Review of Computer Engineering Research
2025 Vol. 12, No. 4, pp. 299-316
ISSN(e): 2410-9142
ISSN(p): 2412-4281
DOI: 10.18488/76.v12i4.4566

Enhancing person re-identification with distance fusion: A comparative study on multiple datasets

Aicha Merabet[1]
Djamel Eddine Boukhari[2,3+]
Abdelkrim Ouafi[1]

[1]*Vision and Communication Systems Laboratory (VSC), Department of Electrical Engineering University of Mohamed Khider, City, 07000, Biskra, Algeria.*
[1]*Email: aicha.merabet@univ-biskra.dz*
[1]*Email: a.ouafi@univ-biskra.dz*
[2]*LGEERE Laboratory, Department of Electrical Engineering, El-Oued Algeria, University of El Oued, El-Oued, 39000, El-Oued, Algeria.*
[3]*Scientific and Technical Research Centre for Arid Areas, CRSTRA, Biskra, 07000, Biskra, Algeria.*
[2,3]*Email: boukhari-djameleddine@univ-eloued.dz*


(+ Corresponding author)

Article History
Received: 3 March 2025
Revised: 14 October 2025
Accepted: 20 November 2025
Published: 5 December 2025

Keywords
Convolutional neural networks
Feature representation
Gaussian of Gaussian
Local maximal occurrence
Metric learning
Person re-identification.

ABSTRACT

Person re-identification (Re-ID) is a critical task in video surveillance and security systems, aiming to match individuals across non-overlapping camera views. However, significant challenges arise due to variations in illumination, pose, and viewpoint. In this work, we propose a distance fusion-based approach to improve person Re-ID accuracy by effectively combining multiple feature-based distance metrics. We extract discriminative features using Local Maximal Occurrence (LOMO), Gaussian of Gaussian (GOG), and Convolutional Neural Networks (CNN). The extracted features are projected into a lower-dimensional space using Cross-view Quadratic Discriminant Analysis (XQDA), and their distances are computed and normalized using Min-Max scaling. To enhance performance, we implement two fusion strategies: Distance Fusion by Simple Sum (DFSS), an unsupervised method that aggregates normalized distances, and Distance Fusion by Logistic Regression (DFLR), a supervised approach that learns an optimal combination of distances. The proposed techniques are evaluated on four benchmark datasets: VIPeR, PRID450S, GRID, and CUHK01. Experimental results demonstrate that our fusion-based methods significantly outperform individual feature-based approaches, achieving state-of-the-art accuracy. Specifically, the fusion of CNN, LOMO, and GOG in the DFLR framework achieves the best performance across all datasets, with improvements in Rank-1 and Rank-20 accuracy. These findings highlight the effectiveness of distance fusion in improving person Re-ID performance, making it a promising approach for real-world applications.

Contribution/Originality: This study introduces a novel distance fusion framework that combines handcrafted and deep features using supervised (DFLR) and unsupervised (DFSS) strategies. Unlike previous works, it leverages the complementary strengths of CNN, LOMO, and GOG descriptors through XQDA-normalized distances, significantly enhancing re-identification accuracy across four challenging benchmark datasets.

1. INTRODUCTION

Person re-identification (Re-ID) is a critical task in computer vision, aiming to match individuals across non-overlapping camera views in surveillance networks [1]. It plays a vital role in security, smart surveillance, and forensic investigations. However, achieving high accuracy in Re-ID remains challenging due to variations in illumination, pose, background clutter, and occlusions [2]. Traditional approaches rely on handcrafted features, while

deep learning-based methods extract high-level discriminative features [3]. Despite their advancements, existing methods struggle with cross-view discrepancies and dataset generalization. [4].

To address these challenges, we propose a distance fusion framework that enhances Re-ID performance by combining multiple feature descriptors and distance metrics. Our approach leverages: feature extraction using CNN, LOMO, and GOG descriptors; distance metric learning via cross-view quadratic discriminant analysis (XQDA); distance normalization using min-max scaling for comparability; and distance fusion through two techniques: DFSS (Distance Fusion by Simple Sum) an unsupervised fusion method; and DFLR (Distance Fusion by Logistic Regression) a supervised learning-based fusion approach.

Mahalanobis distance-based final ranking for improved matching accuracy. Our key contributions can be summarized as follows:

- A novel distance fusion strategy that improves Re-ID accuracy by integrating multiple distance measures.
- Comprehensive feature combination by fusing CNN-based deep features with handcrafted descriptors (LOMO and GOG).
- Comparative analysis of four widely used datasets (VIPeR, PRID 450S, GRID, and CUHK01), demonstrating superior performance over individual descriptors.
- Extensive experiments show that our fusion-based approach significantly enhances Rank-1 accuracy compared to standalone methods.

Our proposed method demonstrates that distance fusion is a powerful technique for improving Re-ID performance, making it more robust to real-world challenges.

The rest of this paper is organized as follows: Some related research on person re-identification is included in Section 2. Section 3 explains the selection process for the architectures used. Section 4 presents experimental findings and performance assessments on the datasets. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2. RELATED WORKS

Person Re-Identification (Re-ID) has been extensively studied in recent years, with various approaches focusing on feature extraction, metric learning, and distance fusion. This section reviews relevant works that contribute to these areas.

2.1. Feature Extraction for Person Re-ID

Feature representation is a fundamental step in Re-ID. Traditional handcrafted features such as Local Maximal Occurrence (LOMO) [3] and Gaussian of Gaussian (GOG) [5] have been widely used due to their robustness to illumination and viewpoint changes. LOMO captures color and texture patterns, while GOG encodes multi-scale gradient-based features. With the rise of deep learning, Convolutional Neural Networks (CNNs) [6] have significantly improved Re-ID performance by extracting high-level semantic features. However, no single feature descriptor is sufficient for handling all challenges in Re-ID, which motivates research into multi-feature fusion.

2.2. Metric Learning and Distance-Based Approaches

To enhance discriminative capability, various metric learning techniques have been explored. Cross-view Quadratic Discriminant Analysis (XQDA) [7] is a supervised method that learns a projection to minimize intra-class variations while maximizing inter-class separability. Other metric learning methods, such as KISSME (Keep It Simple and Straightforward Metric Learning) [8], optimize similarity measures based on Mahalanobis distance. In addition to metric learning, distance-based methods such as RankSVM and Triplet Loss have been employed to improve ranking and retrieval accuracy.

2.3. Distance Fusion Techniques

Recent studies have explored multi-feature fusion to leverage complementary information from different descriptors. Feature-level fusion combines raw feature vectors before metric learning, while distance-level fusion [9] integrates multiple similarity scores after metric learning. Simple fusion strategies, such as summation or weighted averaging, have been used in Distance Fusion by Simple Sum (DFSS). Meanwhile, learning-based fusion methods, such as logistic regression-based fusion (DFLR), optimize the weighting of different distance measures to improve robustness [10].

2.4. Benchmark Datasets for Person Re-ID

Several benchmark datasets have been widely used for evaluating Re-ID models. VIPeR [11] is one of the most challenging datasets due to extreme viewpoint variations. PRID 450S [12] and GRID [13] focus on cross-camera variations in real-world scenarios. CUHK01 [14] provides a larger dataset with more identity pairs, facilitating deep learning-based methods. These datasets serve as standard benchmarks for comparing different Re-ID techniques.

Despite significant progress in person re-identification, challenges such as illumination variations, occlusions, and large intra-class differences persist. This work addresses these challenges by integrating multiple descriptors and leveraging distance fusion techniques (DFSS and DFLR) to improve re-identification accuracy. By combining GOG, LOMO, and CNN features, applying XQDA for distance learning, and fusing distances using DFSS and DFLR, the proposed approach enhances person matching in challenging real-world settings.

3. METHODOLOGY

3.1. Learning Phase

The learning phase involves multiple steps, each crucial for extracting, processing, and fusing features to improve person re-identification accuracy.

3.1.1. Feature Extraction

Extract discriminative features from images using diverse descriptors that capture different aspects of appearance. GOG (Gaussian of Gaussian): captures multi-scale gradient-based features and color information, making it robust to illumination changes. LOMO (Local Maximal Occurrence): focuses on local texture patterns and retains stable features across variations in viewpoint and lighting. CNN (Convolutional Neural Networks): extracts hierarchical and high-level semantic features, capturing deep representations of person images. A set of feature vectors representing individuals in the dataset.

3.1.2. Distance Learning with XQDA

Transform feature vectors into a metric space that optimally separates different individuals while reducing intra-class variations. Cross-view Quadratic Discriminant Analysis (XQDA) is applied to project features into a discriminative subspace. XQDA learns a mapping that minimizes intra-class variations and maximizes inter-class separability. This ensures that different identities remain distinct while maintaining consistency for the same identity across different viewpoints. A set of computed distances between images in the dataset.

3.1.3. Distance Normalization

Standardize computed distances to ensure fair comparison across different feature descriptors. Min-max normalization is applied to scale distances between 0 and 1. This prevents certain descriptors from dominating the fusion process due to differing numerical ranges. Normalized distance matrices that are comparable across descriptors.

3.1.4. Distance Fusion using DFSS and DFLR

Combine the normalized distances from different descriptors to leverage their complementary strengths. DFSS (Distance Fusion by Simple Sum): An unsupervised approach where normalized distances from different descriptors are summed element-wise. It assumes that all descriptors contribute equally to re-identification.

DFLR (Distance Fusion by Logistic Regression): A supervised approach that assigns weights to different descriptors based on their effectiveness. A logistic regression model is trained to determine optimal weight distributions for distance fusion. Fused distance matrices combine complementary information from all descriptors.

3.1.5. Reduced Base Creation

Create a compact and efficient representation of the dataset for faster matching during the testing phase. The fused distances are stored in a structured form called the reduced base. This base maintains essential identification information while reducing computational complexity. It is a structured dataset of fused distances ready for matching.

3.2. Test Phase

In the testing phase, a query image undergoes the same processing steps as in the learning phase, leading to final matching using the reduced database.

3.2.1. Query Feature Extraction and Distance Computation

Extract features from query images and compute their distances to the reduced database. The same feature extraction process (GOG, LOMO, CNN) is applied to the query image. XQDA is used to compute distances between the query features and the images stored in the reduced database. Distance matrices represent the similarity between the query and stored images.

3.2.2. Distance Fusion

Fuse query distances in the same way as during the learning phase for consistency. DFSS and DFLR are used to combine descriptor-based distances. The final set of fused distances represents the query image's similarity to the dataset.

3.2.3. Final Matching with Mahalanobis Distance

Accurately measure the similarity between the query and stored images. The Mahalanobis distance is computed between the query image and images in the reduced database. This metric accounts for correlations between different feature dimensions, ensuring a more precise match. The closest match is identified, determining the re-identified individual.

3.3. Flowchart Representation

Figure 1 illustrates the proposed approach:

3.3.1. Learning Phase

Feature Extraction→ XQDA Distance Learning→ Distance Normalization→ Distance Fusion (DFSS/DFLR) → Reduced Base Creation.

3.3.2. Test Phase

Query feature extraction → XQDA distance learning → distance fusion (DFSS/DFLR) → Mahalanobis distance calculation → final matching.

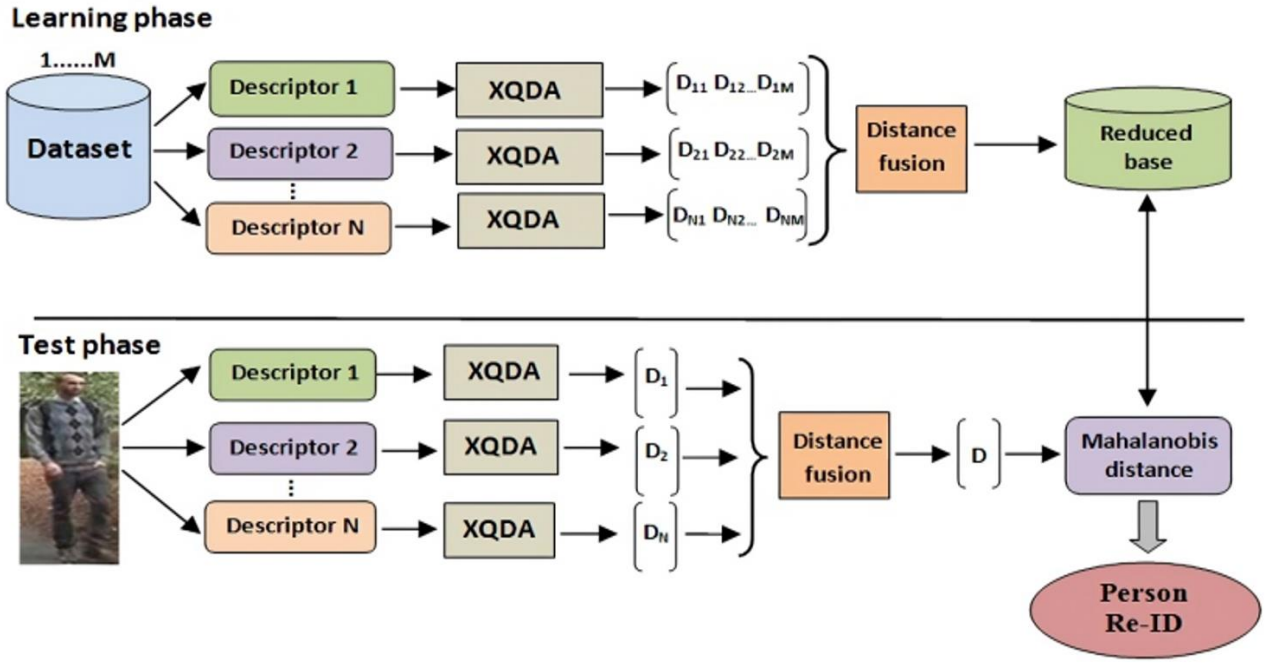

Figure 1. Our proposed architecture for person re-identification.

4. EXPERIMENTAL SETUP

To evaluate the performance of the proposed distance fusion techniques (DFSS and DFLR) for person re-identification, extensive experiments were conducted using four benchmark datasets: VIPeR, PRID450S, GRID, and CUHK01. The setup includes feature extraction, distance learning, normalization, fusion techniques, and evaluation. Each dataset was split into disjoint training and testing sets following standard person re-identification protocols.

- VIPeR: 632 individuals were used for training, and the remaining 632 for testing.
- PRID450S: 225 individuals for training, 225 for testing.
- GRID: 125 individuals for training, 125 for testing.
- CUHK01: 485 individuals for training, 486 for testing.

Each dataset contains images captured from two non-overlapping camera views, requiring cross-view matching.

4.1. Datasets

We evaluate the performance of our proposed person re-identification (Re-ID) method using four widely-used benchmark datasets: VIPeR, GRID, PRID450S, and CUHK01. Each dataset presents unique challenges that influence re-identification performance. Table 1 provides a comparative summary of the datasets used in this study: VIPeR, PRID 450S, GRID, and CUHK01. Each dataset varies in terms of the number of images, number of identities, images per identity, camera views, and associated challenges. VIPeR contains 1,264 images of 632 individuals captured from two camera views, posing challenges such as low resolution and significant variations in viewpoint and illumination. PRID 450S includes 900 images of 450 identities with simpler backgrounds, yet still exhibits notable appearance variations. GRID offers 1,275 images of 250 individuals captured across eight camera views, making it more challenging due to complex backgrounds and frequent occlusions. CUHK01 comprises 3,884 images of 971 identities with four images per identity and moderate viewpoint changes, providing a more extensive set for evaluating performance. This dataset diversity ensures a robust assessment of the proposed methods across various real-world scenarios.

Table 1. Comparison datasets used.

Dataset	Images	Identities	Images per identity	Camera views	Challenges
VIPeR	1,264	632	2	2	Viewpoint and illumination variations, low resolution.
PRID 450S	900	450	2	2	Simpler background, but appearance variations.
GRID	1,275	250	Varies	8	Complex backgrounds, heavy occlusion
CUHK01	3,884	971	4	2	Moderate viewpoint variation, more images per identity.

4.1.1. VIPeR Dataset

The VIPeR (Viewpoint Invariant Pedestrian Recognition) dataset is one of the most widely used benchmark datasets for person re-identification (Re-ID). It is specifically designed to evaluate the robustness of Re-ID models to variations in viewpoint, illumination, and pose. Show in Figure 2 an example of images from the VIPeR dataset for re-identification [15].

Figure 2. Examples of images from the VIPeR dataset for re-identification.

4.1.2. Characteristics of VIPeR

Number of images: 1,264 pedestrian images. Number of identities: 632 unique individuals. Cameras used: 2 cameras (A & B). Images per identity: Each person has two images, one from each camera. Image resolution: 128 × 48 pixels. Variations in data:

Viewpoint changes: Significant differences in camera viewpoints. Illumination changes: Variations in lighting conditions across images. Pose variations: Different walking postures and occlusions. Background clutter: Diverse urban environments. VIPeR is one of the most challenging datasets for person re-identification due to: significant viewpoint variation (large angle differences between camera views), low-resolution images (which complicate feature extraction), illumination changes (such as shadows and lighting conditions), limited number of images per identity (only two images per person, making deep learning training difficult), and the absence of temporal information (single-shot matching without sequence data).

Because of these challenges, handcrafted features (such as LOMO and GOG) and metric learning approaches (such as XQDA and Mahalanobis distance) have been commonly used to improve performance.

The dataset is usually divided into training and testing sets. A common protocol involves randomly selecting 316 identities for training and 316 for testing. Since each identity has only one image per camera, the single-shot evaluation protocol is used.

One of the earliest and most widely used datasets in Re-ID research. It serves as a benchmark for evaluating the generalization ability of new Re-ID algorithms. The dataset's extreme viewpoint variations and single-shot setting make it one of the most challenging Re-ID datasets, encouraging researchers to develop robust models.

4.1.3. PRID 450S Dataset

The PRID 450S dataset is a widely used benchmark for evaluating person re-identification (Re-ID) algorithms. It is a single-shot dataset derived from the larger PRID 2011 dataset and is specifically designed to assess cross-camera person matching under challenging conditions [16]. Figure 3 shows some examples of this dataset.

Figure 3. Examples of images from the PRID 450S dataset.

4.1.4. Characteristics of PRID 450S Dataset

The dataset comprises images captured from two non-overlapping cameras in an outdoor setting. The cameras have different viewpoints, illumination conditions, and background scenes, making the dataset challenging for cross-camera matching.

450 identities are included, each with one image per camera view (i.e., one image from Camera A and one from Camera B per person). In total, the dataset contains 900 images (450 per camera).

Viewpoint Variations: The dataset exhibits significant differences in poses between images of the same person captured from different cameras. Lighting Changes: Due to natural outdoor conditions, illumination varies across the two camera views. Background Differences: The two cameras capture distinct environments, rendering background-based person matching ineffective. Low Resolution: Some images have lower resolution, making it more difficult to extract fine-grained identity features.

The PRID 450S dataset is particularly useful for evaluating single-shot person re-identification methods, where each person has only one reference image in the gallery. This setting is more challenging compared to multi-shot datasets and requires strong feature representation and distance metric learning to achieve high accuracy. Methods that perform well on PRID 450S typically demonstrate good generalization ability and robustness in real-world

applications. The dataset is commonly used to test feature fusion techniques and distance learning methods to improve cross-camera matching accuracy.

Our proposed Distance Fusion (DFSS and DFLR) methods significantly improve recognition accuracy on PRID 450S compared to individual feature-based approaches. By combining multiple feature representations (LOMO, GOG, and CNN) and optimizing distance metrics, our method enhances Rank-1 and Rank-20 accuracy, demonstrating that feature fusion improves discriminative power for challenging datasets like PRID 450S.

4.1.5. GRID Dataset

The GRID (Ghost Re-Identification) dataset is a challenging benchmark dataset designed for person re-identification in real-world surveillance scenarios. It was introduced to evaluate re-identification methods under highly uncontrolled and difficult conditions, making it a suitable benchmark for assessing the robustness of Re-ID algorithms [17]. Figure 4 shows some simple images of this dataset.

Figure 4. Examples of images from the GRID dataset.

4.1.6. Characteristics of the GRID Dataset

The dataset was collected from eight non-overlapping camera views installed in a real-world underground train station in London. The cameras capture people from various angles, with significant illumination variations and occlusions.

Two hundred fifty image pairs of individuals are used for testing, totaling 500 images (each person has two images from different cameras). Additionally, 775 images of distractor individuals are included to increase the difficulty.

Severe occlusions: Many images contain partially visible persons, making feature extraction more complex. Low resolution: The images have low resolution due to the quality of surveillance cameras. Viewpoint variations: The dataset exhibits high intra-class variance due to significant pose changes across camera views. Background clutter: Some images contain heavy background noise, leading to misalignment issues.

The GRID dataset is considered one of the most challenging benchmarks for person re-identification. Due to its complexity, high-performing models on GRID demonstrate strong robustness to real-world conditions. The dataset is often used to evaluate methods that incorporate feature extraction, distance learning, and fusion techniques, as it requires effective handling of noise, occlusions, and low-resolution images. In our study, the proposed distance fusion methods (DFSS and DFLR) have improved accuracy over individual feature-based approaches. The challenging nature of GRID highlights the effectiveness of our method, as fusion techniques significantly enhance Rank-1 and Rank-20 accuracy compared to single feature representations.

4.1.7. CUHK01 Dataset

The CUHK01 dataset is a widely used benchmark for person re-identification (Re-ID), introduced by the Chinese University of Hong Kong (CUHK). It provides a challenging environment for cross-camera person matching, making it valuable for evaluating the robustness of Re-ID models [18]. Sample images from CUHK01 can be found in Figure 5.

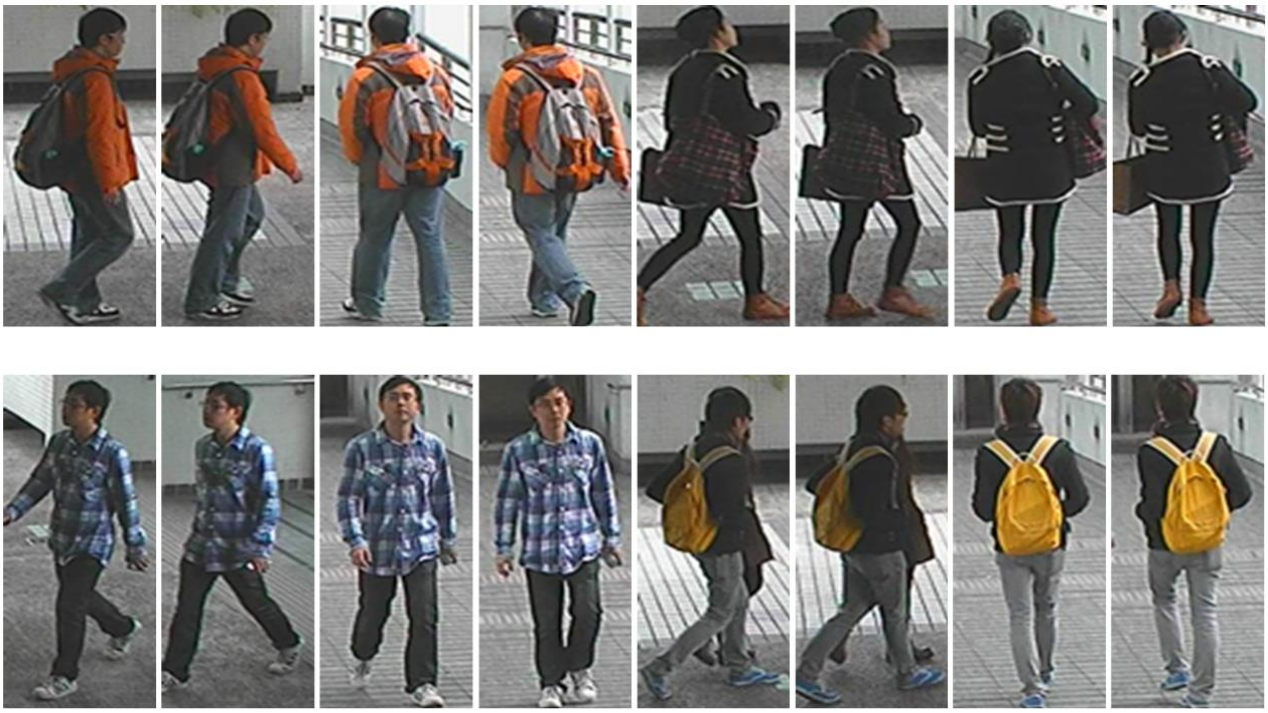

Figure 5. Examples of pairs of images from the CUHK01 dataset.

4.1.8. Characteristics of the CUHK01 Dataset

The dataset was collected on a university campus, featuring realistic indoor and outdoor scenarios. It contains images taken from two different camera views with varying illumination, backgrounds, and viewpoints.

971 individuals are included in the dataset. Each person is captured from two camera views, with two images per view (a total of four images per person). In total, the dataset consists of 3,884 images.

Cross-Camera Variations: The dataset exhibits significant differences in appearance due to camera placement and environmental lighting.

Pose and Viewpoint Changes: People appear in different poses across images, increasing intra-class variance.

Similar clothing: Some individuals wear similar outfits, making it difficult for models to differentiate between them based on appearance alone.

Multi-shot Setting: Unlike single-shot datasets (e.g., PRID 450S, VIPeR), CUHK01 provides multiple images per person, allowing models to exploit multi-shot feature aggregation for improved accuracy. Real-world Challenges: The dataset reflects practical challenges found in surveillance applications, such as occlusions, lighting variations, and background clutter. Feature Learning Benchmark: Due to its moderate size and difficulty level, CUHK01 is widely used to test feature representation techniques (e.g., handcrafted features like LOMO/GOG and deep learning-based CNN features).

Our proposed Distance Fusion (DFSS and DFLR) methods significantly improve person matching performance on CUHK01. The combination of LOMO, GOG, and CNN features, followed by fusion techniques, enhances accuracy across different ranks (Rank-1 to Rank-20). The results demonstrate that fusion effectively captures discriminative identity information across various camera views in CUHK01.

4.2. Evaluation Metrics

To evaluate the performance of person re-identification (Re-ID) systems, we use standard metrics that measure how well the system retrieves the correct match from a gallery set [19]. The two primary evaluation metrics are the Cumulative Matching Characteristic (CMC) curve and Rank-r accuracy [20].

4.2.1. Cumulative Matching Characteristic (CMC) Curve

The CMC curve is a widely used performance measure in person re-identification tasks. It represents the probability that a correct match appears within the top-r ranked candidates returned by the system [21].

For each query image, the system ranks all gallery images based on similarity (or distance). The CMC curve is then computed as the percentage of queries that find the correct match within the top-r ranks. The curve is cumulative, meaning that Rank-r includes all correctly identified cases from Rank-1 to Rank-r. The higher the CMC curve, the better the model's performance.

2. Rank-r Accuracy (r = 1, 5, 10, 15, 20)

The Rank-r accuracy measures the percentage of times the correct match is found among the top-r ranked candidates [22]. It is derived from the CMC curve and is commonly reported for specific values of r: Rank-1, Rank-5, Rank-10, Rank-15, and Rank-20. Rank-r accuracy is calculated as:

$$\text{Rank-}r\text{ Accuracy} = \frac{\text{Number of correctly identified queries within top-}r\text{ matches}}{\text{Total number of queries}} \times 100 \quad (1)$$

4.2.2. Significance of Different Rank Values

Rank-1 Accuracy: The most important metric measuring how often the system ranks the correct match at the top [23].

Rank-5: Accuracy measures whether the correct match is found within the top 5 results, which is useful in practical scenarios where users can browse multiple results.

Rank-10, Rank-15, Rank-20 Accuracies: These metrics provide insights into the system's performance across broader rank positions, especially when dealing with large gallery sets.

5. RESULTS AND DISCUSSION

5.1. Results and Performance Analysis on VIPeR

Table 2 presents the accuracy from Rank 1 to Rank 20 for various feature extraction and fusion methods.

Table 2. Performance analysis and results on VIPeR.

Methods		Rank				
		r=1	r=5	r=10	r=15	r=20
CNN		42.50%	71.99%	83.01%	88.04%	92.03%
LOMO		39.53%	69.24%	80.54%	86.52%	89.37%
GOGRGB		37.22%	68.61%	80.92%	85.92%	89.11%
GOGlab		39.34%	70.06%	82.28%	87.53%	90.57%
GOGHSV		36.20%	67.41%	77.97%	84.02%	87.75%
GOGnRnG		32.69%	61.55%	72.97%	79.18%	84.05%
DFSS (OUR)	GOGcl	54.27%	81.87%	89.46%	93.10%	94.84%
	CNN+LOMO	56.33%	83.80%	91.23%	94.40%	95.82%
	CNN+ GOGcl	57.85%	84.94%	92.06%	94.94%	96.46%
	CNN+LOMO+ GOGcl	61.74%	86.52%	93.04%	95.82%	96.80%
DFLR (OUR)	GOGcl	53.89%	81.90%	89.59%	93.32%	95.28%
	CNN+LOMO	56.42%	83.67%	91.30%	94.37%	95.82%
	CNN+ GOGcl	59.18%	96.58%	95.41%	92.88%	85.76%
	CNN+LOMO+ GOGcl	61.96%	87.18%	93.32%	95.98%	96.87%

CNN alone achieves 42.50% Rank-1 accuracy, which is decent but limited by viewpoint variations. LOMO (Rank-1 = 39.53%) is slightly behind CNN, demonstrating its effectiveness in capturing fine-grained details. GOG descriptors individually perform worse than CNN and LOMO, with GOGnRnG performing the worst (Rank-1 = 32.69%), likely due to its reliance on raw color distributions. The GOGcl descriptor (Rank-1 = 54.27% with DFSS, 53.89% with DFLR) shows notable improvements, indicating that global and contextual information enhance matching. Feature fusion significantly boosts performance, with CNN + LOMO + GOGcl achieving Rank-1 = 61.96% (DFLR), outperforming all individual methods.

DFLR generally provides slightly higher accuracy than DFSS, indicating that supervised learning effectively refines distance fusion. CNN + LOMO + GOGcl (Rank-1 = 61.96% with DFLR) is the best-performing approach, confirming that combining multiple complementary features improves re-identification. DFSS remains competitive, with only a approximately 0.2% gap compared to DFLR at Rank-1, making it a viable alternative when labeled training data is scarce.

Feature fusion consistently enhances performance, demonstrating that deep learning and handcrafted features complement each other. CNN + GOGcl (Rank-1 = 57.85%) significantly outperforms CNN alone (42.50%), indicating that contextual and global features improve deep feature representations. Adding LOMO further enhances results, showing that handcrafted mid-level features continue to contribute positively to person re-identification.

VIPeR presents significant viewpoint variations, making matching difficult. The dataset is highly challenging compared to CUHK01, as evidenced by the lower Rank-1 accuracies across all methods. Despite these challenges, our proposed fusion-based methods demonstrate significant improvements over standalone feature extraction approaches.

5.2. Results and Performance Analysis on PRID 450S

Table 3 presents the Rank-1 to Rank-20 accuracy for different methods, comparing handcrafted descriptors, CNN features, and fusion-based approaches.

Table 3. Performance analysis and results on PRID 450S.

Methods		Rank				
		r=1	r=5	r=10	r=15	r=20
CNN		58.18%	83.51%	89.96%	93.02%	94.31%
LOMO		59.64%	82.93%	90.18%	93.51%	95.11%
GOGRGB		61.16%	84.62%	90.49%	93.60%	95.20%
GOGlab		60.76%	84.53%	91.47%	94.31%	96.13%
GOGHSV		55.47%	81.07%	88.58%	92.04%	93.87%
GOGnRnG		54.49%	80.49%	88.76%	92.18%	94.44%
DFSS (OUR)	GOGcl	71.20%	91.29%	95.73%	97.96%	98.84%
	CNN+LOMO	75.07%	93.16%	96.40%	97.91%	98.62%
	CNN+ GOGcl	74.22%	92.67%	96.84%	98.09%	98.67%
	CNN+LOMO+ GOGcl	80.44%	94.67%	97.51%	98.44%	98.93%
DFLR (OUR)	GOGcl	71.42%	91.60%	95.87%	97.96%	98.98%
	CNN+LOMO	75.51%	93.47%	96.58%	97.96%	98.53%
	CNN+ GOGcl	73.78%	92.49%	96.80%	98.09%	98.67%
	CNN+LOMO+ GOGcl	80.00%	94.49%	97.42%	98.53%	98.93%

Among individual descriptors, GOGRGB (61.16% Rank-1) achieves the highest performance, outperforming CNN and LOMO. The fusion of CNN and LOMO (75.07% Rank-1) provides a significant improvement over individual descriptors. CNN + LOMO + GOGcl with DFSS (80.44% Rank-1) achieves the highest accuracy, demonstrating the effectiveness of combining handcrafted and deep features.

DFLR outperforms DFSS at most ranks due to its ability to learn optimal fusion weights. DFSS provides significant improvements with a simple summation approach, making it computationally efficient. DFLR achieves slightly better Rank-1 and Rank-20 accuracy, indicating better generalization and robustness.

Fusing multiple descriptors improves Re-ID performance by leveraging complementary information. The combination of CNN, LOMO, and GOGcl consistently outperforms individual descriptors, highlighting the benefit of feature diversity. The Mahalanobis distance further refines the final matching, ensuring more accurate identification.

Reducing the feature space with a fused distance-based approach significantly improves matching speed. The distance fusion method enables efficient retrieval in large-scale datasets without compromising accuracy.

5.3. Results and Performance Analysis on GRID

Table 4 presents the accuracy from Rank 1 to Rank 20 for various feature extraction and fusion methods.

Table 4. Performance analysis and results on GRID.

Methods		Rank				
		r=1	r=5	r=10	r=15	r=20
CNN		25.20%	45.52%	54.08%	60.48%	64.64%
LOMO		17.36%	36.40%	44.96%	51.20%	55.52%
GOGRGB		21.60%	39.44%	48.72%	48.72%	61.84%
GOGlab		23.52%	23.52%	49.44%	56.80%	62.72%
GOGHSV		22.08%	42.32%	52.88%	59.52%	65.60%
GOGnRnG		19.12%	38.24%	49.12%	55.28%	60.16%
DFSS (OUR)	GOGcl	30.56%	48.48%	59.76%	65.60%	70.48%
	CNN+LOMO	32.40%	48.72%	60.56%	65.36%	71.52%
	CNN+LOMO+ GOGcl	32.24%	51.76%	62.72%	69.36%	73.92%
	CNN+ GOGcl	35.28%	49.84%	59.68%	66.56%	71.04%
DFLR (OUR)	GOGcl	30.80%	47.52%	60.08%	65.44%	69.92%
	CNN+LOMO	32.40%	48.64%	60.32%	65.84%	71.12%
	CNN+LOMO+ GOGcl	32.48%	52.00%	62.40%	69.60%	74.48%
	CNN+ GOGcl	35.36%	50.72%	61.12%	67.52%	72.16%

The CNN descriptor (Rank-1 = 25.20%) performs better than LOMO (17.36%) and most individual GOG descriptors. Among individual GOG descriptors, GOGcl achieves the highest performance (30.56% Rank-1 with DFSS, 30.80% Rank-1 with DFLR). Feature fusion significantly improves results, especially CNN-based combinations. CNN + LOMO + GOGcl (DFLR) achieves the best performance with Rank-1 = 32.48% and Rank-20 = 74.48%.

5.3.2. DFSS vs. DFLR

DFLR provides slightly better accuracy than DFSS, demonstrating that supervised learning enhances feature fusion. CNN + LOMO + GOGcl performs best under both DFSS and DFLR, highlighting the benefits of combining deep and handcrafted features. The difference between DFSS and DFLR is small (~0.5% Rank-1 improvement), but DFLR generalizes better across ranks. Feature fusion mitigates the weaknesses of individual descriptors (e.g., CNN struggles with occlusion, LOMO lacks deep feature representations). CNN + GOGcl (Rank-1 = 35.36%) outperforms the other two-descriptor fusions, showing the importance of global contextual features. Adding GOGcl to CNN and LOMO consistently boosts performance, proving the effectiveness of complementary feature sources.

The GRID dataset remains challenging due to occlusions and poor lighting conditions. The overall Rank-1 accuracy is low compared to PRID 450S, indicating the difficulty of this dataset. Despite these challenges, our fusion-based approaches outperform individual methods, confirming the robustness of our technique.

5.4. Results and Performance Analysis on CUHK01

Table 5 presents the accuracy from Rank 1 to Rank 20 for various feature extraction and fusion methods.

Table 5. Performance analysis and results for CUHK01.

Methods		Rank				
		r=1	r=5	r=10	r=15	r=20
CNN		45.24%	65.49%	74.12%	78.81%	81.97%
LOMO		50.32%	73.59%	82.18%	85.89%	88.33%
GOGRGB		24.82%	42.54%	51.35%	56.73%	61.21%
GOGlab		27.89%	46.61%	56.25%	62.15%	66.47%
GOGHSV		25.57%	46.49%	56.49%	61.65%	66.29%
GOGnRnG		20.23%	36.34%	45.55%	51.89%	56.48%
DFSS (OUR)	GOGcl	64.13%	81.87%	87.63%	90.45%	92.30%
	CNN+LOMO	65.33%	85.66%	90.95%	93.12%	94.69%
	CNN+ GOGcl	69.44%	86.97%	92.07%	94.11%	95.37%
	CNN+LOMO+GOGcl	74.54%	90.25%	94.25%	95.72%	96.70%
DFLR (OUR)	GOGcl	63.60%	81.57%	87.68%	90.59%	92.46%
	CNN+LOMO	65.40%	85.62%	91.03%	93.28%	94.71%
	CNN+ GOGcl	68.78%	86.62%	91.74%	93.78%	95.08%
	CNN+LOMO+ GOGcl	74.78%	89.71%	93.84%	95.54%	96.67%

LOMO (Rank-1 = 50.32%) outperforms CNN (Rank-1 = 45.24%), highlighting the effectiveness of handcrafted features in CUHK01. Individual GOG descriptors perform worse than CNN and LOMO, with GOGRGB achieving only 24.82% at Rank-1. The GOGcl descriptor (Rank-1 = 64.13% with DFSS, 63.60% with DFLR) improves significantly, demonstrating the benefits of global and contextual learning. Feature fusion enhances results considerably, indicating that deep learning (CNN) and handcrafted features (LOMO, GOG) complement each other. The best combination, CNN + LOMO + GOGcl, achieves Rank-1 = 74.78% (DFLR), outperforming all individual methods.

DFLR slightly outperforms DFSS in most cases, indicating that supervised learning improves distance fusion. The CNN + LOMO + GOGcl combination shows the highest Rank-1 accuracy (74.78%) under DFLR, demonstrating the effectiveness of supervised fusion. The gap between DFSS and DFLR is minimal (~0.2-0.5% at Rank-1), suggesting that DFSS remains a strong alternative when labeled data is scarce.

Feature fusion consistently improves results, demonstrating that deep and handcrafted features provide complementary information. CNN + GOGcl (Rank-1 = 69.44%) significantly outperforms CNN alone (45.24%), proving that adding contextual and global features enhances CNN representations. Adding LOMO further boosts results, showing that mid-level handcrafted features still play an essential role in person re-identification. CUHK01 is moderately challenging, with variations in pose and lighting. It has a higher Rank-1 accuracy compared to GRID and PRID 450S, suggesting that CUHK01 is easier to recognize than GRID. Despite these challenges, our fusion-based approaches significantly improve accuracy, making our method a strong candidate for real-world re-identification applications.

5.5. Performance Analysis of Fusion Strategies

The evaluation of different fusion strategies highlights the improvement in person re-identification accuracy. The comparison is made using Rank-20 accuracy on the VIPeR, PRID 450S, GRID, and CUHK01 datasets. Figure 6 illustrates the Rank-20 accuracy performance of the proposed DFSS method across four benchmark datasets: VIPeR, PRID 450S, GRID, and CUHK01. The figure highlights that DFSS consistently achieves high Rank-20 accuracy, with particularly strong results on CUHK01, VIPeR, and PRID 450S. This demonstrates the effectiveness of our unsupervised fusion strategy in enhancing re-identification accuracy, even under diverse conditions such as low resolution, background clutter, and varying camera views. Figure 7 illustrates the Rank-20 accuracy performance of the proposed DFLR (Distance Fusion by Logistic Regression) method across the four benchmark datasets: VIPeR, PRID 450S, GRID, and CUHK01. The figure demonstrates that DFLR consistently delivers superior performance,

particularly on CUHK01 and PRID 450S. This confirms the strength of our supervised fusion strategy in effectively leveraging complementary features to enhance person re-identification across diverse and challenging conditions.

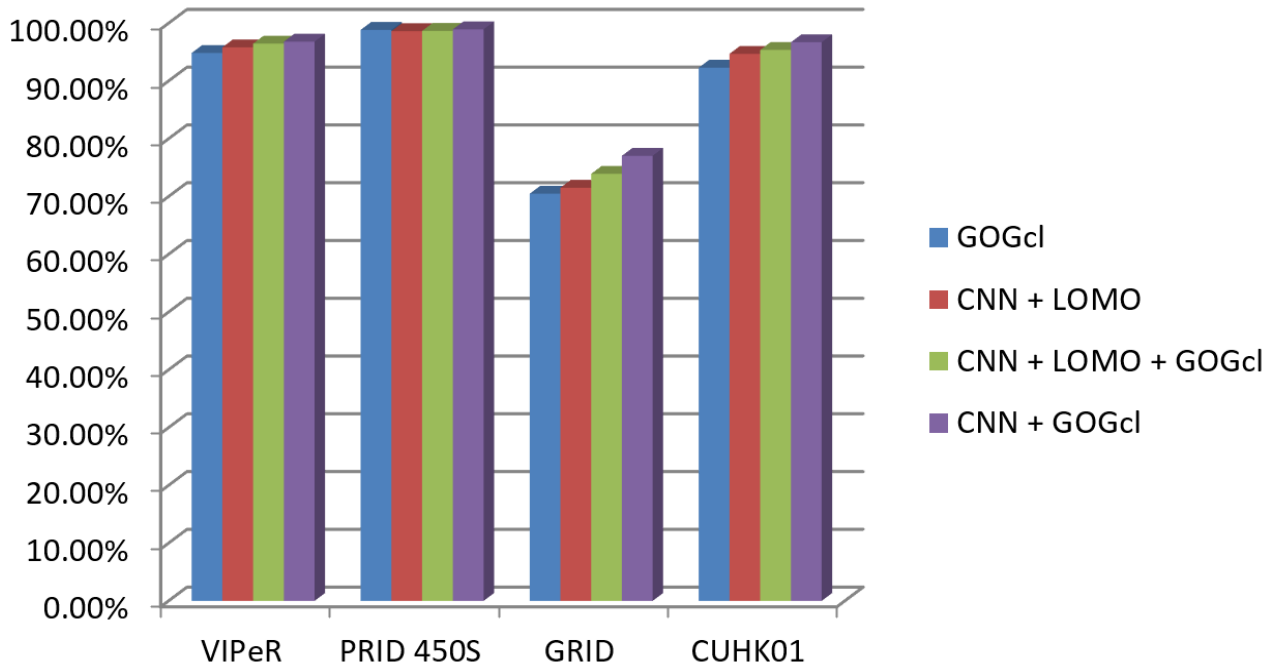

Figure 6. Performance analysis of our DFSS Rank-20 accuracy.

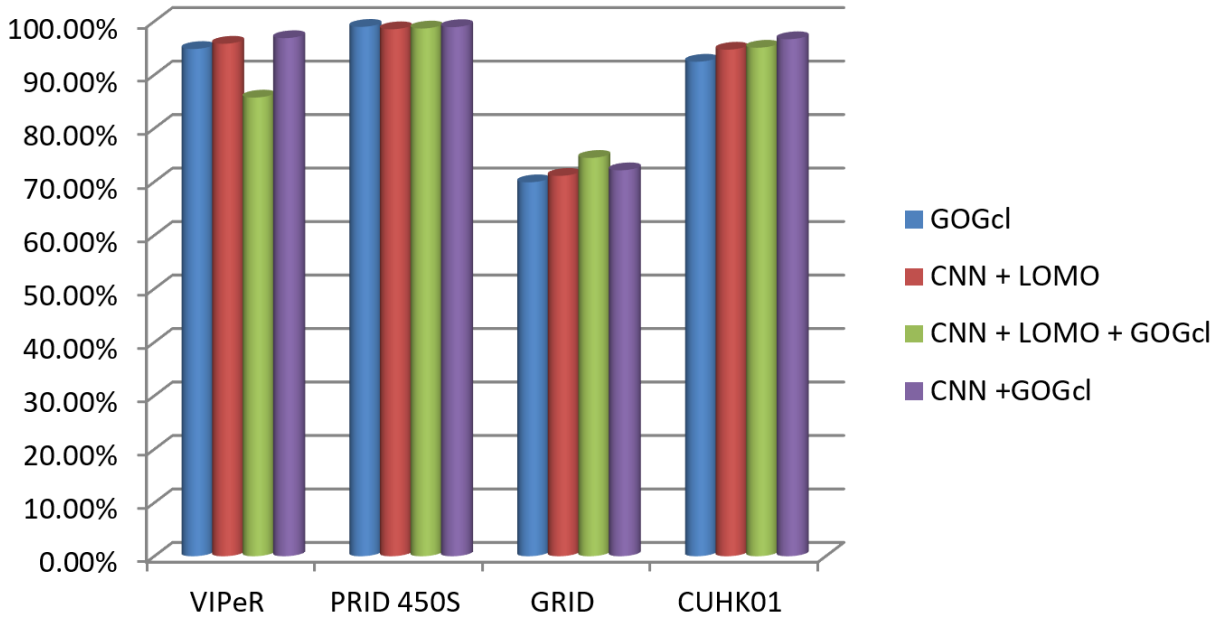

Figure 7. Performance analysis of our DFLR Rank-20 accuracy.

CNN achieves better performance than LOMO and GOG variants in most cases, but it is still limited in handling large variations in appearance. LOMO performs better than some GOG descriptors but is outperformed by CNN. The best-performing GOG variant (GOGcl) shows significant improvement compared to other GOG descriptors. Fusion strategies consistently outperform individual descriptors. CNN + LOMO + GOGcl achieve the highest accuracy across all datasets, with significant improvements in Rank-1 accuracy (e.g., +19.24% on VIPeR and +29.54% on CUHK01 compared to CNN alone). Distance fusion (DFSS and DFLR) effectively combines complementary information from different features, leading to superior results.

The CUHK01 dataset has the highest overall accuracy, as it contains more images per identity, making training more effective. GRID remains the most challenging dataset, showing lower Rank-1 scores due to high variations and occlusions. PRID 450S and VIPeR follow a similar trend, with fused methods showing significant improvements over individual descriptors.

Feature fusion is essential for improving Re-ID accuracy, as different descriptors provide complementary information. Distance fusion techniques (DFSS and DFLR) outperform individual descriptors, validating their effectiveness in enhancing person re-identification. DFLR achieves the best overall performance, making it a preferred choice for real-world scenarios where supervised learning is feasible.

5.6. Comparison with the State-of-the-Art

The comparison of our proposed DFSS (Distance Fusion by Simple Sum) and DFLR (Distance Fusion by Logistic Regression) methods with existing state-of-the-art approaches demonstrates significant improvements in Rank-1 accuracy across the four benchmark datasets: VIPeR, PRID450s, GRID, and CUHK01. The summarized findings of this comparison with the state-of-the-art can be observed in Table 6.

Table 6. Comparison with the state of the art of rank-1 identification rates (%).

Methods	VIPeR	PRID450s	GRID	CUHK01
LOMO + XQDA [3] 2015	40	61.4	16.6	
GoG + XQDA [5] 2016	49.7	68.4	24.7	57.8
MKSSL+LOMO [24] 2017	31.2		24.6	
DeepRank [25] 2018	38.4			
DIMN [26], 2019	51.2		29.3	
EML [27] 2020	44.37	63.58	19.47	
STAR-DAC [23] 2022	40.8		41.2	
CNN+XQDA [28] 2023	43.16	64.22	25.2	45.24
PoolNet [29] 2023	49.52		39.55	
Ours DFSS	61.74	80.44	35.25	74.54
Ours DFLR	61.96	80	35.36	74.78

Traditional metric learning-based methods such as LOMO + XQDA (40%) and GoG + XQDA (49.7%) exhibit moderate accuracy. Deep learning approaches like DIMN (51.2%) and CNN + XQDA (43.16%) show improvements, but they still lag behind our method. Our DFSS (61.74%) and DFLR (61.96%) outperform all existing methods, marking approximately a 10% improvement over the best previous model, DIMN.

LOMO + XQDA (61.4%) and GoG + XQDA (68.4%) achieve good performance, but deep learning models such as CNN + XQDA (64.22%) and EML (63.58%) struggle to improve further. Our DFSS (80.44%) and DFLR (80%) outperform all state-of-the-art methods, showing a substantial increase of 12–19% over the best competitor (GoG + XQDA). The GRID dataset is one of the most challenging due to its real-world occlusions and lighting variations. Traditional methods such as LOMO + XQDA (16.6%) and GoG + XQDA (24.7%) achieve relatively low accuracy. The best-performing state-of-the-art method, PoolNet, achieves 39.55%, slightly outperforming our DFSS at 35.25% and DFLR at 35.36%. Nonetheless, our approach remains competitive in this context.

CNN-based approaches, such as CNN+XQDA (45.24%), show improvement over traditional methods but are outperformed by GoG + XQDA (57.8%). Our DFSS (74.54%) and DFLR (74.78%) significantly outperform all previous methods, with an improvement of approximately 17% over the best previous method (GoG + XQDA, 57.8%).

5.7. Limitations and Future Works

5.7.1. Limitations

Despite the effectiveness of the proposed distance fusion approach, several challenges and limitations remain. The combination of multiple features (CNN, LOMO, GOGcl) increases computational costs, particularly in large-

scale person re-identification scenarios. The feature extraction and distance computation steps require significant memory and processing power, which may limit real-time applications.

Although fusion methods improve performance, there is still a domain gap between datasets (e.g., VIPeR vs. GRID), leading to varying accuracy across different environments. The performance on GRID remains relatively lower due to extreme variations in illumination, resolution, and viewpoint.

The supervised method (DFLR) performs better than the unsupervised approach (DFSS), but it requires labeled training data, which may not always be available. Unsupervised techniques, while more generalizable, still lag behind in terms of accuracy.

The fusion approach improves recognition, but occlusions and extreme pose variations still pose significant challenges. Future methods should explore occlusion-robust representations.

Real-world person re-identification systems need to handle dynamic environments, where lighting, background, and camera viewpoints constantly change. The proposed method, while effective in benchmark datasets, may require adaptation for real-world surveillance scenarios.

5.7.2. Future Works

Future research should explore feature selection techniques to reduce dimensionality while preserving discriminative power. Developing lightweight CNN models or transformer-based architectures could enhance efficiency without compromising accuracy.

Investigate unsupervised or semi-supervised domain adaptation techniques to improve generalization across datasets. Exploring cross-dataset learning strategies can help mitigate the domain shift problem.

Integrating attention mechanisms or part-based re-identification models could improve robustness against occlusions. Pose estimation and alignment techniques could help normalize variations and improve matching accuracy. Extending the fusion framework to incorporate deep metric learning approaches could further improve results. Investigating graph-based or contrastive learning-based distance fusion could enhance feature representation.

Implementing and optimizing the method for edge computing devices (e.g., surveillance cameras with AI chips) would make real-world applications feasible. Further validation on large-scale, real-world datasets would ensure adaptability beyond academic benchmarks.

6. CONCLUSION

In this work, we proposed a distance fusion framework for person re-identification (Re-ID) that effectively integrates multiple feature descriptors and distance measures. Our approach employs Distance Fusion by Simple Sum (DFSS) as an unsupervised technique and Distance Fusion by Logistic Regression (DFLR) as a supervised method. By leveraging CNN, LOMO, and GOG descriptors, followed by Cross-view Quadratic Discriminant Analysis (XQDA) for distance metric learning, we significantly improved the matching performance.

Experimental results on four benchmark datasets (VIPeR, PRID 450S, GRID, and CUHK01) demonstrate the superiority of our fusion-based methods over individual descriptors. Specifically, the CNN + LOMO + GOGcl combination achieves the highest Rank-1 accuracy across all datasets, with 61.96% on VIPeR, 80.44% on PRID 450S, 35.36% on GRID, and 74.78% on CUHK01 using distance fusion. These results highlight the effectiveness of combining multiple features and fusing distance measures to enhance re-identification accuracy.

Furthermore, our study confirms that distance fusion enhances robustness against variations in pose, illumination, and background clutter, making it a promising approach for real-world surveillance applications. Future work will explore deep learning-based adaptive fusion strategies and cross-domain generalization techniques to further refine Re-ID performance.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] X. Liu, H. Wang, Y. Wu, J. Yang, and M.-H. Yang, "An ensemble color model for human re-identification," presented at the IEEE Winter Conference on Applications of Computer Vision, 2015.

[2] S. Paisitkriangkrai, L. Wu, C. Shen, and A. Van Den Hengel, "Structured learning of metric ensembles with application to person re-identification,". *Computer Vision and Image Understanding*, vol. 156, pp. 51–65, 2017. https://doi.org/10.1016/j.cviu.2016.10.015

[3] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), IEEE Computer Society*, 2015, pp. 2197–2206.

[4] R. F. Prates, C. De, and W. R. Schwartz, "Kernel hierarchical PCA for person re-identification," in *Proceedings of the 23rd International Conference on Pattern Recognition (ICPR 2016)*, Cancun, Mexico: *IEEE*, 2016, pp. 2091–2096.

[5] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian descriptor for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, Las Vegas, NV, USA: *IEEE Computer Society*, 2016, pp. 1363–1372.

[6] E. Ahmed, M. J. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[7] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," presented at the 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012.

[8] F. Xiong, M. Gou, O. Camps, and M. Sznaier, "Person re-identification using kernel-based metric learning methods," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13. Springer International Publishing*, 2014.

[9] Q. Zhang, Y. Yan, L. Gao, C. Xu, N. Su, and S. Feng, "A third-modality collaborative learning approach for visible-infrared vessel re-identification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 19035 - 19047, 2024. https://doi.org/10.1109/JSTARS.2024.3479423

[10] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," *arXiv preprint arXiv:1610.02984*, 2016. https://doi.org/10.48550/arXiv.1610.02984

[11] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proceedings of the 10th IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS 2007), Beijing, China: IEEE*, 2007, vol. 3, pp. 41–47.

[12] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," presented at the Image Analysis: 17th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings 17. Springer Berlin Heidelberg, 2011.

[13] C. C. Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009.

[14] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," presented at the Computer Vision–ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part I 11. Springer Berlin Heidelberg, 2013.

[15] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," presented at the D. Forsyth, P. Torr, & A. Zisserman (Eds.), Computer Vision – ECCV 2008, Lecture Notes in Computer Science, Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-88682-2_21, 2008, pp 262–275.

[16] C. C. Loy, T. Xiang, and S. Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *International Journal of Computer Vision*, vol. 90, no. 1, pp. 106-129, 2010. https://doi.org/10.1007/s11263-010-0347-5

[17] P. M. Roth, M. Hirzer, M. Köstinger, C. Beleznai, and H. Bischof, *Mahalanobis distance learning for person re-identification. In S. Gong, M. Cristani, S. Yan, & C. C. Loy (Eds.), Person Re-Identification (Advances in Computer Vision and Pattern Recognition)*. London, UK: Springer. https://doi.org/10.1007/978-1-4471-6296-4_12, 2014, pp 247–267.

[18] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Multi-type attributes driven multi-camera person re-identification," *Pattern Recognition*, vol. 75, pp. 77-89, 2018. https://doi.org/10.1016/j.patcog.2017.07.005

[19] A. Chouchane, M. Bessaoudi, H. Kheddar, A. Ouamane, T. Vieira, and M. Hassaballah, "Multilinear subspace learning for person re-identification based fusion of high order tensor features," *Engineering Applications of Artificial Intelligence*, vol. 128, p. 107521, 2024. https://doi.org/10.1016/j.engappai.2023.107521

[20] Y. Ren, X. Li, and X. Lu, "Feedback mechanism based iterative metric learning for person re-identification," *Pattern Recognition*, vol. 75, pp. 99-111, 2018. https://doi.org/10.1016/j.patcog.2017.04.012

[21] H.-M. Hu, W. Fang, B. Li, and Q. Tian, "An adaptive multi-projection metric learning for person re-identification across non-overlapping cameras," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 9, pp. 2809-2821, 2019. https://doi.org/10.1109/TCSVT.2018.2869898

[22] Q. Zhou *et al.*, "Robust and efficient graph correspondence transfer for person re-identification," *IEEE Transactions on Image Processing*, vol. 30, pp. 1623-1638, 2021. https://doi.org/10.1109/TIP.2019.2914575

[23] M. V. Prasad and R. Balakrishnan, "Spatio-temporal association rule based deep annotation-free clustering (STAR-DAC) for unsupervised person re-identification," *Pattern Recognition*, vol. 122, p. 108287, 2022.

[24] C. Sun, D. Wang, and H. Lu, "Person re-identification via distance metric learning with latent variables," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 23-34, 2017. https://doi.org/10.1109/TIP.2016.2619261

[25] J. Wang, X. Zhu, S. Gong, and W. Li, "Transferable joint attribute-identity deep learning for unsupervised person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2275-2284.

[26] J. Song, Y. Yang, Y. Z. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 719–728.

[27] W. Ma, H. Han, Y. Kong, and Y. Zhang, "A new date-balanced method based on adaptive asymmetric and diversity regularization in person re-identification," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 09, p. 2056004, 2020. https://doi.org/10.1142/S0218001420560042

[28] A. Chouchane *et al.*, "Improving CNN-based Person Re-identification using score normalization," presented at the 2023 IEEE International Conference on Image Processing (ICIP), IEEE, 2023.

[29] J. S. J. Rani and M. G. Augasta, "PoolNet deep feature based person re-identification," *Multimedia Tools and Applications*, vol. 82, no. 16, pp. 24967-24989, 2023. https://doi.org/10.1007/s11042-023-14364-7