

Article

Improving Person Re-Identification with Distance Metric and Attention Mechanism of Evaluation Features

Jieqian Zhou

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China; zhoujieqian@bupt.edu.cn

Abstract: In the present study, we developed a person re-identification network called the Multiple Granularity Attention Cosine Network (MGAC). MGAC utilizes the Multiple Granularity Network (MGN), which combines global and local features and constructs an attention mechanism to add to MGN to form a Multiple Granularity Attention Network (MGA). With the attention mechanism, which focuses on important features, MGA assesses the importance of learned features, resulting in higher scores for important features and lower scores for distracting features. Thus, identification accuracy is increased by enhancing important features and ignoring distracting features. We performed experiments involving several classical distance metrics and selected cosine distance as the distance metric for MGA to form the MGAC re-identification network. In experiments on the Market-1501 mainstream dataset, MGAC exhibited high identification accuracies of 96.2% and 94.9% for top-1 and mAP, respectively. The results indicate that MGAC is an effective person re-identification network and that the attention mechanisms and cosine distance can significantly increase the person re-identification accuracy.

Keywords: person re-identification; attention mechanism; distance metric; global features; local features



Citation: Zhou, J. Improving Person Re-Identification with Distance Metric and Attention Mechanism of Evaluation Features. *Electronics* **2023**, *12*, 4298. <https://doi.org/10.3390/electronics12204298>

Academic Editors: Dah-Jye Lee and Manohar Das

Received: 30 July 2023

Revised: 25 September 2023

Accepted: 8 October 2023

Published: 17 October 2023



Copyright: © 2023 by the author. 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 is a technology that enables users to retrieve images of the same person's identity among gallery by giving a query image of the person [1]. It generally involves feature extraction of the input image, distance metrics of the extracted features, and similarity ranking based on the distance value. A shorter distance between the query image and the gallery images corresponds to a higher degree of similarity and a higher likelihood of it being the person of the same identity [2]. Therefore, feature extraction and distance metrics of images are two important aspects of person re-identification [3–8]. In [9], the importance of good features for identification was reported.

Variations in illumination, occlusion, pose, camera settings, viewpoint, and background clutter in the images for person re-identification pose difficulties and challenges to the implementation of person re-identification [10–30]. In practical applications, the images used for person re-identification are captured by the detection box detected by the detector, not the ground truth box images. This may lead to the potential for inaccurate image detection, poor detection quality, and background clutter, making the implementation of person re-identification more difficult and challenging [31,32]. Re-identification networks, which are usually based on global feature learning, can learn significant appearance features but tend to ignore detailed information. In contrast, re-identification networks that are based on learning local features usually focus on learning detailed feature information but tend to ignore information on global appearance. We selected the multiple granularity network (MGN) [33] as the method for person re-identification, combining global and local feature information to achieve high identification accuracy. A high identification accuracy can be achieved by the MGN; however, it has been discovered through the analysis of images

and the MGN that the MGN can also be improved with distance metric and attention mechanism of evaluation features. An analysis of images from the Market-1501 [31] dataset revealed that because the dataset uses images cropped by detection boxes, the quality of the person framed in the images varies, as do his/her posture, clothing, and the items that he/she carries. The importance of each part of the features for identification also varies. For the MGN that extracts fixed partitioning features, the importance of information extracted at different granularities sometimes varies. With regard to distance metrics, the MGN uses the classical Euclidean distance. Other classical distance metrics [34–41] include the squared Euclidean distance, the Mahalanobis distance, the correlation distance, and the cosine distance, all of which have advantages and disadvantages.

In this study, the MGN [33] was used to extract features using ResNet-50 [42] as the backbone network, and the extracted features were continued with three branches of different granularity. The upper branch extracted fixed global features, the middle branch extracted fixed horizontal bisected features, and the lower branch extracted fixed horizontal trisected features. These different types of semantic granularity information were then combined to build the feature information for person re-identification. The MGN combined global features with local features, which increased identification accuracy. Based on the MGN, we also developed an attention mechanism to learn the importance of different granularity information to global information. Prior to this granularity of information being used for identification, these features were evaluated with the scores learned by the attention mechanism so that the important information would have a high score. All the granularity features were then concatenated for person re-identification to increase the identification accuracy. This type of identification, with the addition of attention mechanisms, is called MGA. Based on MGA, we found a distance metric with the best experimental outcome, i.e., the cosine distance, and we named this identification network using the cosine distance as MGAC.

The contributions of this study are as follows:

The MGN, which combined both global and local features, was selected as the person re-identification network, which yielded a high accuracy in image identification.

An attention mechanism network called MGA was developed to evaluate the importance of different granularity information in the MGN and to increase the accuracy of the re-identification network.

Through experimental analysis of several commonly used classical distance metrics, it was found that the cosine distance was the most helpful for increasing the accuracy of person re-identification in MGA, and we named the corresponding identification network MGAC.

The rest of this paper is organized as follows: In Section 2, we review the relevant literature in three main areas: feature extraction, attention mechanisms, and distance metrics. In Section 3, we introduce the proposed method from the perspectives of the MGN structure, the MGA attention mechanism, and the distance metric formula. In Section 4, we present the experimental results of the proposed method in comparison with those of state-of-the-art experiments; we also discuss the results of the attention mechanisms and the distance metrics. In Section 5, we summarize the study.

2. Related Work

Feature extraction and metric learning are two important aspects of person re-identification. Attention mechanisms bring more focus on the learned features than on the important content or parts of the features. They make the learned features more identifiable and can increase identification accuracy. We reviewed the relevant literature in three aspects: feature extraction, attention mechanisms, and distance metrics.

2.1. Feature Extraction

In person re-identification, feature extraction refers to the process of learning and extracting useful information as features so that the image can be better identified. Some

person re-identification methods are based on global feature extraction. In [43], a deep convolutional network with an initial layer learning the discriminative features of positive and negative pairs and a deeper layer learning the relationships between feature maps was constructed. Thus, the features learned by the network maximized the identification capability and output a similarity score to determine whether the two input images were person of the same identity. In [10], a filter-pairing neural network was proposed that jointly processes illumination, geometric transformations, and occlusion. In [1], the similarity between two images was evaluated using a convolutional layer, a matching layer, and a higher layer. These methods extracted more global features but ignored some of the locally identifiable feature information. Currently, there are many methods based on local feature extraction. In [44], researchers fixed the image and evenly divided it into multiple parts. They extracted features using a convolutional network. The features extracted came from each part of the image corresponding to each other. And the outliers were assigned to the most similar parts. These methods focus on local features and do not incorporate global features. In [45], auxiliary guidance was used to locate each important part of the person in the image, and these important parts were learned to increase the identification accuracy. We selected the MGN [33], which combines local and global features for image discrimination, and achieved satisfying discriminative results. However, features learned by the MGN are fixed features. In the case of pose changes of the person in the image, pictures with poor quality detection, and occlusion of the person, some fixed features learned are interference features. These situations can be improved by adding attention mechanisms.

2.2. Attention Mechanisms

The attention mechanism allows important features to be focused on, scoring the features that are learned. Attention mechanisms are divided into hard and soft attention mechanisms. A hard attention mechanism assigns a score of 0 for unimportant features and 1 for important features. A soft attention mechanism assigns scores ranging from 0 to 1, with low scores for unimportant features and high scores for important features. Attention mechanisms allow important features to be enhanced and distracting features to be ignored, increasing identification accuracy. Regarding person image classification, in [46], global averaging pooling was used to generate class activation mappings to find relationships between feature maps and categories for increasing the image classification accuracy. Attention mechanisms can also be applied to various aspects of artificial intelligence. Regarding person re-identification, in [47], joint learning of soft pixel attention and hard regional attention was used to optimize features and complement information from different feature layers for increasing identification accuracy. In [48], the model was enhanced via class activation mapping so that the person re-identification network learned richer information about image features, increasing the identification accuracy. In [49], the occlusion issue in person re-identification was alleviated by enhancing visible feature regions and suppressing occluded feature regions through an attention mechanism. In [50], many attention mechanisms, such as spatial attention and channel attention, are first-order attention mechanisms, and there is a scarcity of research on high-order attention mechanisms. However, in this study, they created a hybrid high-order attention mechanism and applied it to person re-identification to learn richer image features. In [51], the different attention content learned by different scale features was obtained through an attention pyramid mechanism. Our attention mechanism obtains the importance estimate scores of all the features used for identification by concatenating and learning the global features of the three branches of the MGN. Before each feature is concatenated for identification, they are multiplied by the corresponding scores to obtain the final identification features for person re-identification, increasing the identification accuracy.

2.3. Distance Metrics

Distance metrics are used to calculate the distance or similarity between data points. Distance metrics can be applied in various research fields. In person re-identification, a distance metric is used to calculate the distance or similarity of different image features after the features have been extracted from the images. The classical distance metrics commonly used are the cosine distance, the Euclidean distance, the correlation distance, the Mahalanobis distance, the squared Euclidean distance, etc. In [52], a fine-tuned ResNet-50-based method was developed for person re-identification, wherein the cosine distance is used for the feature distance metric. In [53], the method was combined with feature extraction and distance metrics. The Euclidean distance achieved better experiment results. In [54], the researchers dealt with the complex conditions present in the camera using Mahalanobis distance. In [55], the most useful solution for the site selection of railroad construction was found using a curvilinear model of squared Euclidean distance. In [56], optical character recognition was performed using correlation distance in the template matching method. We experimented with various distances commonly used for feature metrics based on the MGN with the addition of an attention mechanism and identified the optimal distance metric for increasing the accuracy of person re-identification.

3. Methods

This section introduces the MGN framework, the attention mechanism of MGA, and the classical distance metric formula. Table 1 presents the symbols used in the section.

Table 1. Symbols used in the aforementioned methods.

Symbol	Definition
$P_i, i = 1, 2, 3$	Branches of the MGN
$p_j, j = 1, 2, \text{ or } j = 1, 2, 3$	Number of stripes of the branches When the number of branches is $i = 2$, the number of stripes is $j = 1, 2$; when the number of branches is $i = 3$, the number of stripes is $j = 1, 2, 3$.
$z_g^{P_i}, i = 1, 2, 3$	Features obtained via global max pooling for each of the three MGN branch features
$z_{p_j}^{P_i}, i = 2, 3$ when $i = 2, j = 1, 2$; when $i = 3, j = 1, 2, 3$	Features obtained via local max pooling in branches P_2 and P_3 of the MGN
$f_g^{P_i}, i = 1, 2, 3$	Global features $z_g^{P_i}$ continue to be extracted to obtain three 256-dimensional global features
$f_{p_j}^{P_i}, i = 2, 3$ when $i = 2, j = 1, 2$; when $i = 3, j = 1, 2, 3$	Local features $z_{p_j}^{P_i}$ continue to be extracted to obtain five 256-dimensional local features
$conv_{1 \times 1}$	1×1 convolution
BN	batch normalization
ReLU	ReLU activation function
Sigmoid	Sigmoid activation function
$z_{g_i}^{P_i}, i = 1, 2, 3$	Feature obtained after $z_g^{P_i}$ continued feature extraction
z_g	Three global features $z_{g_i}^{P_i}$ are concatenated into one 768-dimensional global feature
w_g	128 – dimensional feature obtained after z_g continued feature extraction
w_w	8 – dimensional feature obtained after w_g continued feature extraction
$w_g^{P_i}, i = 1, 2, 3$	In the w_w dimension, the attention values of global features
$w_{p_j}^{P_i}, i = 2, 3$ when $i = 2, j = 1, 2$; when $i = 3, j = 1, 2, 3$	In the w_w dimension, the attention values of the local feature

Table 1. Cont.

Symbol	Definition
$l_g^{Pi}, i = 1, 2, 3$	Global attention features
$l_{pj}^{Pi}, i = 2, 3$ when $i = 2, j = 1, 2$; when $i = 3, j = 1, 2, 3$	Local attention features
$a(x_{11}, x_{12}, \dots, x_{1n})$	n-dimensional feature vector for distance metrics
$b(x_{k1}, x_{k2}, \dots, x_{kn})$	n-dimensional feature vector for distance metrics
$d(a, b)$	Distance between vectors a and b
Σ	Covariance matrix of vectors a and b
ρ_{ab}	Correlation coefficient of vectors a and b
$D(a)$	Variance of vector a
$D(b)$	Variance of vector b
$Cov(a, b)$	Covariance of vectors a and b
$s(a, b)$	Cosine similarity of vectors a and b
θ	Angle between vectors a and b
$a \bullet b$	Dot product of vectors a and b
$\ a\ $	Length of vector a
$\ b\ $	Length of vector b
MGAE	Euclidean distance used in MGA
MGAM	Mahalanobis distance used in MGA
MGACO	Correlation distance used in MGA
MGAC	Cosine distance used in MGA
MGAS	Squared Euclidean distance used in MGA

3.1. MGN Architecture

Multiple Granularity Network (MGN) is a person re-identification network architecture that combines global and multi-part image information. Global features of the image are used to obtain coarse-grained information, and local features are used to obtain medium-grained and fine-grained information. The MGN first extracts the image features from the input image through the ResNet-50 backbone network and forms three branches behind the ResNet-50, which are then used to continue the feature extraction on the features extracted by the ResNet-50 to obtain different granularity information. These branches share the ResNet-50 backbone, and the parameters are not shared between the branches. Through the backbone network, the upper, middle, and lower branch features are obtained as $P1$, $P2$, and $P3$, respectively. The features obtained by global max pooling of the three branch features are denoted as z_g^{Pi} , where Pi represents the branch, $i = 1, 2, 3$. z_g^{Pi} is used to extract Coarse-grained global semantic features from overall information of image features. Meanwhile, the features of $P2$ and $P3$ are used to perform local max pooling. The $P2$ of the middle branch is used to divide the feature map evenly horizontally into two stripes (upper and lower). This is denoted as z_{pj}^{P2} , where pj represents the stripes, and $j = 1, 2$. z_{pj}^{P2} is used to extract local medium-grained semantic features from the upper and lower parts of image features. The $P3$ of the lower branch is used to divide the feature map evenly horizontally into three stripes (upper, middle, and lower). This is denoted as z_{pj}^{P3} , where pj represents the stripes, and $j = 1, 2, 3$. z_{pj}^{P3} is used to extract local fine-grained semantic features from the upper, middle, and lower parts of image features. The local features of the $P2$ and $P3$ branches are denoted as z_{pj}^{Pi} , $i = 2, 3$; when $i = 2, j = 1, 2$, and when

$i = 3, j = 1, 2, 3$. The features of global branches z_g^{Pi} continue to be extracted separately to obtain three 256-dimensional global features f_g^{Pi} , with $i = 1, 2, 3$. Additionally, local features z_{pj}^{Pi} continue to be extracted for obtaining five 256-dimensional local features f_{pj}^{Pi} , with $i = 2, 3$; when $i = 2, j = 1, 2$, and when $i = 3, j = 1, 2, 3$. During testing phases, the test features are obtained by concatenating three 256-dimensional global features and five 256-dimensional local features from the three branches. Finally, a distance metric is used to rank the similarity of the identified features and find the person of the same identity. The MGN combines global and local identification features, taking into account global information and different granularity information, so that the query image can better search images with the person of the same identity in the gallery. This section only covers the MGN framework and the requirements of the attention mechanism; please refer to the original paper on the MGN [33] for details.

3.2. Attention Mechanisms

The MGN uses eight 256-dimensional global f_g^{Pi} and local f_{pj}^{Pi} features by concatenating as identification features. For the global features, $i = 1, 2, 3$, and for the local features, $i = 2, 3$; when $i = 2, j = 1, 2$, and when $i = 3, j = 1, 2, 3$. These global and local features are of equal importance in identification. However, in an image, the importance of each part of the features for identification varies according to the differences in the image quality of the person framed, the postures of the person, their clothing, and the items they carry. Therefore, the image features are extracted in the MGN through the ResNet-50 backbone network and the three branches composed behind ResNet-50. Among the eight global and local features obtained via max pooling, including the global features z_g^{Pi} and local features z_{pj}^{Pi} , for the global features, $i = 1, 2, 3$, and for the local features, $i = 2, 3$; when $i = 2, j = 1, 2$, and when $i = 3, j = 1, 2, 3$. We select the global features in each branch z_g^{Pi} with $i = 1, 2, 3$ as features and continue to extract features using 1×1 convolution, with each branch feature dimension reduced from 2048 to 256 dimensions. We then perform batch normalization on these features separately. Subsequently, we use the ReLU activation function. These three features do not share parameters and are denoted as z_{gi}^{Pi} with $i = 1, 2, 3$:

$$z_{gi}^{Pi} = \text{ReLU}(\text{BN}(\text{conv}_{1 \times 1}(z_g^{Pi}))) \quad (1)$$

where $\text{conv}_{1 \times 1}$ denotes features undergoing 1×1 convolution, BN denotes features undergoing batch normalization, and ReLU denotes the ReLU activation function. After feature extraction, the three learned global features are concatenated into one 768-dimensional global feature denoted as z_g , which is expressed as

$$z_g = \text{concatenate}(z_{gi}^{Pi}) \quad (2)$$

z_g continues with a 1×1 convolution and is reduced to 128 dimensions and then uses the ReLU activation function. The resulting features are denoted as w_g , which is expressed as

$$w_g = \text{ReLU}(\text{conv}_{1 \times 1}(z_g)) \quad (3)$$

1×1 convolution is further used to learn w_g and is reduced to eight dimensions. Then, the Sigmoid activation function is used to obtain the eight-dimensional feature, which is denoted as w_w , and each dimensional feature represents a learned attention value, which is denoted as w_g^{Pi} or w_{pj}^{Pi} . When the learned value is the attention value of the global feature, it is denoted as w_g^{Pi} , with $i = 1, 2, 3$; when the learned value is the attention value of the local feature, it is denoted as w_{pj}^{Pi} , $i = 2, 3$. When $i = 2, j = 1, 2$. When $i = 3, j = 1, 2, 3$. This can be expressed as

$$w_w = [w_g^{P1}, w_g^{P2}, w_g^{P3}, w_{p1}^{P2}, w_{p2}^{P2}, w_{p1}^{P3}, w_{p2}^{P3}, w_{p3}^{P3}] = \text{Sigmoid}(\text{conv}_{1 \times 1}(w_g)) \quad (4)$$

where Sigmoid denotes the Sigmoid activation function. Each of these eight dimensions is used as the attention value of the eight features of the MGN before identification concatenating, which is the importance value. Their element-wise products with the eight corresponding global f_g^{Pi} and local f_{pj}^{Pi} features are denoted as global attention features l_g^{Pi} and local attention features l_{pj}^{Pi} . For the global attention features, $i = 1, 2, 3$, and for the local attention features, $i = 2, 3$. When $i = 2, j = 1, 2$, and when $i = 3, j = 1, 2, 3$. These are used as the final concatenated features for identification and are expressed as follows:

$$\begin{aligned}
 l_g^{Pi} &= f_g^{Pi} \bullet w_g^{Pi} \\
 l_{pj}^{Pi} &= f_{pj}^{Pi} \bullet w_{pj}^{Pi}
 \end{aligned}
 \tag{5}$$

The method of adding the attention mechanism to each part of the MGN is called the Multiple Granularity Attention Network (MGA) and is shown in Figures 1 and 2. The pseudocode for the attention mechanism of MGA is presented in Algorithm 1.

Algorithm 1: Attention mechanism of MGA

Input: Global features : $z_g^{Pi}, i = 1, 2, 3$
 Global features with reduced dimensions: $z_{gi}^{Pi}, i = 1, 2, 3$
 Concatenated features of global features with reduced dimensions: z_g
 Features with reduced dimensions on $z_g:w_g$
 Features with reduced dimensions on $w_g:w_w$
 Attention value of global features in $w_w:w_g^{Pi}, i = 1, 2, 3$
 Attention value of local features in $w_w:w_{pj}^{Pi}$; when $i = 2, j = 1, 2$, and when $i = 3, j = 1, 2, 3$
 Global features before the attention: $f_g^{Pi}, i = 1, 2, 3$
 Local features before the attention: f_{pj}^{Pi} ; when $i = 2, j = 1, 2$, and when $i = 3, j = 1, 2, 3$
 Output: Global features for identification after the attention: $l_g^{Pi}, i = 1, 2, 3$
 Local features for identification after the attention: l_{pj}^{Pi} ; when $i = 2, j = 1, 2$, and when $i = 3, j = 1, 2, 3$

- 1: $z_{gi}^{Pi} = \text{ReLU}(\text{BN}(\text{conv}_{1 \times 1}(z_g^{Pi})))$
 - 2: $z_g = \text{concatenate}(z_{gi}^{Pi})$
 - 3: $w_g = \text{ReLU}(\text{conv}_{1 \times 1}(z_g))$
 - 4: $w_w = [w_g^{P1}, w_g^{P2}, w_g^{P3}, w_{p1}^{P2}, w_{p2}^{P2}, w_{p1}^{P3}, w_{p2}^{P3}, w_{p3}^{P3}] = \text{Sigmoid}(\text{conv}_{1 \times 1}(w_g))$
 - 5: $l_g^{Pi} = f_g^{Pi} \bullet w_g^{Pi}$
 - 6: $l_{pj}^{Pi} = f_{pj}^{Pi} \bullet w_{pj}^{Pi}$
-

The inputs z_g^{P1}, z_g^{P2} , and z_g^{P3} used by the attention mechanism of MGA are all feature maps of size $H \times W \times C$, where H represents the height of the feature map, W represents the width of the feature map, and C represents the number of channels in the feature map. The complexity of the input is $O(H \times W \times C)$, but the $H \times W \times C$ of all three inputs is $1 \times 1 \times N_{\text{input}}$; thus, the input computational complexity is $O(C_{\text{input}})$. The output features of the attention mechanism of MGA w_w have a feature size of $1 \times 1 \times N_{\text{attention}}$; thus, the output computational complexity is $O(C_{\text{attention}})$.

The output of the attention mechanism of MGA performs the element-wise product for each feature that needs attention. Each value of the attention mechanism output is of size $1 \times 1 \times 1$, and the size of each feature that needs attention is $1 \times 1 \times N_{\text{output}}$; thus, the computational complexity is $O(N_{\text{output}})$. Because $N_{\text{input}} > N_{\text{output}} \gg N_{\text{attention}}$, the complexity of the attention mechanism algorithm is $O(N_{\text{input}})$. $O(N_{\text{input}})$ is of linear order; thus, the computational complexity of the attention mechanism algorithm is low. When using one NVIDIA TITAN X GPU for parallel acceleration processing of data, MGA takes an average of 14 ms to extract features from an image, while MGN takes an average of 13 ms to extract features from an image. This indicates that the algorithm complexity is not very high and does not generate a lot of computation.

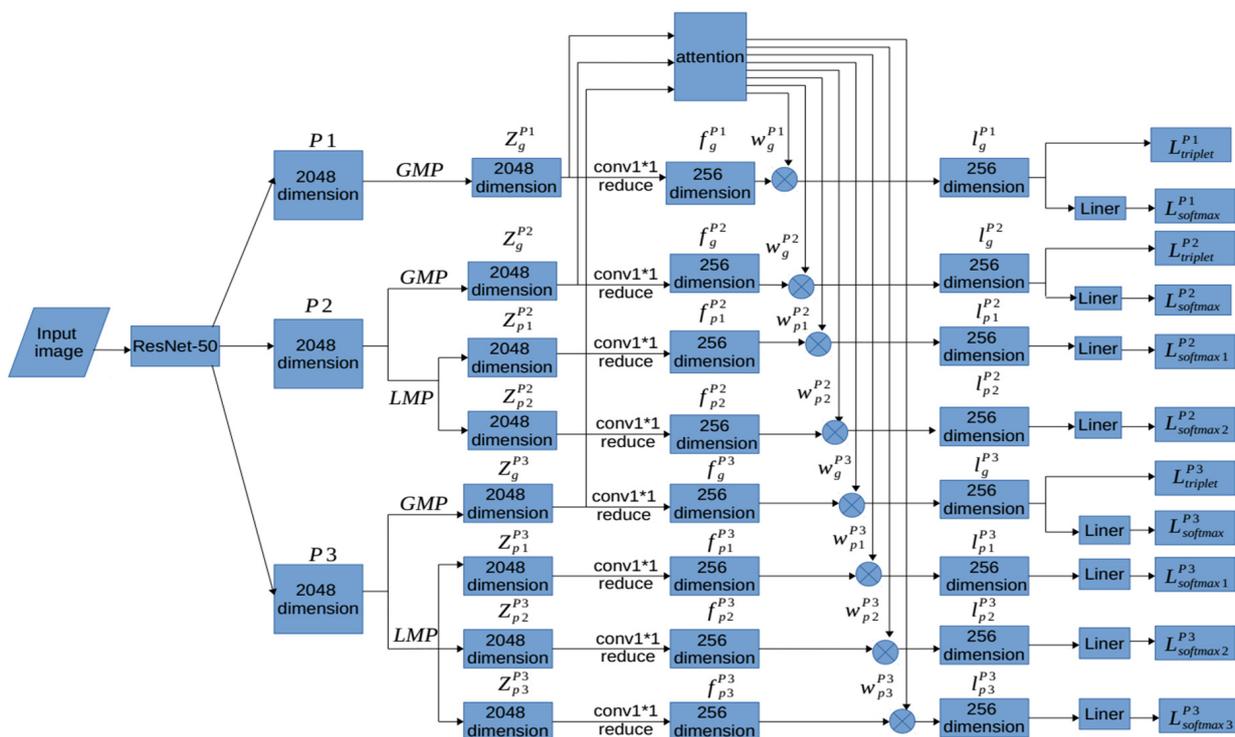


Figure 1. MGA architecture. GMP: Global Max Pooling, LMP: Local Max Pooling, conv1*1 reduce: features undergoing 1×1 convolutional dimension reduction, batch normalization, and ReLU activation function.

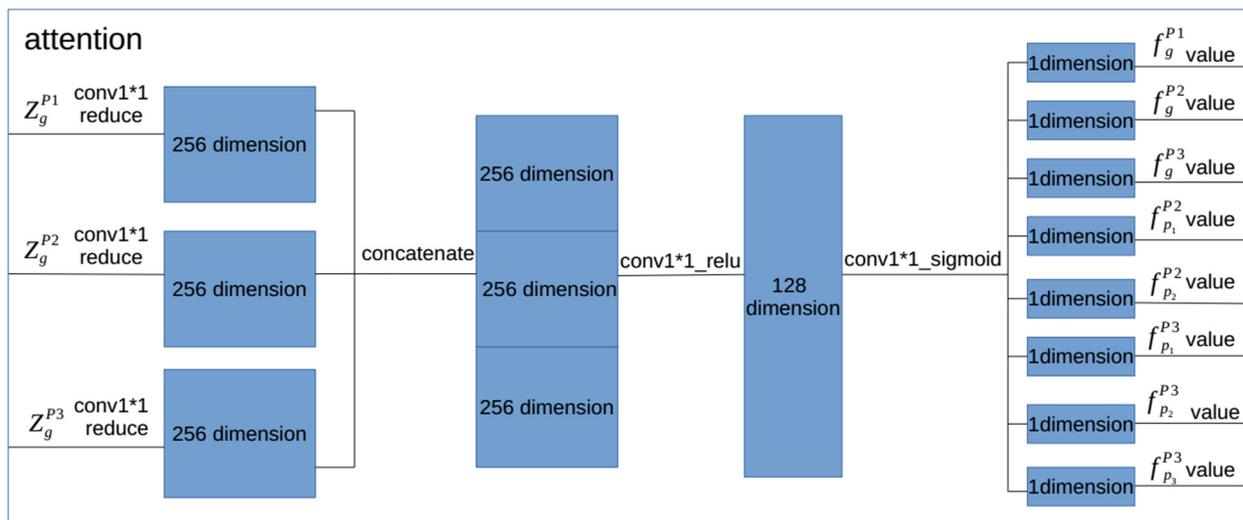


Figure 2. Architecture of the attention mechanism. conv1*1 reduce: features undergoing 1×1 convolutional dimension reduction, batch normalization, and ReLU activation function. conv1*1_relu: features undergoing 1×1 convolutional dimension reduction and ReLU activation function. conv1*1_sigmoid: features undergoing 1×1 convolutional dimension reduction and Sigmoid activation function.

3.3. Distance Metrics

(1) Euclidean distance

The Euclidean distance is a commonly used distance metric. It is the distance between two points in the n-dimensional space. The Euclidean distance between two n-dimensional vectors $a(x_{11}, x_{12}, \dots, x_{1n})$ and $b(x_{k1}, x_{k2}, \dots, x_{kn})$ defined as

$$d(a, b) = \sqrt{(a - b)(a - b)^T} = \sqrt{\sum_{i=1}^n (x_{li} - x_{ki})^2} \tag{6}$$

MGA using the Euclidean distance is called the Multiple Granularity Attention Euclidean Network (MGAE).

(2) Mahalanobis distance

The Mahalanobis distance is measured using covariance, taking into account the relationship of the vectors, and is scale-invariant. There are two n-dimensional vectors $a(x_{11}, x_{12}, \dots, x_{1n})$ and $b(x_{k1}, x_{k2}, \dots, x_{kn})$; their covariance matrix is Σ , and their Mahalanobis distance is defined as

$$d(a, b) = \sqrt{(a - b)^T \Sigma^{-1} (a - b)} \tag{7}$$

MGA using Mahalanobis distance is called the Multiple Granularity Attention Mahalanobis Network (MGAM).

(3) Correlation distance

The correlation distance is measured by the correlation coefficient. The correlation coefficient is a measure of the degree of correlation between two n-dimensional vectors $a(x_{11}, x_{12}, \dots, x_{1n})$ and $b(x_{k1}, x_{k2}, \dots, x_{kn})$ and it has values in the range of $[-1, 1]$. A larger absolute value of the correlation coefficient corresponds to a stronger correlation between the two vectors. A correlation coefficient value of 1 indicates that the two vectors have positive linear dependence. A correlation coefficient value of -1 indicates that the two vectors have negative linear dependence. The correlation coefficient is defined as

$$\rho_{ab} = \frac{Cov(a, b)}{\sqrt{D(a)}\sqrt{D(b)}} = \frac{\sum_{i=1}^n (x_{li} - \bar{x}_l)(x_{ki} - \bar{x}_k)}{\sqrt{\sum_{i=1}^n (x_{li} - \bar{x}_l)^2} \sqrt{\sum_{i=1}^n (x_{ki} - \bar{x}_k)^2}}$$

where $Cov(a, b)$ represents the covariance of the two vectors, and $D(a)$ and $D(b)$ represent the variances of the vectors. The correlation distance is defined as

$$d(a, b) = 1 - \rho_{ab} \tag{8}$$

MGA using the correlation distance is called the Multiple Granularity Attention Correlation Network (MGACO).

(4) Cosine distance

The cosine distance is measured by the cosine of the angle between two vectors. The cosine similarity of the two n-dimensional vectors $a(x_{11}, x_{12}, \dots, x_{1n})$ and $b(x_{k1}, x_{k2}, \dots, x_{kn})$ is defined as

$$(a, b) = \cos(\theta) = \frac{a \bullet b}{\|a\| \cdot \|b\|} = \frac{\sum_{i=1}^n x_{li} x_{ki}}{\sqrt{\sum_{i=1}^n x_{li}^2} \sqrt{\sum_{i=1}^n x_{ki}^2}}$$

where θ represents the angle between the two vectors, $a \bullet b$ represents the dot product of the two vectors, and $\|a\|$ and $\|b\|$ represent the lengths of the two vectors. The cosine similarity represents the relative difference between two vectors in terms of direction. It has values in the range of $[-1, 1]$. A larger cosine similarity value corresponds to a smaller angle between the two vectors and a higher degree of similarity. When the two vectors have exactly the same direction, the cosine similarity is 1. A smaller cosine similarity value corresponds to a larger angle between the two vectors and a lower degree of similarity. When the two vectors are in exactly opposite directions, the cosine similarity is -1 . The cosine distance of two vectors is defined as

$$d(a, b) = 1 - \cos(\theta) \quad (9)$$

MGA using the cosine distance is called the Multiple Granularity Attention Cosine Network (MGAC).

(5) Squared Euclidean distance

The squared Euclidean distance is the square of the Euclidean distance. The Euclidean distance between the two n -dimensional vectors $a(x_{11}, x_{12}, \dots, x_{1n})$ and $b(x_{k1}, x_{k2}, \dots, x_{kn})$ can be defined as

$$d(a, b) = (a - b)(a - b)^T = \sum_{i=1}^n (x_{li} - x_{ki})^2 \quad (10)$$

MGA using the squared Euclidean distance is called the Multiple Granularity Attention Squared Euclidean Network (MGAS).

Through experiments, the best distance metric method based on MGA is the cosine distance, so the cosine distance was selected as the distance metric for our method.

In summary, our person re-identification method is MGAC. The overall framework of our method is shown in Figure 3.

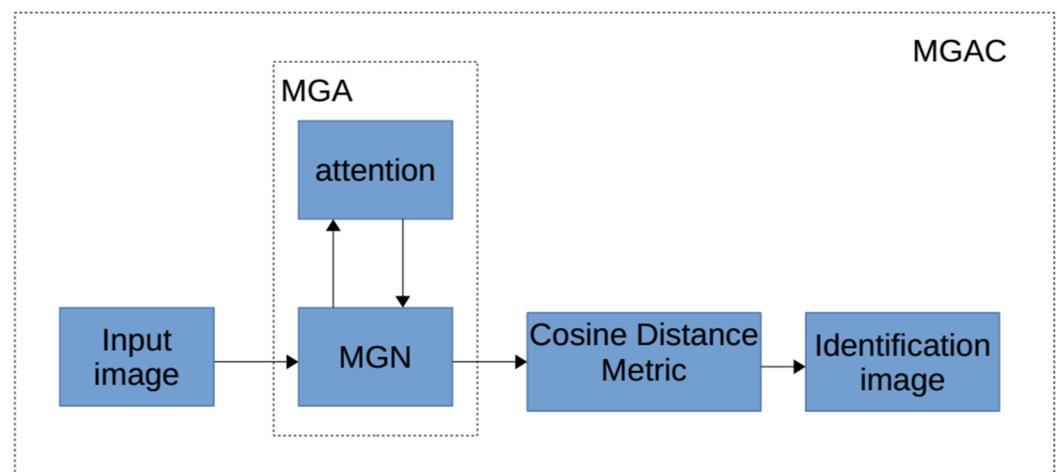


Figure 3. MGAC architecture.

4. Experiments

4.1. Datasets and Protocols

The person re-identification dataset used in the experiments was Market-1501, which is a large-scale dataset. In contrast to other re-identification datasets that use hand-crafted annotated ground truth boxes, its training and test sets use the bounding boxes detected by the detector as annotated boxes. In Market-1501, there are 12,936 training images, test images, and 3368 query images. It has a total of 1501 person identities, with 751 in the training set and 750 in the test set. The dataset was collected by five 1280×1080 HD cameras and a 720×576 SD camera. Each identity is annotated on at least two cameras for cross-camera identification. There are some overlap between some camera areas.

To evaluate the proposed method, we used the mean average precision (mAP) and top-1 of the cumulative match characteristic (CMC) to evaluate and compare the experimental results of the proposed method with those of other methods. We also evaluated the proposed method using the top-1, top-5, and top-10 of CMC.

4.2. Implementation Details

The experiment was implemented using the PyTorch framework, and an Nvidia TITAN X GPU was used for data-parallel acceleration. During the experiment, the resizing of the input images, the initialization of the parameters of ResNet-50, and the data expansion all followed the settings presented in the original paper on the MGN [33], with some improvements. In the sampling, six identities and four images for each identity were randomly selected as training data for each small batch sampling. In the training, ADAM was used as the optimizer, and the initial learning rate was set at 0.0002. Training was performed for a total of 600 epochs. The learning rate decayed to 0.00002 at 300 epochs and to 0.000002 at 500 epochs.

4.3. Comparison with State-of-the-Art Methods

The results of the experiments based on the Market-1501 dataset are presented in Tables 2 and 3. Table 2 presents the results obtained without re-ranking, and Table 3 presents the results obtained with re-ranking.

Table 2. Experimental results for the Market-1501 dataset. The best results of all experiments are shown in bold, and the best results for each implementation are highlighted in grey. G: global feature; MR: multiple regions feature; AG: auxiliary guidance; A: attention.

	Method	Top-1 (%)	mAP (%)
G	DaRe	86.4	69.3
	AOS	86.5	70.4
	OSNet	94.8	84.9
	DG-Net	94.8	86.0
	DML	87.7	68.8
	PAN	82.8	63.4
	SVDNet	82.3	62.1
	SL-ReID	84.5	65.4
	SOMAnet	73.9	47.9
MR	PCB + RPP	93.8	81.6
	MGN	95.7	86.9
	MSCAN	80.3	57.5
	HPM	94.2	82.7
	MultiScale	88.9	73.1
	PAR	81.0	63.4
	DCR	93.8	84.7
MultiRegion	66.4	41.2	
AG	PABR	91.7	79.6
	PGFA	91.2	76.8
	PDC	84.4	63.4
	AACN	85.9	66.9
	PN-GAN	89.4	72.6
	MGCAM	83.8	74.3
	LaST_Cloth	93.1	81.7
	SPReID	92.5	81.3
GLAD	89.9	73.9	
	HA-CNN	91.2	75.7
	Mancs	93.1	82.3
	CASN	94.4	82.8

Table 2. *Cont.*

	Method	Top-1 (%)	mAP (%)
A	IANet	94.4	83.1
	DuATM	91.4	76.6
	CAMA	94.7	84.5
	MHN-6	95.1	85.0
	reID-NAS+	95.1	85.7
	SFNet	95.3	87.7
	MHSA-Net	94.6	84.0
	MGAC	95.1	88.2

Table 3. Results with re-ranking for Market-1501. The best results of the experiment are shown in bold.

Method	Top-1 (%)	mAP (%)
TriNet	86.7	81.1
AOS	88.7	83.3
AACN	88.7	83.0
PSE + ECN	90.3	84.0
LuNet	84.59	75.62
GP-reid	92.2	90.0
CamStyle	89.5	71.5
MHSA-Net	95.5	93.0
DaRe	88.3	82.0
MGN	96.6	94.2
MGAC	96.2	94.9

For person re-identification without re-ranking, the experiments are reported according to the implementation methods of person re-identification, including global features, multiple regions, auxiliary guidance, attention mechanisms, and our method, MGAC.

In the experiments without re-ranking, MGAC achieved the top-1 of 95.1% and the mAP of 88.2%. For global features, DG-Net achieved better experimental results, which were 94.8% in the top-1 and 86.0% in the mAP; in comparison, the top-1 and mAP of MGAC were 0.3% and 2.2% higher. For the multiple regions feature, the MGN achieved the highest experimental results excluding MGAC, with 95.7% in the top-1 and 86.9% in the mAP; in comparison, the top-1 and mAP of MGAC were 0.6% lower and 1.3% higher. For auxiliary guidance, LaST_Cloth achieved better results, which were 93.1% in the top-1 and 81.7% in the mAP; in comparison, the top-1 and mAP of MGAC were 2.0% and 6.5% higher. For the attention mechanism, SFNet achieved better results, which were 95.3% in the top-1 and 87.7% in the mAP; in comparison, the top-1 and mAP of MGAC were 0.2% lower and 0.5% higher.

According to the results, MGAC performed best for the Market-1501 dataset. Its mAP was 1.3% higher than that of the best-known method (MGN), and it achieved better top-1 experimental results than all the other methods except MGN and SFNet. The results indicated that MGAC is useful for improving the MGN with regard to attention and distance metrics.

In the experiments with re-ranking, MGAC achieved the top-1 of 96.2% and the mAP of 94.9%. Of all the experiments with re-ranking, the one with better experimental results was the MGN, which had the top-1 of 96.6% and the mAP of 94.2%; in comparison, the top-1 and mAP of MGAC were 0.4% lower and 0.7% higher. As indicated by the experimental results, the re-ranking significantly improved the results of all the experiments. With the re-ranking, MGAC outperformed all the other methods, except that it had a slightly

lower top-1 than the MGN. The experimental results indicated that MGAC is useful even after re-ranking.

Figures 4 and 5 present experimental results histograms of the mAP and top-1, which were selected from methods proposed in recent years among the best experimental results and MGAC. As shown, for the mAP, MGAC achieved the best results both with and without re-ranking. For top-1, both with and without re-ranking, MGAC had better results than all the other methods except for the MGN.

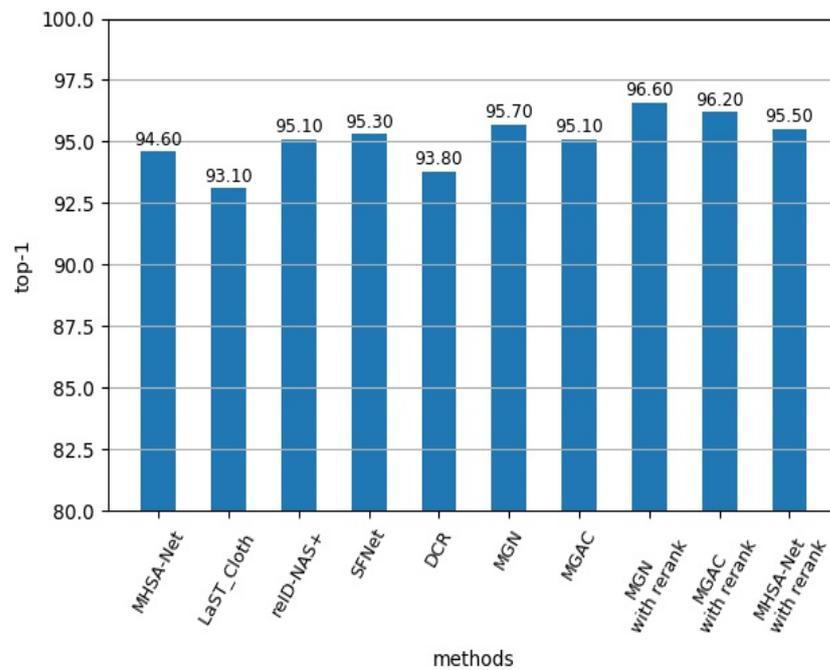


Figure 4. Histogram of the top-1 experimental results of recently proposed methods and MGAC for the Market-1501 dataset.

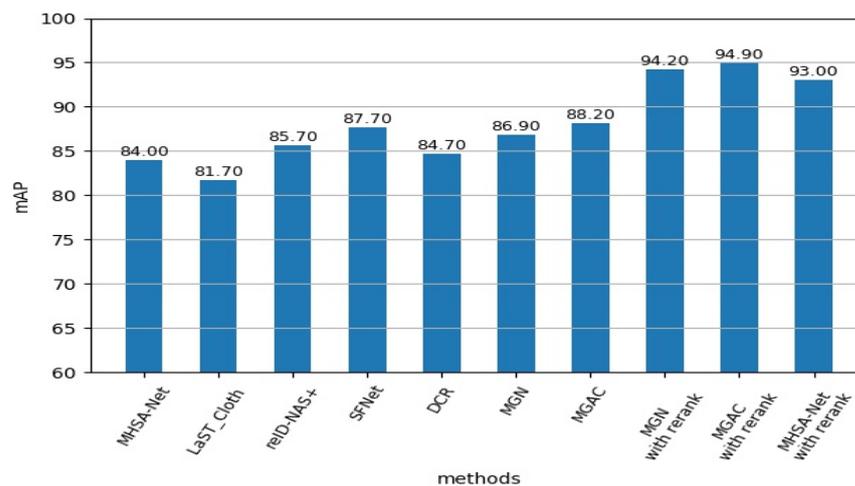


Figure 5. Histogram of the mAP experimental results of recently proposed methods and MGAC for the Market-1501 dataset.

Figures 6 and 7 show curves of the experimental results of MGAC without and with re-ranking for the Market-1501 dataset. The figures present the results of MGAC for 600 training epochs for mAP, top-1, top-3, top-5, and top-10. As shown, the results of MGAC tended to increase steadily as the number of training epochs increased. The best results were achieved at 350 training epochs, and the results remained relatively stable. MGAC achieved better results for mAP and top-1 to top-10. The re-ranking significantly

increased the accuracy of mAP and top and narrowed the gap between mAP and top. The results of the experiments without re-ranking tended to be more stable during the training stage than those of the experiments with re-ranking.

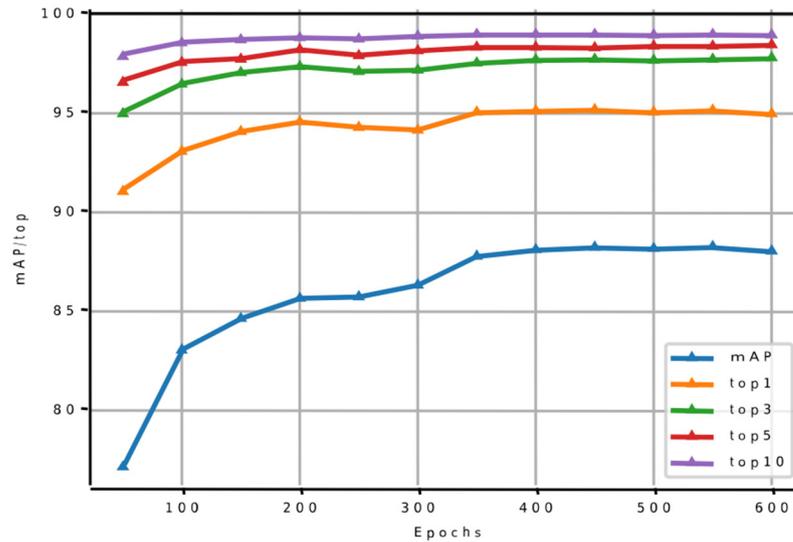


Figure 6. Curves of the experimental results of MGAC for the Market-1501 dataset without re-ranking.

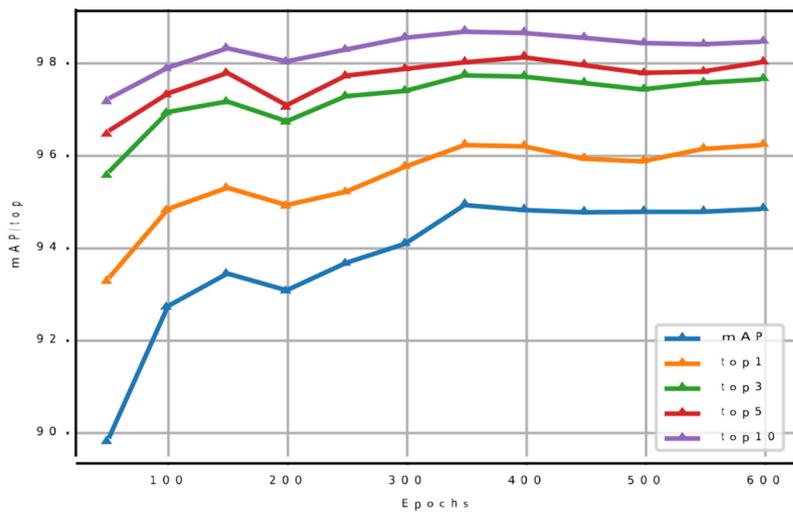


Figure 7. Curves of the experimental results of MGAC for the Market-1501 dataset with re-ranking.

Figure 8 shows curves of the ROC experimental results of MGAC without re-ranking for the Market-1501 dataset. The ROC curve in the figure is drawn by taking into account the identification results of all categories of query images. From the figure, it can be seen that the false positive rate is driven close to 0, the true positive rate is driven close to 1, and the AUC (area under the ROC curve) is approximately equal to 1. These results indicate that MGAC is a good method of person re-identification that can identify the correct images well and distinguish the incorrect images well.

Further discussions of our method with regard to attention mechanisms and distance metrics are presented in the experimental discussion section.

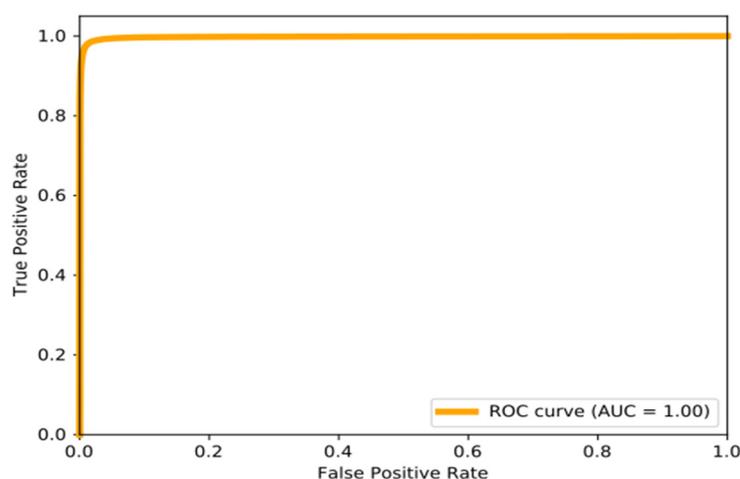


Figure 8. Curves of the ROC experimental results of MGAC for the Market-1501 dataset without re-ranking.

4.4. Discussion

4.4.1. Experimental Results of Attention Mechanisms

In this section, we discuss the two components of our method: the attention mechanism and the distance metrics.

The MGN, with the addition of an attention mechanism, is called MGA. MGA is based on the MGN and learns an attention mechanism that evaluates the importance of each feature learned by the MGN that is finally used for person re-identification, with the important part scoring higher and the unimportant part scoring lower, reducing the impact of distracting features on the person re-identification process. In this section, we discuss attention mechanisms. Table 4 presents the results of the experiments without re-ranking. For the MGN, the top-1 was 95.7%, and the mAP was 86.9%. For MGA, the top-1 was 95.0%, and the mAP was 87.6%. With the addition of the attention mechanism, the mAP was 0.7% higher, and the top-1 was 0.7% lower. Table 5 presents the results of the experiments with re-ranking. For the MGN, the top-1 was 96.6%, and the mAP was 94.2%. For MGA, the top-1 was 96.0%, and the mAP was 94.8%. The mAP of MGA was 0.6% higher than that of MGN, and the top-1 was 0.6% lower. According to the experiments, the mAP results of MGA were improved both with and without re-ranking, suggesting that the proposed method with the attention mechanism is useful for MGN with and without re-ranking.

Table 4. Comparison of experimental results between the MGN and MGA without re-ranking.

Method	Top-1 (%)	mAP (%)
MGN	95.7	86.9
MGA	95.0	87.6

Table 5. Comparison of experimental results between the MGN and MGA with re-ranking.

Method	Top-1 (%)	mAP (%)
MGN	96.6	94.2
MGA	96.0	94.8

Figures 9 and 10 present the experimental results for the mAP and top-1 of MGA without and with re-ranking. As shown, MGA was unstable in the initial training stage, and as the number of epochs increased, its results tended to increase smoothly. During the training, the curves of MGA without re-ranking were smoother than those with re-ranking.

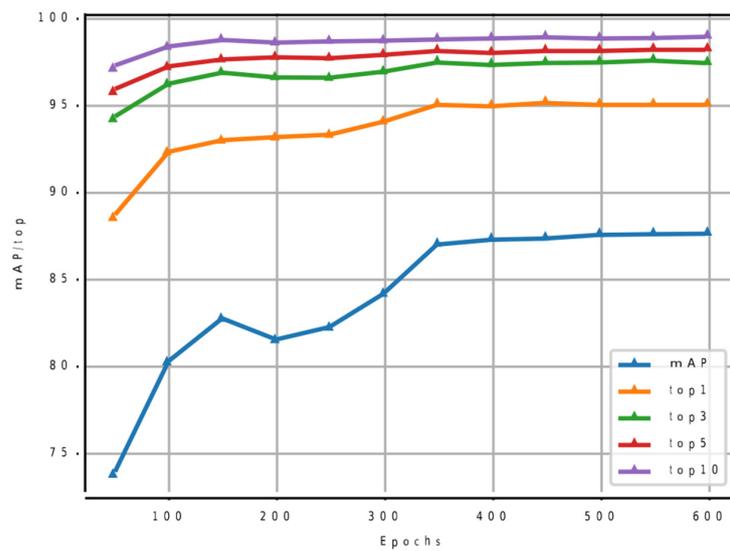


Figure 9. Curves of the experimental results of MGA for the Market-1501 dataset without re-ranking.

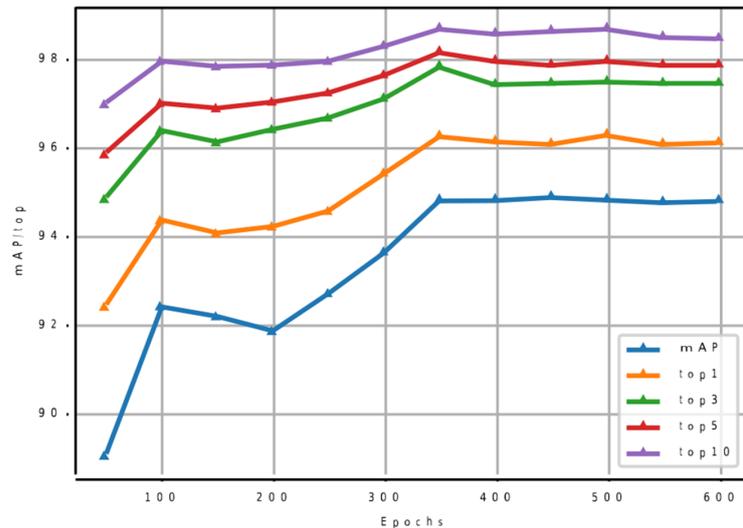


Figure 10. Curves of the experimental results of MGA for the Market-1501 dataset with re-ranking.

4.4.2. Experimental Results for Distance Metrics

The distance metric is an important element in person re-identification. The distance metric used in the MGN was the Euclidean distance. Additionally, classical distance metrics, including the Euclidean distance, Mahalanobis distance, correlation distance, cosine distance, and squared Euclidean distance, were used in the experiments, which were based on the MGA with re-ranking. Table 6 presents the experimental results.

Table 6. Experimental results for the distance metrics. The best results of the experiment are shown in bold.

Method	Top-1 (%)	mAP (%)
MGAE (Euclidean)	96.0	94.8
MGAM (Mahalanobis)	95.9	94.2
MGACO (correlation)	95.9	94.7
MGAC (cosine)	96.2	94.9
MGAS (squared Euclidean)	96.4	94.7

The MGA had the top-1 of 96.0% and the mAP of 94.8% for the Euclidean distance, the top-1 of 95.9% and the mAP of 94.2% for the Mahalanobis distance, the top-1 of 95.9% and the mAP of 94.7% for the correlation distance, the top-1 was 96.2% and the mAP was 94.9% for the cosine distance, and the top-1 was 96.4% and the mAP was 94.7% for the squared Euclidean distance. The distance metric with the best mAP result was the cosine distance; the mAP was 0.1% higher than that for the Euclidean distance. The distance metric with the best top-1 result was the squared Euclidean distance; the top-1 was 0.4% higher than that for the Euclidean distance. Because we used the mAP as the main evaluation result, the cosine distance was selected as the best solution.

Figures 11 and 12 present histograms of the experimental results of the mAP and top-1 of the MGA for various distance metrics. As shown, MGAC, using the cosine distance, and MGAS, using the squared Euclidean distance, achieved the best mAP and top-1 experimental results.

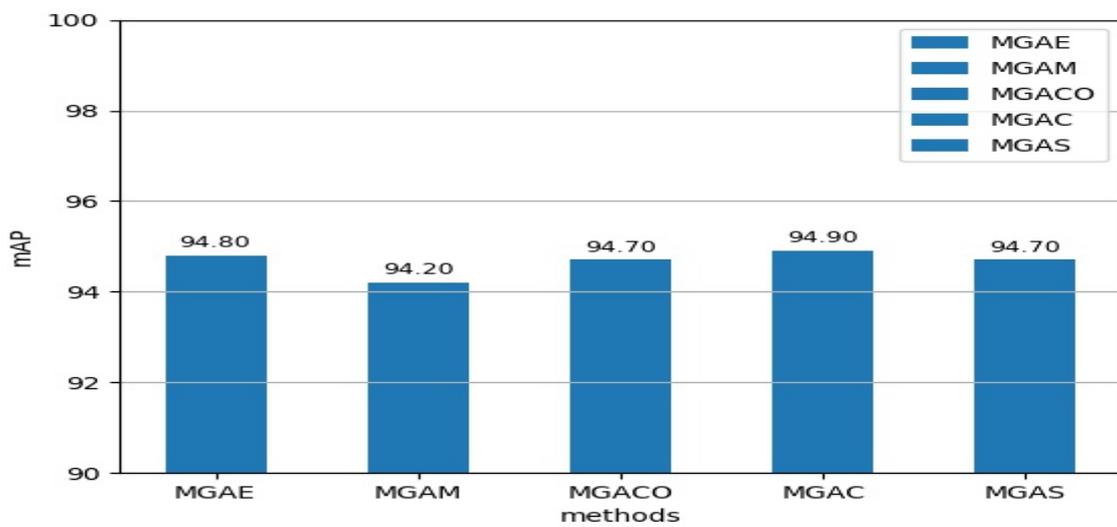


Figure 11. Histogram of the mAP experimental results of MGA obtained using various distance metrics for the Market-1501 dataset.

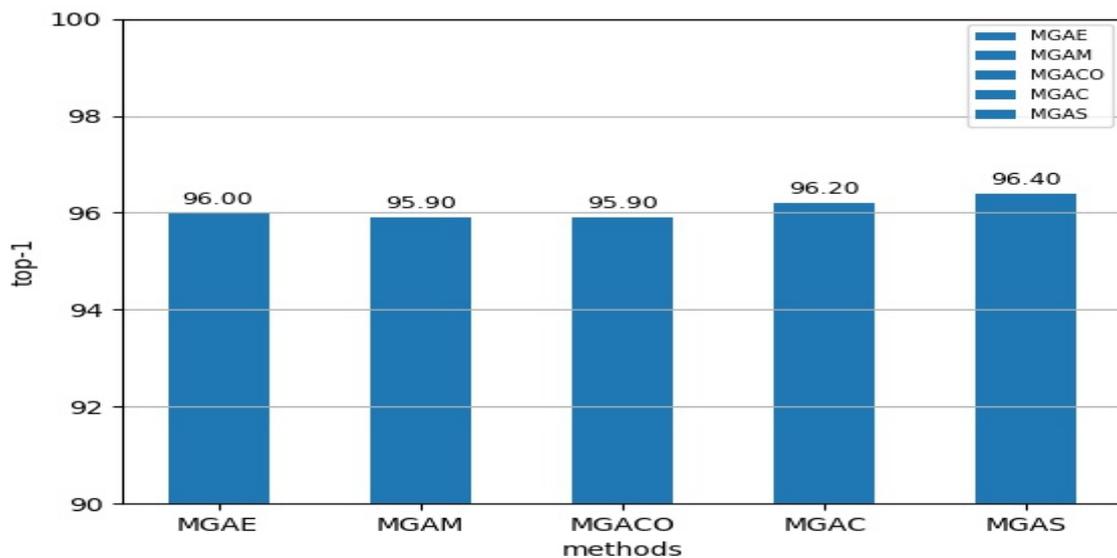


Figure 12. Histogram of the top-1 experimental results of MGA obtained using various distance metrics for the Market-1501 dataset.

4.4.3. Visualization Experiment Results

To validate our method, we visualized the experimental results of our method in Figure 13. The leftmost column is the query person images, and the right rows corresponding to the query images are the result of MGAC identifying person images with the same identity. As can be seen from the figure, MGAC can very well find person images with different angles, different poses, different lighting, and different quality detection boxes. Lines 5 and 6 are partially occluded person images, and there are a few images that are found to be incorrect, but most of the images are found to be correct. This indicates that MGAC can better pay attention to important feature information and ignore interference feature information, achieving better person re-identification accuracy.

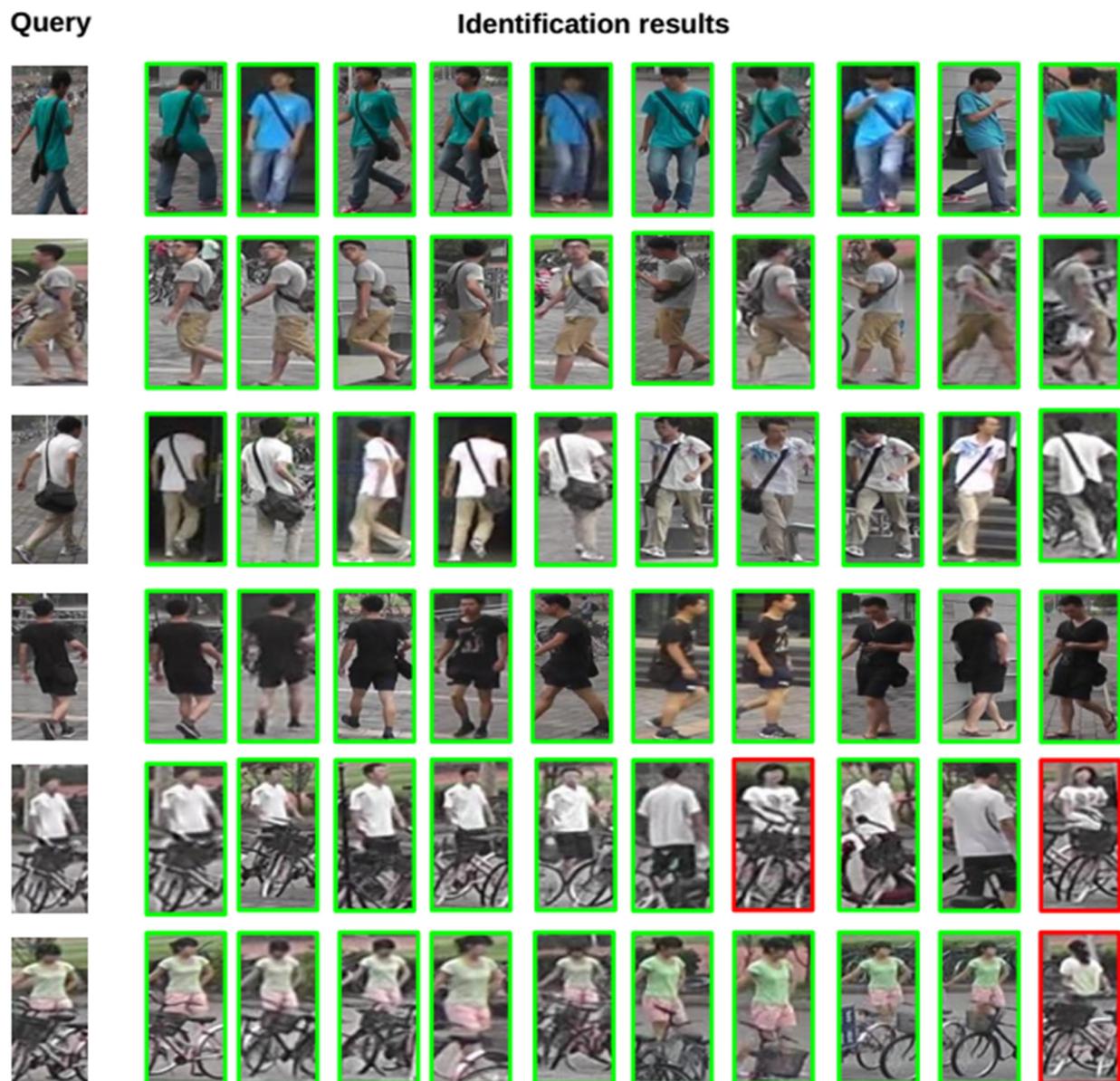


Figure 13. MGAC visualized experimental results. The green boxes are the correct matches. The red boxes are the incorrect matches.

5. Conclusions

An MGAC person re-identification network was constructed by adding attention mechanisms to the MGN to form the MGA network and then using the cosine distance as the distance metric to form MGAC. MGAC used an attention mechanism to evaluate image

features, emphasizing important features and ignoring interfering features. And cosine distance metric be better than other feature distance metric, leading to improved person re-identification accuracy. Experiments involving the Market-1501 dataset indicated that the MGAC network, with the addition of attention mechanisms and the use of the cosine distance, can increase identification accuracy. In the future, we may develop some person re-identification networks using different datasets and develop attention mechanisms and distance metrics according to the features.

Funding: This research received no external funding.

Data Availability Statement: The authors choose not to disclose the data.

Conflicts of Interest: The author declare no conflict of interest.

References

1. Wu, L.; Shen, C.; Hengel, A. PersonNet: Person re-identification with deep convolutional neural networks. *arXiv* **2016**, arXiv:1601.07255.
2. Chen, J.; Zhang, Z.; Wang, Y. Relevance Metric Learning for Person Re-Identification by Exploiting Listwise Similarities. *IEEE Trans. Image Process.* **2015**, *24*, 4741–4755. [[CrossRef](#)]
3. Ding, S.; Lin, L.; Wang, G.; Chao, H. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognit.* **2015**, *48*, 2993–3003. [[CrossRef](#)]
4. Shi, X. Person Re-identification Based on Improved Residual Neural Networks. In Proceedings of the 2021 5th International Conference on Communication and Information Systems (ICCIS), Chongqing, China, 15–17 October 2021; pp. 170–174.
5. Wang, J.; Zhou, S.; Wang, J.; Hou, Q. Deep ranking model by large adaptive margin learning for person re-identification. *Pattern Recognit.* **2018**, *74*, 241–252. [[CrossRef](#)]
6. Fu, M.; Sun, S.; Gao, H.; Wang, D.; Tong, X.; Liu, Q. Liang Improving Person Reidentification Using a Self-Focusing Network in Internet of Things. *IEEE Internet Things J.* **2022**, *9*, 9342–9353. [[CrossRef](#)]
7. Zhang, Z.; Si, T.; Liu, S. Integration convolutional neural network for person re-identification in camera networks. *IEEE Access* **2018**, *6*, 36887–36896. [[CrossRef](#)]
8. Tian, H.; Hu, J. Self-Regulation Feature Network for Person Reidentification. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–8. [[CrossRef](#)]
9. Lin, L.; Wu, T.; Porway, J.; Xu, Z. A stochastic graph grammar for compositional object representation and recognition. *Pattern Recognit.* **2009**, *42*, 1297–1307. [[CrossRef](#)]
10. Li, W.; Zhao, R.; Xiao, T.; Wang, X. DeepReID: Deep filter pairing neural network for person re-identification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 152–159.
11. Van De Weijer, J.; Schmid, C.; Verbeek, J.; Larlus, D. Learning color names for real-world applications. *IEEE Trans. Image Process.* **2009**, *18*, 1512–1523. [[CrossRef](#)] [[PubMed](#)]
12. Liu, H.; Feng, J.; Qi, M.; Jiang, J.; Yan, S. End-to-End comparative attention networks for person re-identification. A Publication of the IEEE Signal Processing Society. *IEEE Trans. Image Process.* **2017**, *26*, 3492–3506. [[CrossRef](#)]
13. Feng, Z.; Lai, J.; Xie, X. Learning view-specific deep networks for person re-identification. *IEEE Trans. Image Process.* **2018**, *27*, 3472–3483. [[CrossRef](#)]
14. Zhou, S.; Wang, J.; Meng, D.; Xin, X.; Li, Y.; Gong, Y.; Zheng, N. Deep self-paced learning for person re-identification. *Pattern Recognit.* **2018**, *76*, 739–751. [[CrossRef](#)]
15. Zheng, W.; Gong, S.; Xiang, T. Re-identification by relative distance comparison. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 653–668. [[CrossRef](#)]
16. Liu, X.; Song, M.; Zhao, Q.; Tao, D.; Chen, C.; Bu, J. Attribute restricted latent topic model for person re-identification. *Pattern Recognit.* **2012**, *45*, 4204–4213. [[CrossRef](#)]
17. Liang, X.; Lin, L. Look into person: Joint body parsing & pose estimation network and a new benchmark. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 871–885.
18. 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](#)] [[PubMed](#)]
19. Wu, L.; Wang, Y.; Li, X.; Gao, J. What-and-where to match: Deep spatially multiplicative integration networks for person re-identification. *Pattern Recognit.* **2018**, *76*, 727–738. [[CrossRef](#)]
20. Chen, Y.; Zhu, X.; Zheng, W.; Lai, J. Person re-identification by camera correlation aware feature augmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 392–408. [[CrossRef](#)] [[PubMed](#)]
21. Yao, H.; Zhang, S.; Hong, R.; Zhang, Y.; Xu, C.; Tian, Q. Deep representation learning with part loss for person re-identification. *IEEE Trans. Image Process.* **2019**, *28*, 2860–2871. [[CrossRef](#)]
22. Yang, F.; Yan, K.; Lu, S.; Jia, H.; Xie, X.; Gao, W. Attention driven person re-identification. *Pattern Recognit.* **2019**, *86*, 143–155. [[CrossRef](#)]

23. Xu, B.; He, L.; Liang, J.; Sun, Z. Learning Feature Recovery Transformer for Occluded Person Re-Identification. *IEEE Trans. Image Process.* **2022**, *31*, 4651–4661. [[CrossRef](#)] [[PubMed](#)]
24. Zhao, R.; Oyang, W.; Wang, X. Person re-identification by saliency learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 356–370. [[CrossRef](#)] [[PubMed](#)]
25. Ma, X.; Zhu, X.; Gong, S.; Xie, X.; Hu, J.; Lam, K.; Zhong, Y. Person re-identification by unsupervised video matching. *Pattern Recognit.* **2017**, *65*, 197–210. [[CrossRef](#)]
26. Gao, Z.; Gao, L.; Zhang, H.; Cheng, Z.; Hong, R.; Chen, S. DCR: A Unified Framework for Holistic/Partial Person ReID. *IEEE Trans. Multimed.* **2021**, *23*, 3332–3345. [[CrossRef](#)]
27. Shu, X.; Wang, X.; Zang, X.; Zang, S.; Zhang, S.; Chen, Y.; Li, G.; Tian, Q. Large-Scale Spatio-Temporal Person Re-Identification: Algorithms and Benchmark. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 4390–4403. [[CrossRef](#)]
28. Zhou, Q.; Zhong, B.; Liu, X.; Ji, R. Attention-Based Neural Architecture Search for Person Re-Identification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 6627–6639. [[CrossRef](#)]
29. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
30. Tan, H.; Liu, X.; Yin, B.; Li, X. MHSA-Net: Multihead Self-Attention Network for Occluded Person Re-Identification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *online ahead of print*. [[CrossRef](#)]
31. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable Person Re-identification: A Benchmark. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 18 February 2016.
32. Peng, W.; Zhong, X.; Zou, C.; Zhang, J.; Ci, Q.; Zhong, L. Label-noise Robust Person Re-identification via Symmetric Learning. In Proceedings of the 7th Annual International Conference on Network and Information Systems for Computers, Guiyang, China, 23–25 July 2021.
33. Wang, G.; Yuan, Y.; Chen, X.; Li, J.; Zhou, X. Learning discriminative features with multiple granularities for person re-identification. In Proceedings of the 26th ACM International Conference on Multimedia (MM), Seoul, Republic of Korea, 22–26 October 2018; pp. 274–282.
34. Su, C.; Zhang, S.; Xing, J.; Gao, W.; Tian, Q. Multi-type attributes driven multi-camera person re-identification. *Pattern Recognit.* **2018**, *75*, 77–89. [[CrossRef](#)]
35. Wang, J.; Li, S. Query-driven iterated neighborhood graph search for large scale indexing. In Proceedings of the 20th ACM International Conference on Multimedia (MM), Nara, Japan, 29 October–2 November 2012; pp. 179–188.
36. Wang, L.; Zhang, Y.; Feng, J. On the Euclidean distance of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 1334–1339. [[CrossRef](#)]
37. Melder, R.A. Some characterizations of city block distance. *Pattern Recognit. Lett.* **1987**, *6*, 235–240. [[CrossRef](#)]
38. McLachlan, G.J. Mahalanobis distance. *Resonance* **1999**, *4*, 20–26. [[CrossRef](#)]
39. Mercioni, M.A.; Holban, S. A Survey of Distance Metrics in Clustering Data Mining Techniques. In Proceedings of the 3rd International Conference on Graphics and Signal Processing (ICGSP), Hong Kong, China, 1–3 June 2019; pp. 44–47.
40. Huang, Y.; Xu, J.; Wu, Q.; Zheng, Z.; Zhang, Z.; Zhang, J. Multipseudo regularized label for generated data in person re-identification. *IEEE Trans. Image Process.* **2018**, *28*, 1391–1403. [[CrossRef](#)]
41. Zheng, Z.; Zheng, L.; Yang, Y. A discriminatively learned cnn embedding for person reidentification. *ACM Trans. Multimed. Comput. Commun. Appl.* **2017**, *14*, 1–20. [[CrossRef](#)]
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
43. Ahmed, E.; Jones, M.; Marks, T.K. An improved deep learning architecture for person re-identification. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3908–3916.
44. Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; Wang, S. Beyond part models: Person retrieval with refined part pooling. In Proceedings of the 2018 European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 480–496.
45. Su, C.; Li, J.; Zhang, S.; Xing, J.; Gao, W.; Tian, Q. Pose-driven deep convolutional model for person re-identification. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 3960–3969.
46. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
47. Li, W.; Zhu, X.; Gong, S. Harmonious attention network for person re-identification. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2285–2294.
48. Yang, W.; Huang, H.; Zhang, Z.; Chen, X.; Huang, K.; Zhang, S. Towards rich feature discovery with class activation maps augmentation for person re-identification. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 1389–1398.
49. Xu, Y.; Zhao, L.; Qin, F. Dual attention-based method for occluded person re-identification. *Knowl. Based Syst.* **2020**, *212*, 106554.1–106554.12. [[CrossRef](#)]
50. Chen, B.; Deng, W.; Hu, J. Mixed high-Order attention network for person re-identification. In Proceedings of the 2019 IEEE International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 371–381.

51. Chen, G.; Gu, T.; Lu, J.; Bao, J.A.; Zhou, J. Person re-identification via Attention Pyramid. *IEEE Trans. Image Process.* **2021**, *30*, 7663–7676. [[CrossRef](#)] [[PubMed](#)]
52. Hu, X.; Chen, Y.; Ma, X.; Liang, Y. Research on person re-Identification method based on fine-tune ResNet50. In Proceedings of the 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 25–27 September 2020.
53. Wu, W.; Yang, Z.; Tao, D.; Zhang, Q.; Cheng, J. On comparing different metric learning schemes for deep feature based person re-identification with camera adaption. In Proceedings of the 2019 IEEE International Conference on Real-time Computing and Robotics (RCAR), Irkutsk, Russia, 4–9 August 2019.
54. Ma, L.; Yang, X.; Tao, D. Person re-identification over camera networks using multi-task distance metric learning. *IEEE Trans. Image Process.* **2014**, *23*, 3656–3670.
55. Lu, L.; Zhou, J. A route-like demand location problem based on squared-euclidean distance. In Proceedings of the IEEE 2016 International Conference on Logistics, Informatics and Service Sciences (LISS), Sydney, Australia, 24–27 July 2016.
56. Shehu, G.; Ashir, A.; Eleyan, A. Character recognition using correlation & hamming distance. In Proceedings of the 2015 Signal Processing and Communications Applications Conference, Malatya, Turkey, 22 June 2015.

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.