

Article

# A Multi-Attention Approach for Person Re-Identification Using Deep Learning

Shimaa Saber <sup>1</sup>, Souham Meshoul <sup>2</sup>, Khalid Amin <sup>1</sup>, Paweł Pławiak <sup>3,4,\*</sup> and Mohamed Hammad <sup>1,\*</sup>

- <sup>1</sup> Information Technology Department, Faculty of Computers and Information, Menoufia University, Shibin El Kom 32511, Egypt; shimaa.saber@ci.menofia.edu.eg (S.S.); k.amin@ci.menofia.edu.eg (K.A.)
- <sup>2</sup> Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia; sbmeshoul@pnu.edu.sa
- <sup>3</sup> Department of Computer Science, Faculty of Computer Science and Telecommunications, Cracow University of Technology, Warszawska 24, 31-155 Krakow, Poland
- <sup>4</sup> Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5, 44-100 Gliwice, Poland
- \* Correspondence: plawiak.pawel@gmail.com (P.P.); mohammed.adel@ci.menofia.edu.eg (M.H.)

**Abstract:** Person re-identification (Re-ID) is a method for identifying the same individual via several non-interfering cameras. Person Re-ID has been felicitously applied to an assortment of computer vision applications. Due to the emergence of deep learning algorithms, person Re-ID techniques, which often involve the attention module, have gained remarkable success. Moreover, people's traits are mostly similar, which makes distinguishing between them complicated. This paper presents a novel approach for person Re-ID, by introducing a multi-part feature network, that combines the position attention module (PAM) and the efficient channel attention (ECA). The goal is to enhance the accuracy and robustness of person Re-ID methods through the use of attention mechanisms. The proposed multi-part feature network employs the PAM to extract robust and discriminative features by utilizing channel, spatial, and temporal context information. The PAM learns the spatial interdependencies of features and extracts a greater variety of contextual information from local elements, hence enhancing their capacity for representation. The ECA captures local cross-channel interaction and reduces the model's complexity, while maintaining accuracy. Inclusive experiments were executed on three publicly available person Re-ID datasets: Market-1501, DukeMTMC, and CUHK-03. The outcomes reveal that the suggested method outperforms existing state-of-the-art methods, and the rank-1 accuracy can achieve 95.93%, 89.77%, and 73.21% in trials on the public datasets Market-1501, DukeMTMC-reID, and CUHK03, respectively, and can reach 96.41%, 94.08%, and 91.21% after re-ranking. The proposed method demonstrates a high generalization capability and improves both quantitative and qualitative performance. Finally, the proposed multi-part feature network, with the combination of PAM and ECA, offers a promising solution for person Re-ID, by combining the benefits of temporal, spatial, and channel information. The results of this study evidence the effectiveness and potential of the suggested method for person Re-ID in computer vision applications.

**Keywords:** ECA; deep learning; PAM; person re-identification; multi-attention



**Citation:** Saber, S.; Meshoul, S.; Amin, K.; Pławiak, P.; Hammad, M. A Multi-Attention Approach for Person Re-Identification Using Deep Learning. *Sensors* **2023**, *23*, 3678. <https://doi.org/10.3390/s23073678>

Academic Editors: Paolo Spagnolo and Alessia Saggese

Received: 3 March 2023

Revised: 29 March 2023

Accepted: 30 March 2023

Published: 2 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Person re-identification (Re-ID) is one of the computer vision tasks that aims to match a target individual across many camera perspectives. It has become an increasingly significant field of research in recent years, particularly in the area of surveillance and security. The main motivation for person Re-ID is to enable effective tracking of individuals in complex and crowded environments, such as airports, train stations, and public places [1,2]. However, the mission of person Re-ID faces several challenges that make it difficult to achieve high levels of accuracy. These challenges include variations in lighting conditions,

occlusions, changes in appearance, and perspective changes [3], as described in Figure 1. Additionally, person Re-ID is a large-scale and complex problem, as it requires searching through large databases of images to find the correct match.

Previous approaches to person Re-ID have included the use of hand-crafted features, metric learning algorithms, and deep learning [1,4,5]. The traditional methods contain hand-crafted feature extraction and distance metrics. Hand-crafted feature extraction is utilized to obtain more discriminative information from the image of a person, by using methods such as color histograms, texture features, scale-invariant feature transform (SIFT), local binary pattern (LBP), and other techniques. Metric learning algorithms have been used to match images using distance metrics, support vector machines (SVMs), neural networks (NN), cross-view quadratic discriminant analysis (XQDA), nearest neighbors (KNN), and other metric learning types [6,7], but this approach requires expert knowledge in feature design and is limited in its ability to capture complex relationships between images. Metric learning algorithms aim to learn a distance metric optimized for person Re-ID, but they still have limitations, such as difficulty learning an appropriate mapping for large-scale datasets. Deep learning, particularly convolutional neural networks (CNNs), has significantly improved the accuracy of person Re-ID algorithms, by learning a feature representation directly from the raw images. However, deep learning approaches also present new challenges, such as the need for large amounts of labeled data and the computational requirements of training large models. The best approach to person Re-ID depends on the specific requirements of the task, but deep learning has had a significant impact on the field.



**Figure 1.** Some difficult issues within the DukeMTMC dataset. (a) Occlusions, (b) illumination differences, (c) pose variations.

CNNs have proven to be an efficacious tool for addressing the issue of person Re-ID. They are capable of learning and capturing the discriminative features of the input images, and can be learned from end-to-end on large datasets [5]. Additionally, CNNs can be fine-tuned for specific datasets, making it possible to improve their performance in challenging scenarios [8]. By leveraging the ability of CNNs to automatically learn and extract features, person Re-ID algorithms have achieved significant improvements in accuracy, making them an important tool for overcoming the perplexing problem of person Re-ID.

In the past decade, person Re-ID has attracted a great deal of interest, due to its utility in a range of computer vision applications, such as video surveillance and person tracking [6]. Re-ID attempts to identify a person of interest across numerous, non-overlapping cameras. Recently proposed methods in person Re-ID tasks show good performance while using the attention mechanism, by focusing on more relevant characteristics [2,6]. In addition, most Re-ID methods depend on global features, that focus on the overall information in the image of a person and ignore the spatial structure of that person, so recently, many Re-ID methods have mainly extracted local features for re-identification, to improve the extracted features [9,10].

Despite the success of person recognition methods, identifying the same person in different cameras remains a difficult task, particularly in scenarios where the features of a person repeatedly change. To tackle this challenge, we provide a new person Re-ID method, that combines attention learning with a pre-trained model, which is a deep CNN that has already been trained to find informative and strong features in images, making the Re-ID process much easier and faster compared to models that are learned from scratch. Our system employs an attention mechanism that combines the position attention module (PAM) and the efficient channel attention (ECA). The PAM captures spatial, temporal, and channel context information, which improves the representation capability of the local features. The ECA reduces the model's complexity while maintaining accuracy, by capturing local cross-channel interactions.

Our contributions in this paper are twofold: (1) we introduce attention learning combined with a pre-trained model, for person Re-ID, which outperforms existing methods, and (2) we present an attention mechanism that combines the PAM and ECA, which improves the representation capability and decreases the complexity of the model, while preserving accuracy.

The remainder of the article is structured as follows: Section 2 covers relevant research in person Re-ID. Section 3 introduces the suggested method. Section 4 displays the research results. Section 5 shows the analysis study. Finally, Section 6 wraps up the paper and suggests future directions.

## 2. Related Work

In recent years, person Re-ID has become a crucial task in video observation, and has gained significant consideration in computer vision. Several approaches, including metric learning, hand-crafted features, and deep learning, have been proposed for this problem. In this part, we provide a summary of the most recent and relevant research in this area, with a focus on deep learning methods.

### 2.1. Hand-Crafted Feature-Based Person Re-ID

Manual feature extraction and metric learning design are person Re-ID's traditional methods; they rely on detecting low-level appearance features from the requisite image characteristics, such as shapes, colors, and textures [11]. Support vector machines (SVMs), neural networks (NN), nearest neighbors (KNN), and others, are metric learning types that minimize the distance between traits of the same person. Feature descriptors and metric learning are two independent stages. Liao et al. [9] presented a method that incorporates effective feature detection with metric learning. They suggested local maximal occurrence (LOMO) as a traits descriptor, that represents the image by extracting the histogram for colors using the texture histogram and sliding window with scale-invariant local ternary mode. Also, they used cross-view quadratic discriminant analysis (XQDA) for matching between features. Yang et al. [11] presented a method for extracting the features dependent on colors, that are called salient color names-based color descriptors (SCNCD), and they used the KISSME technique for metric learning. SCNCD divides the image into six parts equally and then computes the histogram for different spaces of color on all parts, to make the definitive extracted features sensitive to changes in illumination.

### 2.2. Hybrid Feature-Based Person Re-ID

The hybrid method combines deep learning with metric learning. The authors extract the features by utilizing a convolutional neural network and metric learning for classification. Saber et al. [6] used VGG-Net as a person representation, which provides a deep learning mechanism for person identification, and they selected the most estimated layers, to gain a useful feature description for the person. Subsequently, for person matching, a support vector classifier (SVC) was used, which eliminated the issue of using a small dataset. Jayapriya et al. [10] used CNN to extract traits from sequential information. This strategy combined the prioritized chromatic texture image (PCTimg) with the original

images, then entered them into the CNN to detect the traits. XQDA is employed for the classification. Wang et al. [12] developed a Siamese model, that employed XQDA to learn a discriminant metric, and extracted traits from deep networks to obtain spatiotemporal information about the person.

### 2.3. Deep Learned Feature-Based Person Re-ID

Deep learning is based on neural network algorithms and has become a prevalent offshoot of machine learning [13]. Deep learning algorithms employ multiple transformation layers with intricate constructions, in an effort to demonstrate high-level characteristics in data. In contrast to traditional methods, deep learning methods incorporate feature descriptions and similarity measures into an entire model. There are different kinds of architectures for deep learning-based methods, like attention-based methods and part-based methods.

Attention-based methods aim to carefully choose high-interest areas from input data, while disregarding other areas, with weak or no discriminative features. Attention modules concentrate on extracting regions with extremely distinguishing characteristics. Guodong et al. [14] proposed a hybrid architecture for CNN, that allows the network to concentrate on global and local discriminatory features for a person's image. They introduced a method called feature mask network (FMN). Wei et al. [15] established the global-local-alignment descriptor (GLAD) network, that appreciates the skeletons and splits the image by using the deeper cut. GLAD is intended to detect both local traits from separated images and global traits from the whole body. Masked graph attention network (MGAT) is a network designed by Bao et al. [16], that concentrates on the relationship between individual images and their labels, while ignoring the global mutual information present in the full sample set. The MGAT is dependent on a plenary network that extracts features, where nodes can concentrate on the characteristics of others in a directly navigable mode in the form of a mask matrix, with label information for guiding.

Part-based Re-ID approaches, elicit image areas to discover distinctive part-level features, established on accurate part-level cues that are often neglected when retrieving global traits. Part-based convolutional baseline (PCB) network was suggested in [17], which uses uniform segmentation on the convolution layer to interpret part-level data, by dividing the entire body into six horizontally running stripes in the feature map. Each component feature vector is supplied to a classifier, which generates an ID-prediction loss, that is independent for each part. Tian et al. [18] proposed a joint learning network that focuses on learning more distinctive and powerful features. They applied a global branch to learn the most distinctive global-level traits, and they divided the extracted map of traits into N parts, which are taken as inputs into a distinctive network that comprehends the local-level features. Afterward, they generate a local loss by combining N-part losses. They can then obtain a desirable total loss by combining local and global losses. A Siamese multiple granularity network (SMGN), with two major branches, was proposed by Li et al. [19], for learning the local and global characteristics of a person independently. The retrieved features of the two branches are combined as multiple features for personal images, and multiple loss functions are employed to enhance their performance.

From the above discussion, it is seen that previous studies have tried to enhance the person Re-ID performance using different methods. However, most of these methods have limitations, and do not perform well on large datasets. Our proposed method overcomes these limitations, by combining attention learning with a pre-trained model, which outperforms the existing methods on large datasets. The main difference between the proposed work and the related work, is that the proposed method combines the PAM and ECA, to extract features from temporal, spatial, and channel contexts. This is a new approach that has not been explored in previous studies. The proposed multi-part feature network, with the combination of PAM and ECA, has great potential to solve the problem of person Re-ID successfully, as it combines the benefits of temporal, spatial, and channel information. To summarize, the proposed method differs from previous studies, in that it combines the PAM and ECA to extract features from multiple contexts, with a high potential to achieve

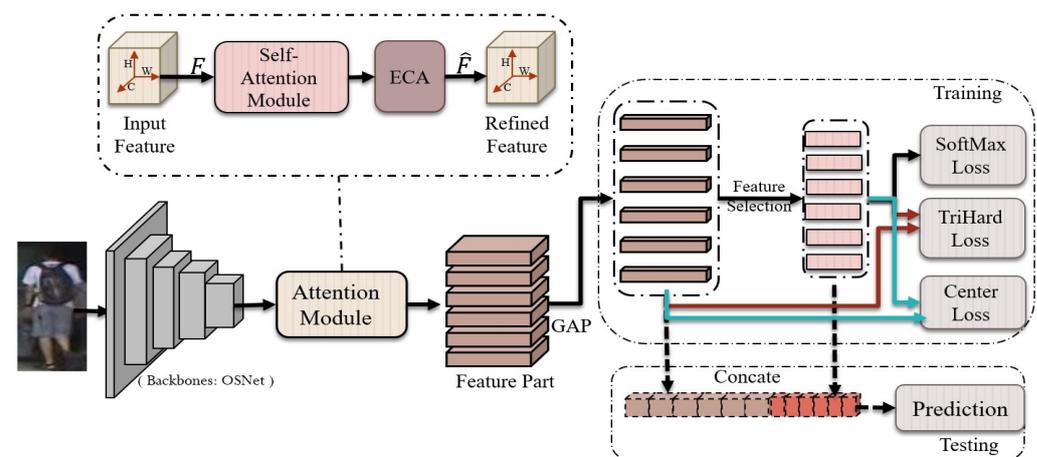
better results than existing methods. Table 1 summarizes the main differences between the proposed work and related work in the field of person Re-ID.

**Table 1.** Summary of the main differences between the proposed work and related work.

Approach	Main Focus	Techniques Used
Hand-crafted	Detecting low-level appearance features	SVM, NN, KNN, feature descriptors, metric learning
Hybrid	Combining deep learning with metric learning	CNN, metric learning, support vector classifiers
Deep learned	Using deep learning algorithms	Attention-based, part-based methods, CNN
Proposed	Multi-part, enhancing accuracy and robustness	Position attention module, efficient channel attention, multi-part feature network

### 3. Methodology

In this section, we depict the overall structure of a multi-part feature network for a person Re-ID task, that can independently learn extensive information from different parts of features, and the features from these parts can be merged for prediction. Then, we describe the two attention modules that are utilized to reduce the impact of irrelevant background, while concentrating on discriminative features of a person's appearance. Finally, we describe the loss functions that are utilized. OSNet [20] acts as the foundation for our network structure, as shown in Figure 2.



**Figure 2.** The architecture of the multi-part feature network.

#### 3.1. Baseline Configuration

We utilized OSNet [20] as a feature extractor for combining heterogeneous and homogeneous features, as well as a relatively lightweight network capable of developing performance, while avoiding over-fitting. OSNet [20] is built by stacking the bottleneck layer by layer, to decrease the parameter numbers, thereby lowering the computational cost.

#### 3.2. Position Attention Module

In the person scenario, we observed that distinctive trait representations are fundamental for person Re-ID, which may be achieved by understanding contextual information. To extract contextual information from local traits, we utilized a position attention module (PAM), which extracts much information derived from local characteristics, thereby improving their ability to represent the features.

The structure of the position attention module (PAM) [21], which is made for detecting and collecting the relevant pixels in the spatial domain, is depicted in Figure 3. The feature  $F \in R^{C \times H \times W}$ , where  $C$  is the number of channels,  $H$  is the spatial dimension height, and  $W$

is the spatial dimension width of an input tensor. We first feed the feature maps in the first branches into a convolution layer, to produce the new feature maps  $F1 \in R^{C/16 \times H \times W}$ , then we reshape  $F1$  to  $R^{C/16 \times N}$ , where  $N$  is the number of the pixels, which is equal to  $H \times W$ . To obtain  $F2$  for the second branch, we apply the same mechanism as for the first branch. Following that, we multiply the transpositions of  $F2$  and  $F1$  using matrix multiplication, and then utilize a softmax layer to compute the attention map  $S \in R^{N \times N}$ . Then, we execute matrix multiplication between  $S$  and the reshaping of the input feature, to get the feature to  $R \in R^{C \times H \times W}$ . Ultimately, the definitive output  $\in R^{C \times H \times W}$  is obtained by applying the batch normalization and then executing an operation of element-wise sum with the input features. Generally, in the original PAM, the third branch began with the 2D convolution layer, and we removed this layer to decrease the training time and increase the accuracy of our Re-ID method.

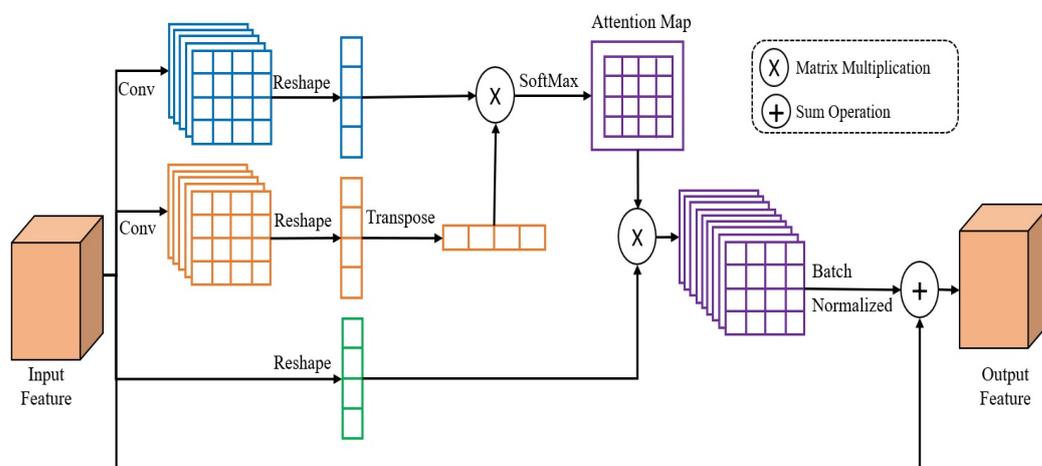


Figure 3. PAM attention module.

### 3.3. Efficient Channel Attention (ECA)

The channel attention module has shown significant potency to enhance the effectiveness of deep CNN. Channel attention is utilized to ameliorate the features of different channels, by simulating the significance of all channels in the feature. One of these channel attention modules is efficient channel attention (ECA). ECA detects interactions on the local cross-channel, by analyzing the channel and its neighbors. ECA minimizes the parameter numbers and reduces the model’s complexity, while maintaining precision.

ECA’s structure was proposed in [22]. To begin, as illustrated in Figure 4, a global average pooling (GAP) method is used, to reduce the size dimension of the input feature. After that, the weights of the channel are derived by a 1D convolution with a kernel size of three. Lastly, a sigmoid function is used, to obtain the final attention weights. Channels’ local interactive information can be reserved in this manner.

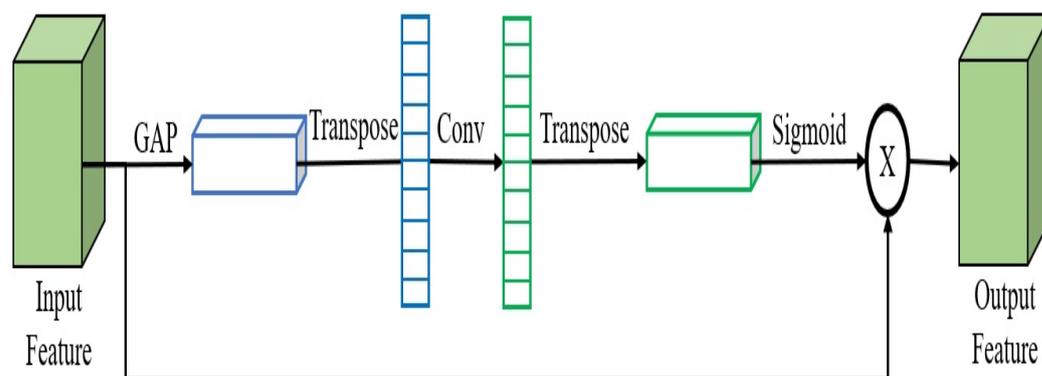


Figure 4. ECA attention module.

### 3.4. Loss Functions

As the final description for the person Re-ID features, we concatenate the feature vectors from the GAP and feature selection. Our loss function is gathered from ID loss (softmax loss) [23] for the six parts of the selected feature, and from a hard-margin triplet loss [24] and a center loss [25], for the concatenated feature. As demonstrated in Figure 2, each classifier forecasts the identification of the input image, namely,

$$L_{total} = \sum L_{id} + \beta \times L_{H-triplet} + \alpha \times L_{center} \quad (1)$$

where  $\beta$  and  $\alpha$  are weighting factors.

The cross-entropy loss (softmax loss) for the learned features,  $f_i$ , with label smoothing [23], is given as:

$$L_{id} = - \sum_{i=1}^N \times q_{yi} \times \log \frac{e^{(w_i \times f_i + b_i)}}{\sum_{j=1}^C e^{(w_j \times f_j + b_j)}} \quad (2)$$

where  $N$  is the batch size,  $C$  is the identity class number,  $f_i$  is the extracted feature,  $w_i$  and  $b_i$  are the weighted and bias for class  $i$ , respectively, and  $q_{yi}$  is the ground truth of the labels.

By obtaining many centers for all identity classes, hard triplet loss [24] outperforms softmax loss. However, the max function is required, to find the closest center for each identity class, and it is not smooth, thus the function can be sensitive between several centers. Smoothing of the max function in the softmax loss, can be utilized to enhance robustness. The hard triplet loss for the learned feature  $f_i$ , is given as:

$$L_{H-triplet} = - \sum_{i=1}^N \log \frac{e^{\lambda(S_{i,y_i} - \delta)}}{e^{\lambda(S_{i,y_i} - \delta)} + \sum_{j \neq y_i} e^{\lambda S_{i,j}}} \quad (3)$$

where  $\lambda$  is compensated to optimize a smoothed triplet loss,  $\delta$  is a predefined margin, and  $S_{(i,j)}$  is the similarity between feature  $f_i$  and the class  $j$ .

The center loss [25] is used to decrease intra-class variance between each sample in the mini-batch, while maintaining the features of the various classes separately. It can also reduce the distance within the class, so the compression of the samples within the class can be realized. The center loss function is written as follows:

$$L_{center} = \frac{1}{2} \sum_{i=1}^N \|f_i - C_{y_i}\|_2^2 \quad (4)$$

where  $f_i$  is the detected feature, and  $C_{y_i}$  is the updated deep feature.

## 4. Experimental Results and Discussion

In this section, we will carry out comprehensive experiments to confirm the viability of the suggested procedure. This section is arranged as follows: 1. provides three common datasets; 2. explains the specifics of implementation; 3. elucidates the protocols employed to test our strategy; and 4. compares the introduced approach to competing approaches on the relevant datasets.

### 4.1. The Utilized Datasets

To evolve and test the introduced model, we employed three diverse common datasets, as shown in Table 2, which are the fundamental datasets employed for the person Re-ID task.

CUHK03 [26]: was the first considerable dataset for a person Re-ID task. Images in this dataset contain the person detected by manual labeling and deformable part models (DPM). It contains 1467 identities, captured by two non-overlapping cameras.

Market-1501 [27]: was gathered by six separate cameras, at Tsinghua University. It contains 1501 identities, and images in this dataset contain the person detected by manual

labeling and deformable part models (DPM), it also has 2793 false images because of the DPM detector.

DukeMTMC [7]: is one of the large-scale datasets. Eight cameras were utilized in the DukeMTMC dataset, to track multiple targets. It contains 1812 persons, and the person in each image is manually labeled.

**Table 2.** Description of the employed datasets in this research.

Dataset	Year	Camera	Person	Images	Challenges
CUHK03	2014	2	1467	13,164	Alignment variation, occlusion, missing body parts
Market-1501	2015	6	1501	32,217	Variations in scale, illumination, and pose, occlusion, background noise, alignment variation, occlusion, missing body parts
DukeMTMC	2017	8	1812	36,441	

#### 4.2. Specifics of Implementation

Our introduced network was tested on a PC that uses NVIDIA RTX3060 12GB. OS-Net [20] was pre-trained on ImageNet [28], where we omitted the GAP layer and fully connected layers. To be more specific, all image sizes were changed to  $384 \times 128$  before being entered into the network. For training, our proposed network extracted features, and the optimizer of the network was the stochastic gradient descent (SGD) algorithm, with a learning rate of 0.04, decay rate of 0.1, and momentum of 0.9. In the training set, CUHK03 has 767 individuals and 7368 photos, while in the testing set, there are an additional 700 individuals and 6732 images. Market-1501 has 751 persons in the training set, with 12,936 images, and another 750 persons in the testing set, with 16,482 images. DukeMTMC has 702 identities, with 16,522 images in the training set, and another 702 identities in the testing set, with 16,426 images.

#### 4.3. Metric Protocol

We employed the single-shot approach in our experiment, which allows a thorough comparison. The cumulative matching characteristic (CMC) [29] and mean average precision (mAP) [30] were utilized to evaluate the person Re-ID performance. To improve performance even further, we added the re-ranking method [31], dependent on k-reciprocal encoding, to our method. The re-ranking operation was utilized in the testing phase.

#### 4.4. Evaluation on the Used Datasets

The introduced method appears to have excellent results compared to the preceding methods. Prior to discussing its accuracy on the three datasets, the introduced approach is evaluated against the state-of-the-art methods as follows:

**Market-1501 database:** Table 3 shows the competitive fineness results for the proposed technique and other person Re-ID methods, using the Market-1501 dataset. Our proposed method affords enhanced outcomes compared to the other methods. The introduced method achieves 95.93%, compared to the highest score achieved by DM-OSNet [8], of 95.61%. However, DRL-Net [32] achieves the highest mean average precision (mAP) score, of 89.9%, while the proposed method achieves a score of 87.57%. By utilizing the re-ranking [31], the proposed method achieves even higher results, with a rank-1 accuracy of 96.41% and an mAP of 94.15%. The results demonstrate that the proposed method performs well compared to other state-of-the-art techniques. However, there are still some limitations, as some methods perform better in certain aspects, such as DRL-Net for mAP. The findings of this study could have important implications for the development of more accurate person re-identification systems in real-world applications.

**DukeMTMC-reID database:** This has more challenges than in the Market-1501 dataset, due to the greater number of camera views and noisy backgrounds, that gives more variation within classes. Table 3 also presents the results of various person re-identification methods on the DukeMTMC dataset, including our proposed method, with and without re-ranking. In terms of rank-1 accuracy, our introduced method achieves 89.77%, compared to the highest score achieved by AET-Net [33], of 89.5%. However, AET-Net [33] achieves the highest mean average precision (mAP) score, 80.1%, while the proposed method achieves a score of 78.62%. Using the re-ranking technique, our proposed method improved its performance to 94.08% and 92.22%, in rank-1 and mAP, respectively.

**CUHK03 database:** Table 4 presents the results of various person re-identification methods on the CUHK03 dataset, including our proposed method, with and without re-ranking. Our proposed method achieves impressive performance on both labeled and detected types, outperforming other methods by 3.01% and 2.65%, respectively. Moreover, the use of the re-ranking technique results in a substantial improvement in performance, with an increase of 18% for labeled and 17.51% for detected types. These results demonstrate the effectiveness of our proposed method and highlight its potential to improve on state-of-the-art person re-identification methods. When comparing our proposed method to other state-of-the-art methods, it is clear that our approach presents several strengths. For instance, our method outperforms the widely used PCB [17] method by a significant margin, achieving an improvement of 10.55% in rank-1 accuracy for detected types. Additionally, our proposed method outperforms the HAN [34] method by 26.71% in mAP, for labeled types. However, our method does have some limitations, such as being computationally expensive, due to the high-dimensional feature extraction required. Despite these limitations, our proposed method demonstrates superior performance, and the results indicate that it has the potential to be a useful tool for person re-identification in real-world scenarios.

**Table 3.** Network technical change comparison on the Market-1501 dataset.

Method	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
PCB [17]	92.3	77.4	81.8	66.1
RJLN [18]	93.7	81.9	85.5	73.1
PIE [35]	87.33	69.25	80.84	64.09
AlignedReID++ [36]	91.8	79.1	82.1	69.7
PGFA [37]	91.2	76.8	82.6	65.5
Deep Person [38]	92.31	79.58	80.90	64.80
FMN [14]	85.99	67.12	74.51	56.88
FMN+re-rank [14]	87.92	80.62	79.52	72.79
FPO [39]	91.81	79.23	81.0	78.0
DCNN [40]	90.2	82.7	80.6	64.1
HAN [34]	91.6	76.7	80.7	65.5
UANet [41]	91.3	76.5	82.1	65.2
UnityStyle [42]	91.8	76.5	80.38	64.32
SMGN [19]	94.1	79.2	86.1	75.3
SMGN + re-rank [19]	95.5	80.3	87.1	76.0
GCN [43]	88.65	74.15	-	-
ARFM [44]	88.02	76.13	81.53	65.94

Table 3. Cont.

Method	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
AL-APR [45]	89.01	74.38	80.52	63.67
DUNet [46]	91.6	75.90	82.1	66.5
NFML [47]	95.3	86.4	89.2	76.2
EDAGAN [48]	85.36	64.52	74.19	51.90
CooRL [49]	89.5	74.3	78.9	65.2
Tri-GCN [2]	92.98	80.5	83.23	66.8
HOB-net [50]	94.7	86.3	88.2	77.2
SFBM [51]	95.3	85.4	88.6	74.5
VACNet [33]	95.1	86.1	89.5	77.1
twinsReID [52]	93.7	85.4	88.6	78.2
DM-OSNet [8]	95.61	87.36	89.18	78.26
TAFN [53]	94.7	86.2	85.9	74.8
MS-LS-RK [54]	92.3	88.3	86.5	81.7
AM0BH [55]	94.6	85.9	89.2	76.7
RANGEv2 [56]	94.7	86.8	87.0	78.2
DRL-Net [32]	94.7	89.9	88.1	76.6
AET-Net [33]	94.8	87.5	89.5	80.1
Our method	<b>95.93</b>	<b>87.57</b>	<b>89.77</b>	<b>78.62</b>
Our method+re-rank	<b>96.41</b>	<b>94.15</b>	<b>94.08</b>	<b>92.22</b>

Table 4. Network technical change comparison on the CUHK03 dataset (labeled and detected).

Method	CUHK03			
	Labeled		Detected	
	Rank-1	mAP	Rank-1	mAP
PCB [17]	-	-	61.3	54.2
PIE [35]	-	-	45.88	41.21
RJLN [18]	-	-	66.6	60.9
FMN [14]	41.0	38.1	42.6	39.2
FMN+re-rank [14]	46.0	47.6	47.5	48.5
HAN [34]	46.5	46.1	47.5	45.5
FPO [39]	65.60	60.16	63.07	56.31
UANet [41]	62.6	57.7	58.9	52.6
DUNet [46]	54.6	52.2	51.6	49.9
Tri-GCN [2]	68.29	61.59	65.86	58.21
HOB-net [50]	70.2	67.5	69.2	66.8
RANGEv2 [56]	64.3	67.4	61.6	64.6
Our method	<b>73.21</b>	<b>67.34</b>	<b>71.85</b>	<b>66.16</b>
Our method+re-rank	<b>91.21</b>	<b>91.4</b>	<b>89.36</b>	<b>89.29</b>

## 5. Research Analysis

In this section, we analyze the parameters of the Market-1501 dataset, including the effect of image size, the number of image parts, batch size, loss type, attention module type, and epoch number.

### 5.1. Comparison of Loss Function Change

In the training stage, our loss function gathers from cross-entropy, the hard triplet loss, and the center loss. To inspect the effect of the loss function, we performed experiments in which we performed cross-entropy loss with the triplet loss, center loss, or a combination of them, to confirm the efficacy of employing multiple losses. Table 5 showcases the performance of the proposed multiple loss function combinations on the three different datasets—Market-1501, DukeMTMC, and CUHK03 (both labeled and detected). As seen in the experimental results, utilizing many losses causes the network to exhibit varying degrees of accuracy enhancement on the three datasets, when compared to using only the softmax loss. For instance, using the combination of losses, outperformed the competition by 0.99% and 1.11% in rank-1 and mAP, respectively, on the DukeMTMC dataset. Similarly, the CUHK03 dataset increased by 1.14% and 1.64%, for labeled and detected sets, respectively, using the proposed method. In combination, the loss functions are fused, making them interactive, resulting in improved performance at the cost of speed, and the network converges towards greater performance. Generally, the results demonstrate that the proposed method is effective in enhancing the accuracy of the network and has the potential to improve state-of-the-art person re-identification.

**Table 5.** Performance of Re-ID models under different loss functions (× loss not employed, ✓ loss employed).

Loss			Market-1501		DukeMTMC		CUHK03			
	Tri-Hard	Center	Rank-1	mAP	Rank-1	mAP	Labeled		Detected	
Cross-Entropy			Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
✓	×	×	94.86	86.17	88.78	77.6	72.07	66.23	70.21	64.92
✓	×	✓	95.01	86.32	89.00	77.79	72.25	66.41	70.34	65.20
✓	✓	×	95.74	87.12	89.53	78.21	72.94	66.91	71.66	65.79
✓	✓	✓	95.93	87.57	89.77	78.62	73.21	67.34	71.85	66.16

### 5.2. Comparison of Attention Change

Many cutting-edge methodologies for person Re-ID tasks, make use of attention modules. To extract global features, we added the attention module, which consists of PAM and ECA, into the network. To investigate the efficiency of the suggested attention module in our framework, we conducted experiments on the Market-1501 dataset. Six structures are compared: only the network without the attention, the network with only one attention module (PAM or ECA), the network with changing the order of the attention modules (PAM after or before ECA), the network with the average of the attention modules (PAM and ECA), and the complete network. Table 6 shows the experimental results for the Market-1501 dataset, comparing the use of attention mechanisms with different configurations, against a baseline without attention mechanisms. The configurations of the attention module are denoted by the numbers in parentheses. As seen in the table, the use of attention mechanisms improves the network's performance, achieving higher rank-1 and mAP scores compared to the baseline. Specifically, the best performing configuration is (1) (2), which utilizes both PAM and ECA attention mechanisms, achieving a rank-1 score of 95.93% and mAP of 87.57%, which represents a significant improvement, of 2.42% and 6.92%, respectively, compared to the baseline. Additionally, the results show

that the PAM attention mechanism contributes more to the improvement than the ECA attention mechanism. Configuration (1)—which uses only PAM—achieved a higher rank-1 score than configuration (2) (1), which uses only ECA. This suggests that PAM is more effective in capturing long-range dependencies between features. The experimental results demonstrate the effectiveness of using attention mechanisms in improving the performance of person re-identification networks. The use of PAM and ECA attention mechanisms with appropriate configurations, can significantly improve the rank-1 and mAP scores, which are important performance metrics for person re-identification systems.

**Table 6.** Quantitative comparison of the attention module type on the Market-1501 dataset (where the number is the order of the attention modules).

Attention		Market-1501	
PAM	ECA	Rank-1	mAP
-	-	93.04	80.04
(1)	-	95.81	86.97
-	(1)	93.13	80.11
(1)	(2)	95.93	87.57
(2)	(1)	95.54	86.3
(1)	(1)	95.46	86.96

### 5.3. Comparison of Using Different Pre-Trained Models

To investigate the usefulness of the baseline that we chose, we compare the results of several baselines on different datasets. Our baselines for comparison are various versions of OSNet [20] and VGG16. Table 7 presents the experimental results obtained for the Market-1501, DukeMTMC, CUHK03-labeled, and CUHK03-detected datasets. Each row of the table corresponds to a different baseline network, while each column shows the rank-1 and mAP scores for a specific dataset. Our results demonstrate that the addition of suggested branches improves the performance of all baseline networks, with the most significant gains observed in the  $OSNet_{X_1}$  network. In particular, our method achieved a rank-1 accuracy of 95.93% and a mAP score of 87.57% on the Market-1501 dataset, outperforming all other baseline networks. Our results also show that the VGG16 baseline network performed relatively poorly, with a rank-1 accuracy of only 90.25% and a mAP score of 74.86%. When comparing the results of each baseline network to the proposed method, it is evident that our method outperformed all baseline networks on all datasets, except for DukeMTMC, where  $OSNet_{X_1}$ , with our suggested addition, achieved the best performance. These results highlight the effectiveness of our proposed method in enhancing the performance of existing baseline networks. Furthermore, we observed some interesting trends and patterns in our data. For example, we found that  $OSNet_{X_1}$  performed significantly better than other OSNet baselines after adding all the branches, except for rank-1 for DukeMTMC. Additionally, the  $OSNet_{X_{0.75}}$  and  $OSNet_{X_1}$  networks achieved higher performance than  $OSNet_{X_{0.5}}$  and  $OSNet_{X_{0.25}}$ , respectively, suggesting that larger networks may better capture complex features in person re-identification. Overall, our study provides valuable insights into the effectiveness of our proposed method and the relative performance of different baseline networks in person re-identification.

**Table 7.** Performance of proposed strategy of RE-ID under different baselines.

Baseline	Market-1501		DukeMTMC		CUHK03-Labeled		CUHK03-Detected	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
VGG16	90.25	74.86	74.92	55.47	57.21	51.44	51.07	45.39
OSNet_IBN_X_1	95.00	84.54	89.06	76.54	68.36	63.62	65.79	60.23
OSNet_X_0.25	88.8	72.61	81.83	64.14	55.71	49.99	51.64	46.90
OSNet_X_0.5	93.4	82.64	87.98	73.19	69.5	63.06	64.29	58.17
OSNet_X_0.75	95.42	86.18	89.96	77.2	72.93	67.35	68.57	62.92
OSNet_X_1	95.93	87.57	89.77	78.62	73.21	67.34	71.85	66.16

#### 5.4. Comparison of Network Architectural Change

To further interpret the results presented in Table 8, we can observe that the introduced strategy for applying the attention module after layer 4, provided the best performance in terms of rank-1, rank-5, rank-10, rank-20, and mAP scores. This indicates that the retrieved feature should include both coarse and fine information for a person's representation, to make the attention module more successful. The results also show a clear trend of increasing performance with deeper layers, as adding the attention module after layers 3 and 4 improves performance, compared to adding it after layer 2. Moreover, the rank-1 score of 95.93%, achieved by applying the attention module after layer 4, is particularly noteworthy, as it represents a significant improvement over the other positions tested. These results demonstrate the effectiveness of the proposed strategy for integrating attention mechanisms into person re-identification models and suggest that future work in this area should explore the use of attention modules in conjunction with deeper network architectures.

**Table 8.** Comparison of performance when changing attention position.

Position	Market-1501				
	Rank-1	Rank-5	Rank-10	Rank-20	mAP
Layer 2	87.55	94.71	96.17	97.95	68.6
Layer 3	93.72	97.39	98.52	99.17	82.17
Layer 3 and 4	94.2	97.45	98.4	99.05	82.12
Layer 4	95.93	98.13	99.02	99.44	87.57

#### 5.5. Comparison of Image Size Change

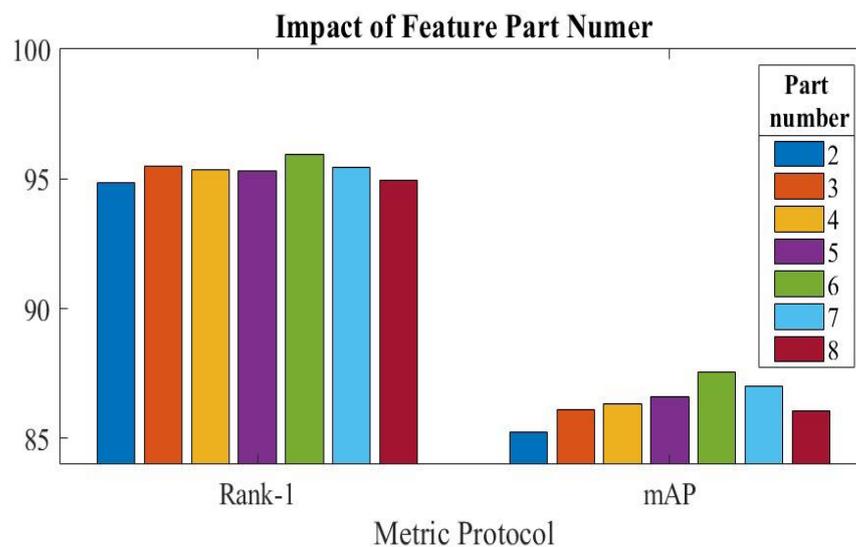
To better understand the impact of image size on the performance of the proposed method, we conducted experiments using different image sizes, and evaluated the results in terms of rank-1, rank-5, rank-10, rank-20, and mAP scores, as shown in Table 9. It can be seen, that resizing the image to  $384 \times 128$  provided the best performance in terms of rank-1 accuracy, with a score of 95.93%. The other image sizes had rank-1 scores ranging from 95.24% to 95.86%. This suggests that a larger image size can capture more detailed information about the person's appearance, leading to better recognition performance. It is noteworthy that the choice of image size can also impact the overall computational cost of the system, and this factor should be considered when selecting the optimal image size for a given application.

**Table 9.** Performance of Re-ID models under changing image sizes.

Image Size		Market-1501				
Height	Width	Rank-1	Rank-5	Rank-10	Rank-20	mAP
256	128	95.51	98.01	99.02	99.32	86.58
	192	95.24	98.31	98.9	99.35	86.12
320	128	95.25	98.13	98.96	99.32	87.97
	192	95.6	98.4	99.05	99.38	86.55
384	128	95.93	98.13	99.02	99.44	87.57
	192	95.86	98.31	99.05	99.44	87.15

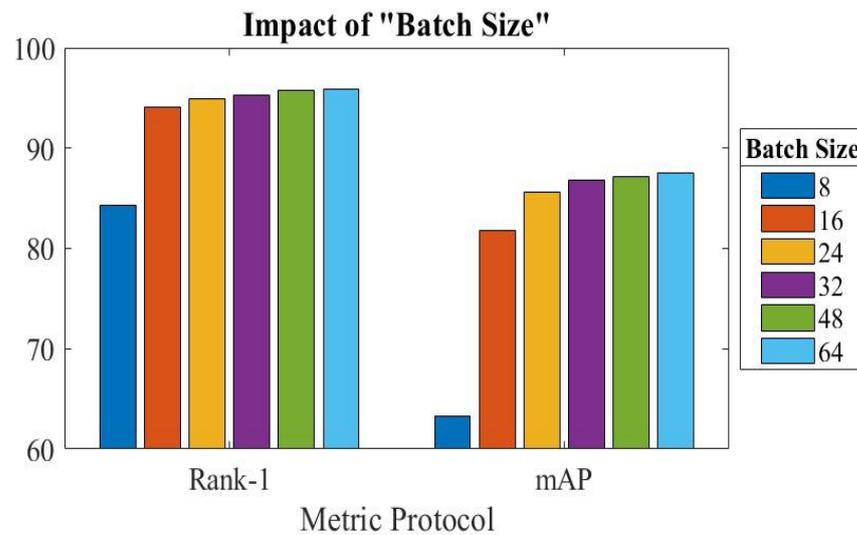
### 5.6. Comparison of Feature Part Number

We explored the impact of the part number of feature selection on overall Re-ID performance and tested it using the Market-1501 dataset. We attempted to train the provided model with a varied number of feature selection components. The output feature is a global feature if the part number is set to 1. Having six parts exhibits the best performance on the Market-1501 dataset, according to the results presented in Figure 5. The Re-ID performance begins to fall with adding further parts, indicating that too many components load the model training and therefore lower performance.

**Figure 5.** Performance of our proposed models under different feature part numbers.

### 5.7. Comparison of Batch Size Change

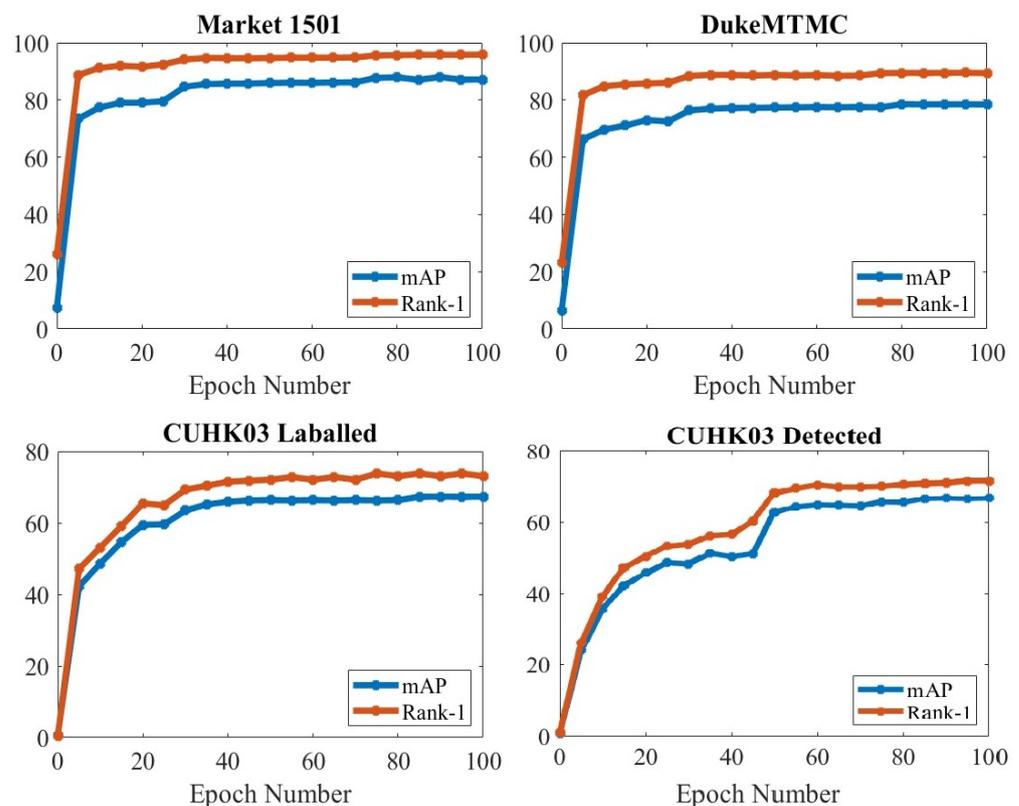
Here, we examined the effects of modifying the batch size in the training stage, where the batch size represents the number of images fed into the network. To examine the impacts of various batch sizes on the efficiency of our introduced network, comparative experiments were conducted. The largest batch size that could be used was 64, because of GPU limitations. Figure 6 illustrates the results of the experiment. As seen, performance changes as the batch size changes. The accuracy of the Market-1501 dataset may reach its highest value when the batch size is 64. Comparing the improvement to a batch size of 48, it is slight. As a result, performance varies by altering the batch size, and accuracy will continue to improve. We draw the conclusion that the processing of the samples' derived features can be helped by increasing the batch size.



**Figure 6.** Effect of changing size of batch on Market-1501.

### 5.8. Impact of Numbers of Epoch

The empirical results for our introduced network during the training stage are illustrated in Figure 7, to test the effect of changing the number of epochs. Three different datasets were utilized in the experiment. This experiment comprised 100 training epochs and was evaluated every 5 training epochs. As illustrated in Figure 7, both rank-1 and mAP performance improve by increasing epoch numbers in the training stage, but the difference is slight until epoch 35 in the Market-1501, DukeMTMC, and CUHK03-labeled datasets, unlike the CUHK03-detected dataset, which needs to reach epoch 55 before the change becomes small.



**Figure 7.** Effect of different numbers of epochs on the Market-1501, DukeMTMC, and CUHK03 (labeled and detected) datasets.

## 6. Conclusions

This research presents a multi-part feature network for individual Re-ID, which combines the position attention module with efficient channel attention, to improve the robustness and discrimination of the features. The suggested attention mechanism utilizes temporal, spatial, and channel context information, to extract a broader variety of contextual information from local features, hence enhancing their capacity for representation. Under the restrictions of numerous losses, the methods we propose can produce resilient feature representations. Extensive testing on three datasets revealed that the proposed strategy outperformed state-of-the-art techniques and was highly generalizable. The results indicate that the suggested strategy enhances both quantitative and qualitative methods for re-identifying individuals. In the future, we intend to investigate and expand the introduced method, to improve the precision and efficacy of person Re-ID.

**Author Contributions:** Conceptualization, S.S. and M.H.; methodology, S.S.; software, S.S. and M.H.; validation, K.A. and P.P.; formal analysis, M.H. and S.M.; investigation, S.S.; resources, S.M.; data curation, M.H.; writing—original draft preparation, S.S. and M.H.; writing—review and editing, S.M., P.P., and K.A.; visualization, M.H.; supervision, K.A. and M.H.; project administration, M.H. and S.M.; funding acquisition, S.M. and P.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number PNURSP2023R196, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to acknowledge the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R196), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ming, Z.; Zhu, M.; Wang, X.; Zhu, J.; Cheng, J.; Gao, C.; Yang, Y.; Wei, X. Deep learning-based person re-identification methods: A survey and outlook of recent works. *Image Vis. Comput.* **2022**, *119*, 104394. [[CrossRef](#)]
2. Saber, S.; Amin, K.; Pławiak, P.; Tadeusiewicz, R.; Hammad, M. Graph convolutional network with triplet attention learning for person re-identification. *Inf. Sci.* **2022**, *617*, 331–345. [[CrossRef](#)]
3. Wang, J.; Li, P.; Zhao, R.; Zhou, R.; Han, Y. CNN Attention Enhanced ViT Network for Occluded Person Re-Identification. *Appl. Sci.* **2023**, *13*, 3707. [[CrossRef](#)]
4. Lavi, B.; Serj, M.F.; Ullah, I. Survey on deep learning techniques for person re-identification task. *arXiv* **2018**, arXiv:1807.05284.
5. Liu, M.; Zhang, Y.; Li, H. Survey of Cross-Modal Person Re-Identification from a Mathematical Perspective. *Mathematics* **2023**, *11*, 654. [[CrossRef](#)]
6. Saber, S.; Amin, K.M.; Adel Hammad, M. An efficient person re-identification method based on deep transfer learning techniques. *Int. J. Comput. Inf.* **2021**, *8*, 94–99. [[CrossRef](#)]
7. Gou, M.; Karanam, S.; Liu, W.; Camps, O.; Radke, R.J. Dukemtmc4reid: A large-scale multi-camera person re-identification dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 10–19.
8. Zhou, Y.; Liu, P.; Cui, Y.; Liu, C.; Duan, W. Integration of Multi-Head Self-Attention and Convolution for Person Re-Identification. *Sensors* **2022**, *22*, 6293. [[CrossRef](#)] [[PubMed](#)]
9. Liao, S.; Hu, Y.; Zhu, X.; Li, S.Z. Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 July 2015; pp. 2197–2206.
10. Jayapriya, K.; Jacob, I.J.; Mary, N.A.B. Person re-identification using prioritized chromatic texture (PCT) with deep learning. *Multimed. Tools Appl.* **2020**, *79*, 29399–29410. [[CrossRef](#)]
11. Yang, Y.; Yang, J.; Yan, J.; Liao, S.; Yi, D.; Li, S.Z. Salient color names for person re-identification. In Proceedings of the Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; pp. 536–551.

12. Wang, S.; Zhang, C.; Duan, L.; Wang, L.; Wu, S.; Chen, L. Person re-identification based on deep spatio-temporal features and transfer learning. In Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 24–29 July 2016; pp. 1660–1665.
13. Wang, K.; Wang, H.; Liu, M.; Xing, X.; Han, T. Survey on person re-identification based on deep learning. *CAAI Trans. Intell. Technol.* **2018**, *3*, 219–227. [[CrossRef](#)]
14. Ding, G.; Khan, S.; Tang, Z.; Porikli, F. Feature mask network for person re-identification. *Pattern Recognit. Lett.* **2020**, *137*, 91–98. [[CrossRef](#)]
15. Wei, L.; Zhang, S.; Yao, H.; Gao, W.; Tian, Q. GLAD: Global–local-alignment descriptor for scalable person re-identification. *IEEE Trans. Multimed.* **2018**, *21*, 986–999. [[CrossRef](#)]
16. Bao, L.; Ma, B.; Chang, H.; Chen, X. Masked graph attention network for person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019.
17. Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; Wang, S. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 480–496.
18. Tian, Y.; Li, Q.; Wang, D.; Wan, B. Robust joint learning network: Improved deep representation learning for person re-identification. *Multimed. Tools Appl.* **2019**, *78*, 24187–24203. [[CrossRef](#)]
19. Li, D.X.; Fei, G.Y.; Teng, S.W. Learning large margin multiple granularity features with an improved siamese network for person re-identification. *Symmetry* **2020**, *12*, 92. [[CrossRef](#)]
20. Zhou, K.; Yang, Y.; Cavallaro, A.; Xiang, T. Learning generalisable omni-scale representations for person re-identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5056–5069. [[CrossRef](#)] [[PubMed](#)]
21. Chen, T.; Ding, S.; Xie, J.; Yuan, Y.; Chen, W.; Yang, Y.; Ren, Z.; Wang, Z. Abd-net: Attentive but diverse person re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8351–8361.
22. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
23. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
24. Qian, Q.; Shang, L.; Sun, B.; Hu, J.; Li, H.; Jin, R. Softtriple loss: Deep metric learning without triplet sampling. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6450–6458.
25. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 499–515.
26. Li, W.; Zhao, R.; Xiao, T.; Wang, X. Deepreid: Deep filter pairing neural network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 152–159.
27. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1116–1124.
28. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
29. Gray, D.; Brennan, S.; Tao, H. Evaluating appearance models for recognition, reacquisition, and tracking. In Proceedings of the IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), Clearwater Beach, FL, USA, 15–17 January 2007; Volume 3, pp. 1–7.
30. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2009**, *88*, 303–308. [[CrossRef](#)]
31. Zhong, Z.; Zheng, L.; Cao, D.; Li, S. Re-ranking person re-identification with k-reciprocal encoding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1318–1327.
32. Jia, M.; Cheng, X.; Lu, S.; Zhang, J. Learning disentangled representation implicitly via transformer for occluded person re-identification. *IEEE Trans. Multimed.* **2022**. [[CrossRef](#)]
33. Shi, W.; Zhang, Y.; Zhu, S.; Liu, Y.; Coleman, S.; Kerr, D. VAC-Net: Visual Attention Consistency Network for Person Re-identification. In Proceedings of the 2022 International Conference on Multimedia Retrieval, Newark, NJ, USA, 27–30 June 2022; pp. 571–578.
34. Li, W.; Zhu, X.; Gong, S. Scalable person re-identification by harmonious attention. *Int. J. Comput. Vis.* **2020**, *128*, 1635–1653. [[CrossRef](#)]
35. Zheng, L.; Huang, Y.; Lu, H.; Yang, Y. Pose-invariant embedding for deep person re-identification. *IEEE Trans. Image Process.* **2019**, *28*, 4500–4509. [[CrossRef](#)]
36. Luo, H.; Jiang, W.; Zhang, X.; Fan, X.; Qian, J.; Zhang, C. Alignedreid++: Dynamically matching local information for person re-identification. *Pattern Recognit.* **2019**, *94*, 53–61. [[CrossRef](#)]
37. Miao, J.; Wu, Y.; Liu, P.; Ding, Y.; Yang, Y. Pose-guided feature alignment for occluded person re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 542–551.

38. Bai, X.; Yang, M.; Huang, T.; Dou, Z.; Yu, R.; Xu, Y. Deep-person: Learning discriminative deep features for person re-identification. *Pattern Recognit.* **2020**, *98*, 107036. [[CrossRef](#)]
39. Tang, Y.; Yang, X.; Wang, N.; Song, B.; Gao, X. Person re-identification with feature pyramid optimization and gradual background suppression. *Neural Netw.* **2020**, *124*, 223–232. [[CrossRef](#)]
40. Li, Y.; Jiang, X.; Hwang, J.N. Effective person re-identification by self-attention model guided feature learning. *Knowl.-Based Syst.* **2020**, *187*, 104832. [[CrossRef](#)]
41. Jeong, D.; Park, H.; Shin, J.; Kang, D.; Paik, J. Uniformity attentive learning-based Siamese network for person re-identification. *Sensors* **2020**, *20*, 3603. [[CrossRef](#)]
42. Liu, C.; Chang, X.; Shen, Y.D. Unity style transfer for person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 6887–6896.
43. Kim, G.; Shu, D.W.; Kwon, J. Robust person re-identification via graph convolution networks. *Multimed. Tools Appl.* **2021**, *80*, 29129–29138. [[CrossRef](#)]
44. Zhu, S.; Zhang, Y.; Coleman, S.; Wang, S.; Li, R.; Liu, S. Semi-supervised learning for person re-identification based on style-transfer-generated data by CycleGANs. *Mach. Vis. Appl.* **2021**, *32*, 1–16. [[CrossRef](#)]
45. Chang, H.; Qu, D.; Wang, K.; Zhang, H.; Si, N.; Yan, G.; Li, H. Attribute-guided attention and dependency learning for improving person re-identification based on data analysis technology. *Enterp. Inf. Syst.* **2021**, *17*, 1941274. [[CrossRef](#)]
46. Li, R.; Zhang, B.; Teng, Z.; Fan, J. A divide-and-unite deep network for person re-identification. *Appl. Intell.* **2021**, *51*, 1479–1491. [[CrossRef](#)]
47. Wang, J.; Zhang, J.; Wen, X. Non-full multi-layer feature representations for person re-identification. *Multimed. Tools Appl.* **2021**, *80*, 17205–17221. [[CrossRef](#)]
48. Wang, Y.; Jiang, K.; Lu, H.; Xu, Z.; Li, G.; Chen, C.; Geng, X. Encoder-decoder assisted image generation for person re-identification. *Multimed. Tools Appl.* **2022**, *81*, 10373–10390. [[CrossRef](#)]
49. Peng, J.; Jiang, G.; Wang, H. Cooperative Refinement Learning for domain adaptive person Re-identification. *Knowl.-Based Syst.* **2022**, *242*, 108349. [[CrossRef](#)]
50. Chen, D.; Wu, P.; Jia, T.; Xu, F. HOB-net: High-order block network via deep metric learning for person re-identification. *Appl. Intell.* **2022**, *2022*, 1–14. [[CrossRef](#)]
51. Zhang, J.; Yang, W. Spatial Foreground Bigraph Matching for Generalizable Person Re-identification. In Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2022: 31st International Conference on Artificial Neural Networks, Bristol, UK, 6–9 September 2022; pp. 273–285.
52. Jin, K.; Zhai, J.; Gao, Y. TwinsRelD: Person re-identification based on twins transformer’s multi-level features. *Math. Biosci. Eng.* **2023**, *20*, 2110–2130. [[CrossRef](#)] [[PubMed](#)]
53. Ahn, H.; Hong, Y.; Choi, H.; Gwak, J.; Jeon, M. Tiny Asymmetric Feature Normalized Network for Person Re-Identification System. *IEEE Access* **2022**, *10*, 131318–131330. [[CrossRef](#)]
54. Chen, H.; Zhao, Y.; Zhang, L. Person Re-identification Based on CNN with Multi-scale Contour Embedding. In Proceedings of the Artificial Intelligence: Second CAAI International Conference, CICA 2022, Beijing, China, 27–28 August 2022; pp. 560–571.
55. Sabri, S.I.; Randhawa, Z.A.; Doretto, G. Joint Discriminative and Metric Embedding Learning for Person Re-identification. In Proceedings of the Advances in Visual Computing: 17th International Symposium, ISVC 2022, San Diego, CA, USA, 3–5 October 2022; pp. 165–178.
56. Wu, G.; Zhu, X.; Gong, S. Learning hybrid ranking representation for person re-identification. *Pattern Recognit.* **2022**, *121*, 108239. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.