b) with a ResNet-50 backbone to extract features while evaluating our proposed. We use two evaluation metrics to assess the performance of person Re-ID methods across all datasets. The first metric is the Cumulative Matching Characteristic (CMC). Viewing re-ID as a ranking problem, we report the cumulative matching accuracy at Rank@1. The second metric is the mean Average Precision (mAP). We evaluate the proposed re-ranking method on the top 100 candidates in the initial result list (**M** = 100). For **K**, which is the number of nearest neighbors, we select **K** = 6 for the MSMT17 and Occluded-DukeMTMC datasets, while for Market1501, we choose **K** = 4. The selection of these hyperparameters **M** and **K** is further analyzed in Section 4.3. All experiments used the PyTorch framework on an GeForce RTX 4090 GPU with 32GB of RAM and an Intel(R) Core(TM) i9-10900X processor.

4.2 Main Results

4.2.1 Quantitative Results

As shown in Table 1, our method **consistently outperforms the traditional Stage 1 alone** in both Rank@1 and mAP metrics, regardless of the weight selection strategies used. Among the three strategies, *Inverse Distance Power Weighting (p = 2)* demonstrates the best Rank@1 performance, while *Exponential Decay Weighting* shows uniform improvement in both Rank@1 and mAP. Specifically, with *Inverse Distance Power Weighting*, our method achieves improvements of **1.7%**, **9.8%**, and **22.0%** in Rank@1 on the Market-1501, MSMT17, and

Table 2: Comparison of the computation time and memory usage in re-ranking methods.

Method	Market1501		Occluded-DukeMTMC	
	Evaluate time (↓)	GPU Memory (↓)	Evaluate time (↓)	GPU Memory (↓)
AQE (Chum et al., 2007)	**7.8s**	10.55GB	8.8s	12.53GB
α-QE (Radenović et al., 2018)	7.9s	10.55GB	8.8s	12.53GB
k-reciprocal (Zhong et al., 2017)	146.0s	5.63GB	150.8s	5.62GB
GNN (Zhang et al., 2020)	8.6s	4.75GB	10.97s	4.96GB
GCR (Zhang et al., 2023)	34.5s	20.97GB	35.7s	23.24GB
Our	8.5s	**1.16GB**	**6.1s**	**1.11GB**

Table 3: Performance of our method on Market-1501 and Occluded-DukeMTMC across different top **M** candidates.

Top-M	Market-1501			Occluded-DukeMTMC		
	Rank-1	mAP	Time(ms)	Rank-1	mAP	Time(ms)
20	96.0	88.3	1.41	58.9	46.7	1.58
40	96.1	89.8	1.47	62.6	50.6	1.70
60	96.1	90.3	1.59	64.7	52.8	1.73
80	96.1	90.6	1.70	66.1	54.2	1.85
100	96.1	90.8	1.76	66.9	55.3	1.93
120	96.1	90.9	1.88	67.6	56.1	2.03
140	96.1	90.9	1.94	68.1	56.8	2.17
160	96.1	91.0	2.03	68.6	57.4	2.25

Table 4: Performance of our proposed on Market-1501 and Occluded-DukeMTMC across different **K** values.

K	Market-1501		Occluded-DukeMTMC	
	Rank-1	mAP	Rank-1	mAP
2	95.9	89.3	69.3	53.2
3	95.9	90.2	68.1	54.3
4	96.1	90.8	67.8	54.9
5	95.7	91.0	66.9	55.2
6	95.8	91.1	66.9	55.3
7	95.7	91.0	65.8	55.0
8	95.4	90.9	65.5	54.7
9	95.2	90.7	64.3	54.3
10	95.2	90.6	63.8	53.9

Occluded-DukeMTMC datasets, respectively. Moreover, when comparing our method to other post-processing methods, it delivers higher Rank@1 accuracy, particularly on the MSMT17 and Occluded-DukeMTMC datasets. Our approach focuses on re-ranking the top **M** candidates based on multi-view features, making it particularly effective when evaluated on the Rank@1 metric. Our method improves person re-identification, especially in datasets with occlusions and complex variations, such as MSMT17 and Occluded-DukeMTMC. The proposed weight selection strategies enhance feature representation, leading to better identity discrimination.

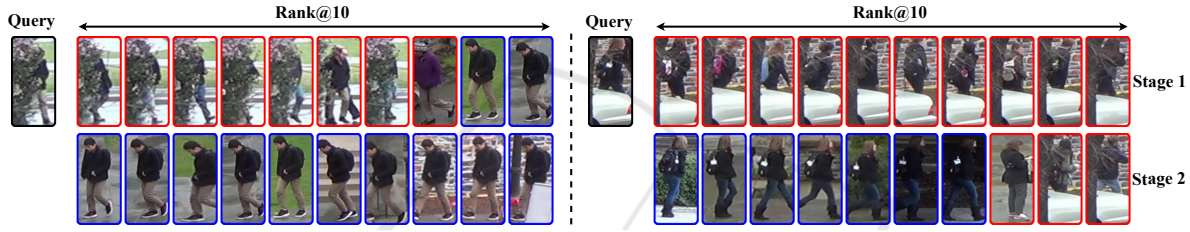
To compare the computational cost of our method with other approaches, we evaluate memory usage and evaluation time (measuring the time five times and taking the average result). All methods are implemented on a GPU (except for the k-reciprocal (Zhong et al., 2017), executed on the GPU in Stage 1 and the CPU in Stage 2). We do not account for the image feature extraction time; instead, we only measure each method's time to process queries and perform re-ranking. The results in Table 2 demonstrate that our method requires significantly less memory than other methods, utilizing only about 1GB of GPU memory. Our approach involves only the time needed to generate multi-view features without requiring additional memory to store new features or auxiliary components. Additionally, with competitive processing time, our method shows strong applicability for large-scale retrieval systems.

4.2.2 Qualitative Results

Figure 6 visualizes some results after using our re-ranking method. Stage 1 shows the initial results based on direct feature distance computation, while Stage 2 presents the re-ranking results using multi-view features. The outcomes demonstrate the robustness of our method in re-identifying images with changes in viewpoint and background across the Market1501 and MSMT17 datasets. Even more impressively, on the Occluded-DukeMTMC dataset, our re-identification approach accurately handles cases

(a) Market1501.

(b) MSMT17.



(c) Occluded-DukeMTMC.

Figure 6: The results from the experiments demonstrate the effectiveness of the proposed reranking method. In each subfigure (Market1501, MSMT17, and Occluded-DukeMTMC), the query images are on the left, followed by columns showing the top 10 most retrieved results. The results from Stage 1 are on the top row of each dataset, while the reranking results from Stage 2 are on the bottom row. Blue and red boxes indicate true positives and false positives, respectively.

where the persons are occluded. These results show that our method effectively aggregates information, significantly enhancing retrieval performance during re-ranking.

4.3 Ablation Studies and Analysis

In this section, we perform extensive ablation studies on the Market1501 and Occluded-DukeMTMC datasets to examine the impact and sensitivity of various hyperparameters.

4.3.1 Effect of Different K

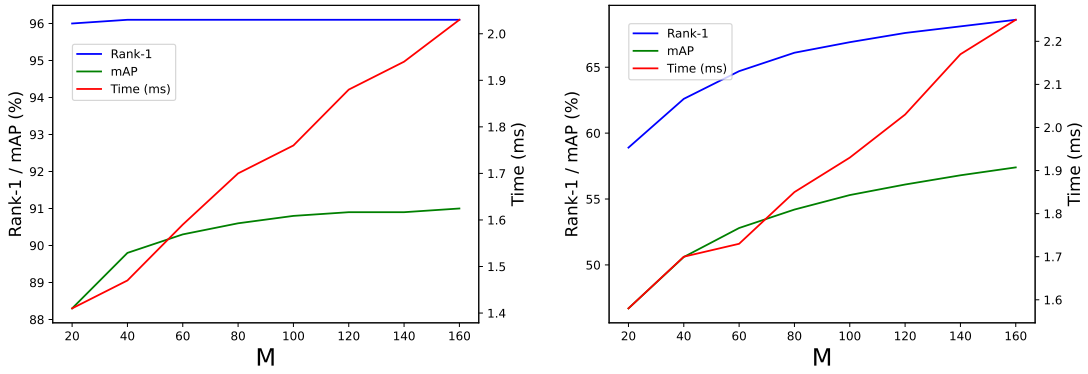
To analyze the impact of the number of nearest neighbors **K** we conducted experiments on the Market-1501 and Occluded-DukeMTMC datasets. Table 4 and Figure 8 demonstrate the trade-off between Rank@1 and mAP as **K** increases from 2 to 10. Specifically, Rank@1 tends to decrease as **K** increases, while mAP rises to **K** = 6 , after which it starts to decline. For the Occluded-DukeMTMC dataset, we chose **K** = 6 as it provides the best balance between Rank-1 and mAP. For the Market-1501 dataset, we selected **K** = 4 to maintain Rank-1 performance, as this metric decreases rapidly with increasing **K**.

Table 5: Performance of our proposed on Market-1501 and Occluded DukeMTMC across different α values.

α	Market-1501		Occluded-DukeMTMC	
	Rank-1	**mAP**	**Rank-1**	**mAP**
0.0	94.5	85.9	48.7	42.6
0.2	95.0	87.7	53.7	47.0
0.4	95.7	89.1	58.1	50.9
0.6	95.9	90.0	63.0	53.6
0.8	96.0	90.5	66.0	54.9
1.0	96.1	90.8	66.9	55.3

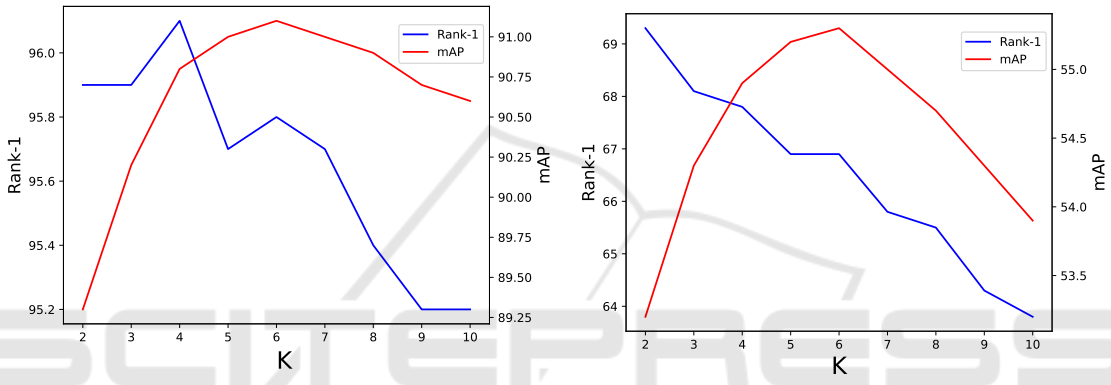
4.3.2 Effect of Different Top M Candidates

The selection of the top **M** candidates, as observed in Table 3 and Figure 7, illustrates the trade-off between accuracy and query time. The query time refers to the duration required to perform a single query, measured five times, with the average value reported. Choosing



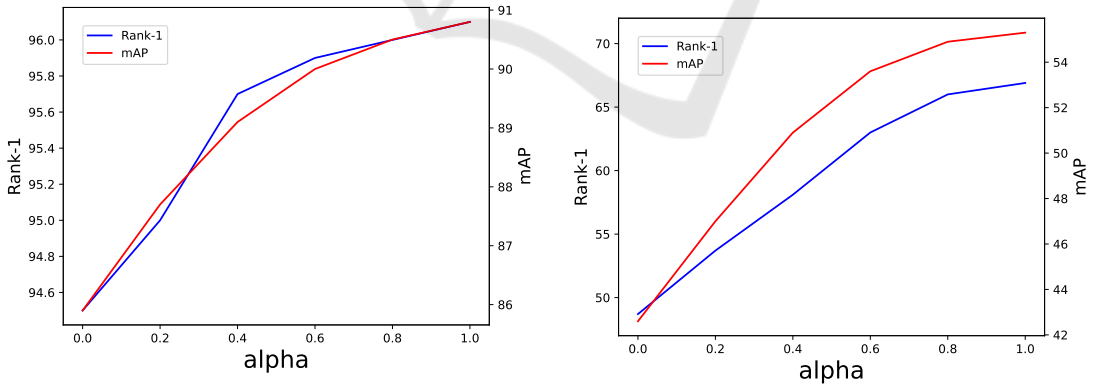
(a) Market-1501. (b) Occluded-DukeMTMC.

Figure 7: Results when select different top **M** candidates.



(a) Market-1501. (b) Occluded-DukeMTMC.

Figure 8: Results when changing **K** for finding nearest neighbors.



(a) Market-1501. (b) Occluded-DukeMTMC.

Figure 9: Results when sweeping across α for linearly combining multi-view and single-view features.

a higher **M** value means re-ranking more candidates using multi-view features, which improves accuracy but also requires more time for feature aggregation. Therefore, we choose $\mathbf{M} = 100$ in this study to balance accuracy and query time.

4.3.3 Combining Multi-View and Single-View Features

In this section, we explore combining the generated multi-view feature and the original single-view feature by linearly combining f and $f^{(mv)}$ using $\alpha \in \{0.0,$

Table 6: The results of our method applied to different baselines.

Method	Market1501		Occluded-DukeMTMC		MSMT17	
	Rank@1	mAP	Rank@1	mAP	Rank@1	mAP
CLIP-ReID (CNN)	94.7	88.1	54.2	47.4	-	-
+ **Ours**	95.7 (↑ **1.0**)	91.8 (↑ **3.7**)	68.1 (↑ **13.9**)	59.5 (↑ **12.1**)	-	-
CLIP-ReID (ViT)	93.3	86.4	60.8	53.5	-	-
+ **Ours**	93.7 (↑ **0.4**)	87.8 (↑ **1.4**)	64.8 (↑ **4.0**)	60.1 (↑ **6.6**)	-	-
BoT (R101-ibn)	95.4	88.9	-	-	81.0	59.4
+ Ours	96.1 (↑ **1.7**)	92.2 (↑ **3.3**)	-	-	86.3 (↑ **5.3**)	67.5 (↑ **8.1**)

Table 7: Experimental results evaluating our method with different indexing techniques, highlighting the trade-offs between query time and re-identification accuracy.

Type	Market-1501				Occluded-DukeMTMC			
	Time (s)		Rank@1	mAP	Time (s)		Rank@1	mAP
	CPU	GPU			CPU	GPU		
Normal	597.21	8.46	96.1	90.8	423.26	6.02	69.1	55.3
IndexIVFPQ	357.84 ↓ 40.08%	7.76 ↓ 8.27%	95.1 ↓ 1.0	88.4 ↓ 2.4	237.31 ↓ 43.93%	5.11 ↓ 15.12%	62.3 ↓ 6.8	51.7 ↓ 3.6
IndexIVFFlat	354.63 ↓ 40.62%	7.19 ↓ 15.01%	95.5 ↓ 0.6	89.0 ↓ 1.8	232.42 ↓ 45.09%	4.74 ↓ 21.26%	61.4 ↓ 7.7	51.4 ↓ 3.9
IndexLSH	515.09 ↓ 13.75%	-	96.2 ↑ 0.1	88.0 ↓ 2.8	358.89 ↓ 15.21%	-	70.7 ↑ 1.6	53.2 ↓ 2.1

0.2, 0.4,..., 1.0}:

$$f^* = (1-\alpha) \times f + \alpha \times f^{(mv)} \quad (5)$$

We found that varying α impacts both Rank@1 and mAP results across the Market-1501 and Occluded-DukeMTMC datasets. The results in Figure 9 and Table 5 show that the larger the value of α, the higher the performance. This indicates that the more significant the contribution of multi-view features, the better the re-ranking performance. Therefore in all experiments we choose α = 1 and ignore the contribution of f.

4.3.4 Effectiveness in Different Baselines

In addition to the results on the baseline BoT Resnet-50, we also experimented with our method on different backbones. Table 6 shows the results on various baselines, including BoT Resnet101-ibn (Luo et al., 2019b), CLIP-ReID Resnet-50, and ViT-B/16 (Li et al., 2023). The results demonstrate that our method can be easily applied to other pre-trained models, leading to significant improvements in both Rank@1 and mAP.

4.3.5 Time Cost Comparison with Different Feature Index Methods

In real-world applications such as surveillance and security, indexing is crucial in managing large-scale galleries. Most available indexing structures involve a trade-off between query time and accuracy. Therefore, in this study, we evaluate our method using various indexing types (Johnson et al., 2019; Douze et al., 2024). Table 7 presents our experimental results. Across both datasets, using IndexIVFPQ and IndexIVFFlat demonstrates a similar trade-off: a slight decrease in re-identification accuracy but a significant reduction in evaluation time on the CPU. On the other hand, when using IndexLSH, the time improvement is not substantial; however, there is a minor increase in Rank@1 accuracy compared to the standard approach. These experiments provide deeper insights into the practical application of our method in real-world query systems.

5 DISCUSSION AND CONCLUSION

5.1 Limitations

In this study, we primarily focus on exploring the potential of a multi-view feature-based approach for the re-ranking stage without optimizing the re-ranking method, resulting in less competitive mAP scores. Additionally, since our methods select features in an unsupervised manner, the performance depends on the pre-trained models. Therefore, the performance of our method relies on the performance of the pre-trained person re-identification models. Finally, we have yet to investigate the effectiveness of multi-view features in other retrieval tasks, which could be an exciting direction for future research.

5.2 Conclusion

In this study, we proposed a two-stage hierarchical person re-identification approach combining single-view and multi-view features. Introducing the K-nearest Weighted Fusion (*KWF*) method addressed the challenges posed by view bias and significantly improved re-ranking performance without requiring additional fine-tuning or annotations. Experimental results on Market-1501, MSMT17, and Occluded-DukeMTMC datasets demonstrate that our method outperforms existing re-ranking techniques in Rank-1 accuracy while maintaining computational efficiency. Our approach improved substantially on challenging datasets with occlusions, highlighting its robustness and practical applicability. This work advances the re-ranking in person re-identification and opens avenues for future research in adaptive feature aggregation and the application of multi-view representations to other domains. By optimizing feature representation, we aim to contribute to developing more accurate and efficient retrieval systems for real-world applications.

ACKNOWLEDGMENTS

This research is funded by University of Information Technology-Vietnam National University of Ho Chi Minh city under grant number D1-2024-70.

REFERENCES

Arandjelović, R. and Zisserman, A. (2012). Three things everyone should know to improve object retrieval. In *2012 IEEE conference on computer vision and pattern recognition*, pages 2911–2918. IEEE.

Bai, S. and Bai, X. (2016). Sparse contextual activation for efficient visual re-ranking. *IEEE Transactions on Image Processing*, 25(3):1056–1069.

Bai, X., Yang, X., Latecki, L. J., Liu, W., and Tu, Z. (2010). Learning context-sensitive shape similarity by graph transduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):861–874.

Che, Q.-H., Nguyen, L.-C., Luu, D.-T., and Nguyen, V.-T. (2025). Enhancing person re-identification via uncertainty feature fusion method and auto-weighted measure combination. *Knowledge-Based Systems*, 307:112737.

Chen, W., Xu, X., Jia, J., Luo, H., Wang, Y., Wang, F., Jin, R., and Sun, X. (2023). Beyond appearance: a semantic controllable self-supervised learning framework for human-centric visual tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15050–15061.

Chum, O., Philbin, J., Sivic, J., Isard, M., and Zisserman, A. (2007). Total recall: Automatic query expansion with a generative feature model for object retrieval. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE.

Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.-E., Lomeli, M., Hosseini, L., and Jégou, H. (2024). The faiss library.

Hai Phan, A. N. (2022). Deepface-emd: Re-ranking using patch-wise earth mover's distance improves out-of-distribution face identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Hirzer, M., Roth, P. M., Köstinger, M., and Bischof, H. (2012). Relaxed pairwise learned metric for person re-identification. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12*, pages 780–793. Springer.

Johnson, J., Douze, M., and Jégou, H. (2019). Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547.

Leng, Q., Hu, R., Liang, C., Wang, Y., and Chen, J. (2015). Person re-identification with content and context re-ranking. *Multimedia Tools Appl.*, 74(17):6989–7014.

Li, S., Sun, L., and Li, Q. (2023). Clip-reid: Exploiting vision-language model for image re-identification without concrete text labels. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37:1405–1413.

Liao, S., Hu, Y., Zhu, X., and Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2197–2206.

Liu, C., Loy, C. C., Gong, S., and Wang, G. (2013). Pop: Person re-identification post-rank optimisation. In *2013 IEEE International Conference on Computer Vision*, pages 441–448.

Liu, Y., Shang, L., and Song, A. (2018). Adaptive re-

ranking of deep feature for person re-identification. *arXiv preprint arXiv:1811.08561*.

Liu, Y., Shen, J., and He, H. (2020). Multi-attention deep reinforcement learning and re-ranking for vehicle re-identification. *Neurocomputing*, 414:27–35.

Luo, C., Chen, Y., Wang, N., and Zhang, Z.-X. (2019a). Spectral feature transformation for person re-identification. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4975–4984.

Luo, H., Gu, Y., Liao, X., Lai, S., and Jiang, W. (2019b). Bag of tricks and a strong baseline for deep person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Miao, J., Wu, Y., Liu, P., Ding, Y., and Yang, Y. (2019). Pose-guided feature alignment for occluded person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Ni, H., Li, Y., Gao, L., Shen, H. T., and Song, J. (2023). Part-aware transformer for generalizable person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 11280–11289.

Qin, D., Gammeter, S., Bossard, L., Quack, T., and Van Gool, L. (2011). Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In *CVPR 2011*, pages 777–784. IEEE.

Radenović, F., Tolias, G., and Chum, O. (2018). Fine-tuning cnn image retrieval with no human annotation. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1655–1668.

Sarfraz, M. S., Schumann, A., Eberle, A., and Stiefelhagen, R. (2018a). A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 420–429.

Sarfraz, M. S., Schumann, A., Eberle, A., and Stiefelhagen, R. (2018b). A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 420–429.

Shen, X., Lin, Z., Brandt, J., Avidan, S., and Wu, Y. (2012). Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3013–3020. IEEE.

Somers, V., De Vleeschouwer, C., and Alahi, A. (2023). Body part-based representation learning for occluded person re-identification. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1613–1623.

Wei, L., Zhang, S., Gao, W., and Tian, Q. (2018). Person transfer gan to bridge domain gap for person re-identification. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 79–88.

Wieczorek, M., Rychalska, B., and Dabrowski, J. (2021). On the unreasonable effectiveness of centroids in image retrieval. In Mantoro, T., Lee, M., Ayu, M. A., Wong, K. W., and Hidayanto, A. N., editors, *Neural Information Processing*, pages 212–223, Cham. Springer International Publishing.

Wu, F., Yan, S., Smith, J. S., and Zhang, B. (2018). Joint semi-supervised learning and re-ranking for vehicle re-identification. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 278–283. IEEE.

Xu, Y., Jiang, Z., Men, A., Wang, H., and Luo, H. (2021). Multi-view feature fusion for person re-identification. *Knowledge-Based Systems*, 229:107344.

Yan, C., Shan, S., Wang, D., Li, H., and Chen, X. (2015). View-adaptive metric learning for multi-view person re-identification. In *Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part II 12*, pages 688–702. Springer.

Yu, R., Zhou, Z., Bai, S., and Bai, X. (2017). Divide and fuse: A re-ranking approach for person re-identification. *arXiv preprint arXiv:1708.04169*.

Zhang, X., Jiang, M., Zheng, Z., Tan, X., Ding, E., and Yang, Y. (2020). Understanding image retrieval re-ranking: A graph neural network perspective. *arXiv preprint arXiv:2012.07620*.

Zhang, Y., Qian, Q., Wang, H., Liu, C., Chen, W., and Wang, F. (2023). Graph convolution based efficient re-ranking for visual retrieval. *IEEE Transactions on Multimedia*, 26:1089–1101.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., and Tian, Q. (2015). Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124.

Zhong, Z., Zheng, L., Cao, D., and Li, S. (2017). Re-ranking person re-identification with k-reciprocal encoding. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3652–3661.