information

Article

Improving Performance in Person Reidentification Using Adaptive Multiple Loss Baseline

Zhongmiao Huang, Liejun Wang *, Yongming Li, Anyu Du and Shaochen Jiang

College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
* Correspondence: wljxju@xju.edu.cn; Tel.: +86-13999816618

Abstract: Currently, deep learning is the mainstream method to solve the problem of person reidentification. With the rapid development of neural networks in recent years, a number of neural network frameworks have emerged for it, so it is becoming more important to explore a simple and efficient baseline algorithm. In fact, the performance of the same module varies greatly in different positions of the network architecture. After exploring how modules can play a maximum role in the network and studying and summarizing existing algorithms, we designed an adaptive multiple loss baseline (AML) with a simple structure but powerful functions. In this network, we use an adaptive mining sample loss (AMS) and other modules, which can mine more information from input samples at the same time. Based on triplet loss, AMS loss can optimize the distance between the input sample and its positive and negative samples and protect structural information within the sample. During the experiment, we conducted several group tests and confirmed the high performance of AML baseline via the results. AML baseline has outstanding performance in three commonly used datasets. The two indicators of AML baseline on CUHK-03 are 25.7% and 26.8% higher than BagTricks.

Keywords: deep learning; person reidentification; baseline

Citation: Huang, Z.; Wang, L.; Li, Y.; Du, A.; Jiang, S. Improving Performance in Person Reidentification Using Adaptive Multiple Loss Baseline. *Information* **2022**, *13*, 453. https://doi.org/10.3390/info13100453

Academic Editor: Francesco Camastra

Received: 15 July 2022
Accepted: 20 September 2022
Published: 26 September 2022

1. Introduction

With the development of deep neural networks and the demand for application scenarios, person reidentification (Re-ID) technology has better development prospects. What this research usually needs to solve is occlusion, similar appearance, and illumination change. Several novel and effective Re-ID models [1,2] based on deep learning have been proposed. In addition, Re-ID models [3–5] based on attention mechanisms have also achieved many encouraging results. A popular method is to capture the global and local features of pedestrians by attention mechanism. Yang et al. found that the semantic information obtained by global features also included interference information (background interference information, etc.). To solve this problem, Chen et al. designed a hybrid higher-order attention network to utilize complex higher-order statistical information through the attention mechanism. Chen et al. proposed an attention but diversity network (ABD-Net) to apply a complementary mechanism to the attention mechanism.

However, researchers often focus on achieving a more robust model and ignore the baseline research. The current baseline research is insufficient to support high-precision model research. Specifically, BagTricks [6] is a baseline with high performance. We show the visualization results of BagTricks in Figure 1. We observed that in Figure 1a,c, the target pedestrian is blocked by another pedestrian, which brings the wrong retrieval to the search results. In Figure 1b, the clothes of the target pedestrian are very similar to another person, which makes BagTricks [6] retrieve some wrong visualization results.

After studying and summarizing the existing algorithms, we designed an adaptive multiple loss baseline with a simple structure but powerful functions for Re-ID.

There are two reasons why we design a simple but powerful baseline. Firstly, to extract rich and representative global and local features, most researchers work on constructing

deep convolutional neural networks [7–9]. Zhang et al. [7] designed a second-order nonlocal attention model (SONA), which effectively represents the local information of the target pedestrian through second-order features. In addition, Zheng et al. [8] posed a pyramid network to integrate global with local information of input pictures. Alemu et al. [9] proposed the idea of limiting samples, which can alleviate the problem of appearance similarity to a certain extent. Zhang et al. [10] designed a semantic dense arrangement framework (DSA) to effectively alleviate the occlusion phenomenon encountered in the process of pedestrian recognition. Zhang et al. [11] designed global relational awareness (RAG) to extract contextual semantic information through research. Usually, researchers apply the new method to a strong baseline to achieve a high retrieval readiness network. Through comparative experiments, we found that the performance of the network is very different when the same module is applied to different baselines.

Figure 1. Ranking performance of BagTricks on DukeMTMC-ReID. The green box represents the correct target pedestrian image retrieval, and the red box represents the wrong retrieval results. (**a**) Sample 1; (**b**) Sample 2; (**c**) Sample 3.

Second, we conducted a detailed survey of current articles on Re-ID baselines [6,12,13]. Specifically, BagTricks [6] is a high-performance baseline that combines six tricks. Xiong et al. [12] put batch normalization behind global pooling to improve network performance. In particular, Sun et al. [13] designed a baseline to extract pedestrian features based on partial convolution. Ye et al. [14] posed a robust AGW baseline, which uses the nonlocal attention [15] module based on renet-50 [16]. In addition, we find that the great difference between these baselines is that the loss functions are different. The retrieval accuracy of the model on small datasets is not satisfactory, and the network performance is relatively poor. Researchers usually use ID loss and triplet loss [17] to build models, as the triplet loss can increase the distance between the input image and the negative sample. Wu et al. [18] posed a view that loss of the triplet state would disrupt the internal information of the sample, and the existence of hard negative samples may lead to model collapse. We also introduced an adaptive mining sample loss (AMS) based on the triplet loss. AMS loss can automatically give the appropriate distance to the sample group, which can effectively avoid the misjudgment of samples (negative samples are misjudged as positive samples). We use triplet loss and AMS in the designed baseline, and the trained model has high retrieval accuracy.

In summary, the manuscript has the following contributions:

- Based on the triplet loss, the designed AMS loss can greatly improve the performance of the model. The simple but robust characteristics make the network have not only high accuracy but also strong practicability.
- We posed a robust and simple baseline, which achieves 82.3% mAP and 85.6% Rank-1 on the dataset of CUHK-03. This result is 25.7% and 26.8% higher than the current strong baseline BagTricks [6], respectively.
- We also carried out comparative and ablation experiments, such as embedding some novel methods or replacing the backbone, to prove that the baseline proposed is valid for Re-ID tasks.

2. Related Work

In this part, we mainly investigate the loss function used in the current Re-ID model. It is certain that the loss function plays an obvious role in model optimization, and the performance gap of models trained by different loss functions is obvious. Therefore, we must first use the appropriate loss function to design an excellent Re-ID model.

In Re-ID, ID loss and triplet loss [17] are often used because ID loss can accurately assign input samples to their classes. The model can narrow the distance between the input and the positive sample under the effect of triplet loss, and it can also increase the distance between the input and the negative sample. However, the performance of the trained network using only ID loss is not enough. Therefore, we introduce a metric learning method, which can adaptively learn the metric distance of the target sample. According to the similarity between input samples, they can classify the input images into image categories with high similarity. Most researchers combine ID with triplet loss, and they find that this strategy can train a good network model. However, the triplet loss function only pays attention to the positive samples, but it is easy to ignore the internal structure of the samples. Wu et al. [18] believed that triplet loss might destroy the internal information of samples, and the existence of hard negative samples may lead to model collapse. At the same time, the purpose of the recognition task is to retrieve the most similar images to the input sample from the gallery.

A triplet loss has the following three parts, namely, negative samples, positive samples, and target samples. Target samples and positive samples belong to the same identity; on the contrary, target samples and negative samples do not belong to the same identity, and triplet loss can reduce the distance between target samples and positive samples and increase the distance between target samples and negative samples. The AMS loss can automatically give the appropriate distance to the sample group, which can effectively avoid the misjudgment of samples (negative samples are misjudged as positive samples). AMS loss provides a safe distance for the sample group by learning a hypersphere for each class. We use triplet and AMS loss in the designed baseline, and the trained model has high retrieval accuracy.

In addition, a simple and high-accuracy model is more popular in practical applications. However, many models studied by the academic community are complex, and researchers often pile up some modules to extract more features of the target pedestrian. These models are not practical because of their high complexity. It is worth mentioning that for Tricks [19], Luo et al. summarized and applied some practical skills, which improved the accuracy without increasing or adding small network complexity. Therefore, we applied these training tricks to the proposed strong baseline.

3. The Proposed Baseline

In this part, we present a concrete framework of the designed baseline and the adaptive multiple loss.

3.1. The Pipeline of Baseline

AML baseline is an uncomplicated but effective Re-ID network. At present, most researchers choose the resnet-50 [16] as the backbone model; this is because the backbone

network can effectively prevent gradient explosion and make the network converge rapidly. The baseline we designed still uses the resnet-50 backbone network, which is convenient for comparison with other advanced algorithms. Figure 2 shows the main frame of the baseline presented in this manuscript. There are four key parts in the baseline: ResNet-50, GeM, BN layers, and ID + Triplet + AMSL (loss function part).

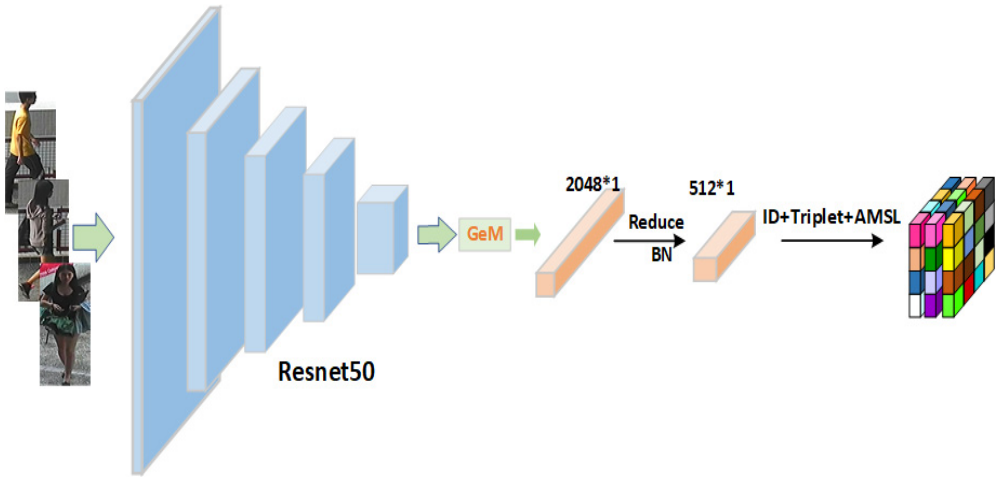


Figure 2. Basic structure of the proposed baseline.

In addition, we used some small but very useful techniques on the baseline model. The performance of the baseline is effectively improved without changing the network structure and model complexity. Training tricks include: last stride [13], random erasing [19], data augmentation [20], warmup learning rate [21], and label smoothing [22]. Specifically, we use the strategy of decreasing the learning rate. In other words, with the increase in epoch, we regularly reduce the learning rate to obtain a convergent model. The specific settings are shown in Table 1.

Table 1. The setting of learning rate.

Epoch	Lr	Epoch	Lr
1–10	$5.0 \times 10^{-4} \times 0.1$ t		
11–50	5.0×10^{-4}	161–210	4.0×10^{-6}
51–90	1.0×10^{-4}	211–250	8.0×10^{-7}
91–160	2.0×10^{-5}	251–300	1.6×10^{-7}

3.2. Adaptive Mining Sample Loss

In this part, we mainly introduce the adaptive multivariate loss function used for the baseline. Among them, we focus on AMS loss because it can concentrate the positive and negative samples on the same hypersphere to effectively improve the performance of the model. AML loss adaptively weights negative sample loss to distinguish hard negative samples, which can alleviate the problem of appearance similarity to a certain extent. Specifically, if the negative samples are particularly similar to the input image features, AMSL will assign greater weight to the loss values of these negative samples. If the positive samples are particularly similar to the input image features, AMSL assigns a smaller weight to the loss values of these negative samples, which enables the model to identify hard positive samples and extract sample features.

We first explain some formulaic symbols and what they represent to understand AML loss. $X = \{(x_i, y_i)\}_{i=1}^{N}$ is the training set. (x_i, y_i) represents the sample and label with sequence number i. The total number of training samples is C and $y_i \in [1, 2, \ldots, C]$. $\left\{\left(x_i^c\right)\right\}_{i=1}^{N_c}$

represents a collection of all samples. $P^*_{c,i}$ represents positive samples and $N^*_{c,i}$ represents the negative samples. The safe distance between samples is α.

$$L_m(x_i, x_j; f) = y_{ij}{d_{ij}}^2 + (1 - y_{ij})|\alpha - d_{ij}|^2 \tag{1}$$

where the cosine distance between the samples and another is $d_{ij} = |f(x_i) - f(x_j)|$. $L_m(x_i, x_j; f)$ loss can keep the negative sample points away from the input image.

Given an input image X_i, our purpose is to make the input sample closer to the positive sample point and obtain a hypersphere with radius α-m. We also need to train all the positive sample points related to input samples.

$$L_{amsl}(x_i, x_j; f) = (\alpha + y_{ij}(d_{ij} - m) + (y_{ij} - 1)d_{ij}) \tag{2}$$

where $y_{ij} \in \{0, 1\}$. When $y_i = y_j$, $y_{ij} = 1$; otherwise, $y_{ij} = 0$.

The basic equation of the baseline algorithm in this section is L_{amsl} in Equation (3). Positive sample loss can concentrate all positive samples of the input image in a sphere with radius α-m. In addition, we extract the positive sample information by calculating the Euclidean distance ${d_{ij}}^n$ between the positive sample and the input sample, and, finally, assign weights to the positive sample loss by *softmax*. Below is the specific calculation formula:

$$L_p(x_i^c; f) = \sum_{x_j^c \in P^*_{c,i}} \frac{\exp({d_{ij}}^n)}{\sum_{x_j^c \in P^*_{c,i}} \exp({d_{ij}}^n)} L_{amsl}(x_i, x_j; f) \tag{3}$$

Our purpose is to keep the x_i^c away from the negative sample, so the negative sample loss can achieve a minimum safety distance α. In addition, there is a lot of hard sample information that is difficult to extract from the dataset. Therefore, we adaptively assign weights to negative sample loss by the *softmin* function. The specific calculation formula is as follows:

$$L_n(x_i^c; f) = \sum_{x_j \epsilon N^*_{c,i}} \frac{\exp(-{d_{ij}}^n)}{\sum_{x_j^c \epsilon |N^*_{c,i}|} \exp(-{d_{ij}}^n)} L_{amsl}(x_i^c; x_j^c; f) \tag{4}$$

We find that the partial derivative of L_{amsl} is always 1. This shows that we apply the weighted strategy to model training. The specific calculation formula is as follows:

$$\left| \frac{\partial L_{amsl}\left(x_i^c; x_j^c; f\right)}{\partial f(x_i)} \right| = \left| \frac{\alpha + d_{ij} - y_{ij}\beta}{\partial f(x_i)} \right| = \left| \frac{2(f(x_i) - f(x_j))}{2|f(x_i) - f(x_j)|} \right| = 1 \tag{5}$$

Then we calculate both positive and negative sample losses by the AMSL function and optimize them together.

$$L_{amsl}(x_i^c; f) = L_p(x_i^c; f) + L_n(x_i^c; f) \tag{6}$$

Finally, we use the AML loss function based on ID loss and triplet loss, and the specific strategies are as follows:

$$L_{aml} = L_{id} + w_1 L_{amsl} + w_2 L_{triplet} \tag{7}$$

where L_{amsl} is the final adaptive multiple loss function of the proposed baseline, L_{id} is the cross-entropy loss function, and $L_{triplet}$ is the triplet loss function. We need to fine-tune the coefficients w_1 and w_2.

4. Results and Discussion

The experimental framework is Pytorch1.3, and the server is Tesla V100 GPUs. This manuscript evaluates AML baselines on three large public datasets: Market-1501 [23], DukeMTMC-ReID [24], and CUHK-03 (detected) [25]. Firstly, this paper compares the

performance of AML baseline with the latest method. Then, this paper shows the related hyperparameter processing experiments and ablation experiments.

4.1. Datasets and Evaluation Metrics

Datasets: Through collection and investigation, the authors of this manuscript decided to evaluate AML baseline through three public datasets. This paper is detailed in Table 2. Specifically, Market-1501 contains 32,668 images of 1501 tagged samples with six cameras. As a large dataset, DukeMTMC-ReID has 36,411 images with 1404 identities from eight countries. The training set has 16,522 images with 702 identities, and the test set has 19,889 images with 702 identities. CUHK-03 is a Re-ID dataset with 14,088 images of 1467 identities.

Table 2. The specific introduction of the dataset.

Datasets	Train ID	Images	Test ID	Images	Sum	Cameras
Market-1501	751	12,936	750	19,732	32,668	6
DukeMTMC-ReID	702	16,522	702	19,889	36,411	8
CUHK-03	767	7356	700	6732	14,088	2

Evaluation Metrics: Person Re-ID tasks often use mAP and the cumulative matching function (CMC) to evaluate the designed model. The CMC curve can determine whether the candidate image is included in the first k items of the list (the retrieval results that belong to the same identity as the input image), and it can calculate the correct matching rate of the same pedestrian in the first k items of the list. However, the CMC curve only focuses on whether the first k items in the ranking list have correct retrieval results; it does not take the recall rate into account. To solve this problem, some researchers proposed mAP as a supplement. mAP can intuitively show the average accuracy of correct retrieval.

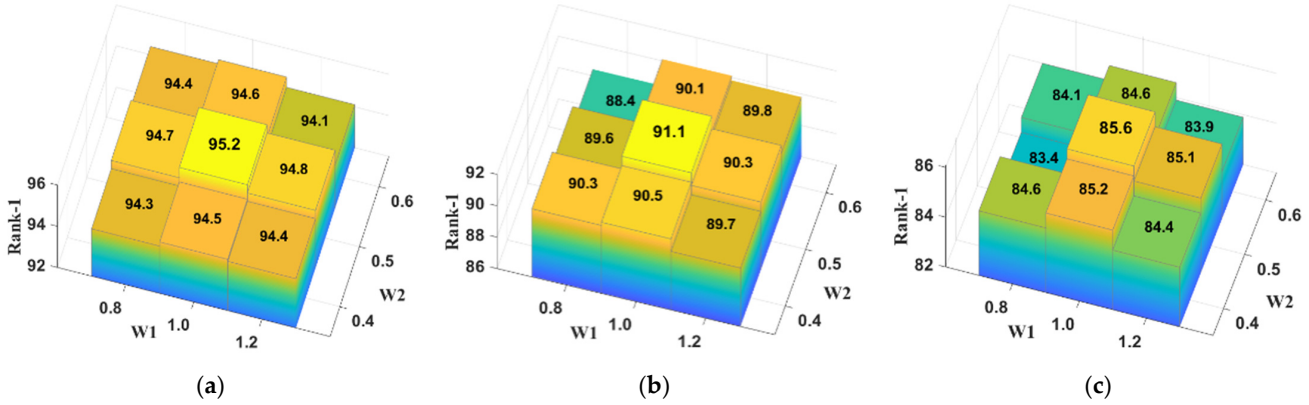
4.2. Hyperparameter Optimization

It is important to note that our baseline has the characteristics of low complexity and short training time. This paper analyzes the parameters of the loss function section. Because there are four parameters, this paper adopts the following strategy. Firstly, we use ID loss and the AMSL loss function to determine the appropriate values of α and m through comparative experiments. Then we use the triplet loss function and analyze w_1 and w_2 . Finally, this paper combines the triplet loss and AMS loss to analyze α and m again.

For example, we fix the parameter $w_1 = 0.5$. Then, the value of w_2 is changed, and the best experimental results are selected through many experiments. In this paper, the approximate value of w_2 is found by changing 0.1 at a time. After that, this paper changes 0.05 at a time to obtain the final value of w_2. Finally, the parameter w_1 is fixed, and the optimum value of w_2 is found. Figure 3 shows the specific optimization results. It can be intuitively found that the coefficients w_1 and w_2 in Figure 3 have a significant influence on the baseline model. We notice that when $w_1 = 1.0$ and $w_2 = 0.5$, AML baseline has the best performance.

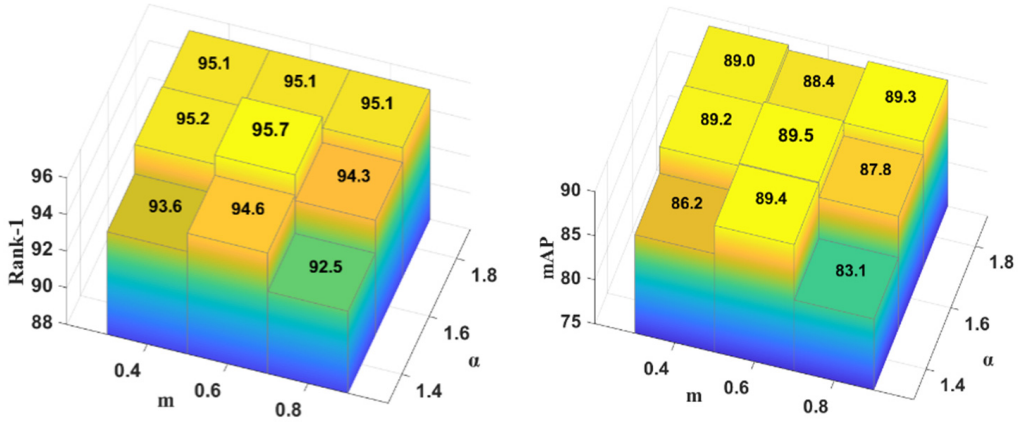
Coefficient w_1 and w_2 : In the loss function part, it is found that different combinations of coefficients have a great influence on the experimental results. Therefore, we need to adjust the coefficient of triplet loss and AML loss appropriately. We set the initial values to $w_1 = 0.5$ and $w_2 = 0.5$ and obtain the final parameter values by using this control variable method [26].

Hyperparameter m and α: In addition, parameters m and α in the loss function section also have significant effects on the performance of AML baselines. The size of α-m determined by the data set represents the radius of the hypersphere. This manuscript makes appropriate adjustments to ensure the training of high-performance models. Similarly, we use the control variable method here to adjust these two parameters. The optimized results are shown in Figures 4–6.



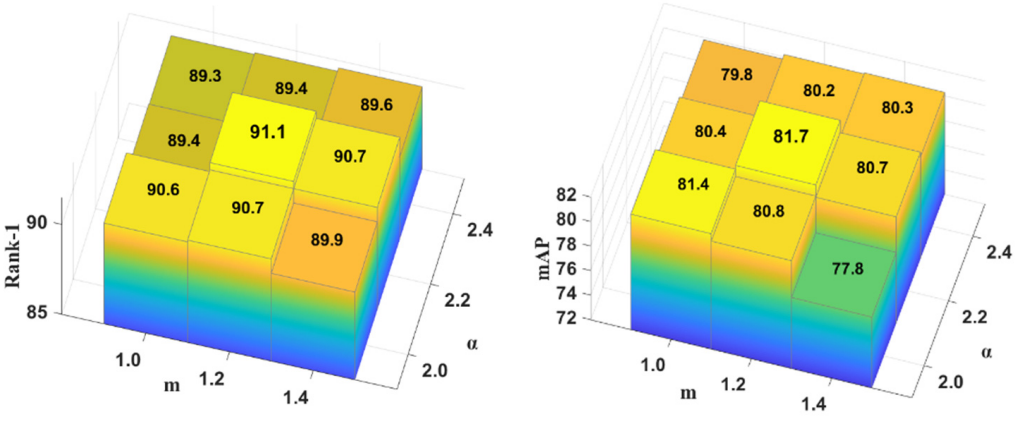
(**a**) (**b**) (**c**)

Figure 3. Optimization of AML baseline loss functions on three datasets: (**a**) the Market-1501 dataset; (**b**) the DukeMTMC-ReID dataset; (**c**) the CUHK-03 (detected) dataset.



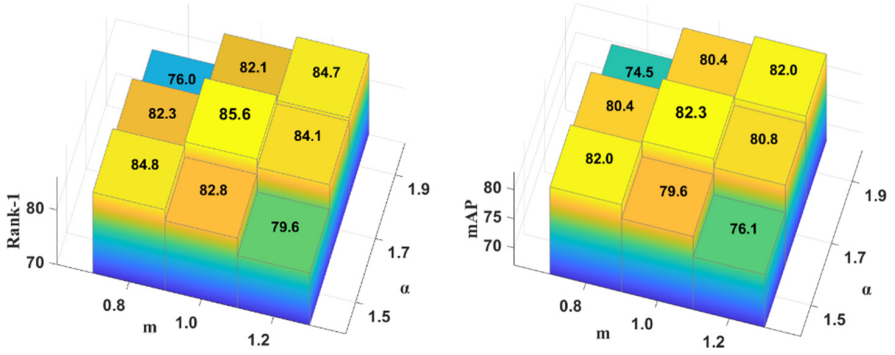
(**a**) (**b**)

Figure 4. Optimization of AML baseline loss functions on Market-1501 datasets. (**a**) Results of different hyperparameters on Rank-1; (**b**) Results of different hyperparameters on MAP.



(**a**) (**b**)

Figure 5. Optimization of AML baseline loss functions on DukeMTMC-ReID datasets. (**a**) Results of different hyperparameters on Rank-1; (**b**) Results of different hyperparameters on MAP.



(**a**) (**b**)

Figure 6. Optimization of AML baseline loss functions on CUHK-03 (detected) datasets. (**a**) Results of different hyperparameters on Rank-1; (**b**) Results of different hyperparameters on MAP.

Results on Market-1501 dataset: When the values of hyperparameters α and m are 0.6 and 1.6, respectively, AML has the best baseline performance, namely Rank-1 reaches 95.7%, and mAP reaches 89.5%. Figure 4 shows the optimization results of hyperparameters α and m on Market-1501. We can find that no matter whether the distance between the hyperparameters α and m is too large or too small, the AML baseline performance will decline.

Results on DukeMTMC-ReID dataset: When the values of hyperparameters α and m are 1.2 and 2.2, respectively, AML has the best baseline performance, namely Rank-1 reaches 91.1%, and mAP reaches 81.9%. Figure 5 shows the optimization results of hyperparameters α and m on this bigger dataset. Similarly, we can find that no matter whether the distance between the hyperparameters α and m is too large or too small, the AML benchmark accuracy will decline.

Results on CUHK-03 (detected) dataset: When the values of hyperparameters α and m are 1.0 and 1.7, respectively, AML baseline has the best baseline performance, namely Rank-1 reaches 91.1%, and mAP reaches 82.3%. Figure 6 shows the optimization results of hyperparameters α and m on this smaller dataset. Similarly, we can find that no matter whether the distance between the hyperparameters α and m is too large or too small, the AML benchmark accuracy will decline.

4.3. Compared with Advanced Baselines and Different Loss Functions

In this section, we compare AML baseline in two cases. The first case is to compare AML baseline with the current more advanced baseline algorithms; for example, AWTL [27] baseline uses the triplet loss function to optimize the baseline network, GP [12] and PCB [13] uses ID loss to train the model, and BagTricks [6] uses ID, triplet, and center loss to optimize the baseline. AMSL [28] uses ID and AMS loss to optimize the baseline model. The second case is to compare AML baseline with different loss functions such as ID, triplet, AMS, circle [29], and contrastive [30] loss. Table 3 shows the specific results.

On Market-1501: Since the current algorithms perform well on this dataset, the improvement in the AML baseline on this dataset is not obvious. The AML baseline reaches 89.5% on mAP value and 95.7% on Rank-1 value. The evaluating indicator of AML baseline is 1.1% and 0.9% higher than BagTricks. In addition, AML baseline is 1.4% in mAP and 0.6% in Rank-1 higher than AMSL baseline. For the results using different loss functions, the combination of ID, triplet, and AMS loss in AML baseline is 1.1% in mAP and 0.6% in Rank-1 higher than that of ID, triplet, and circle loss.

On DukeMTMC-ReID: AML baselines also have high retrieval accuracy on this dataset with more pedestrian identities. The AML baseline reaches 81.7% on mAP and 91.1% on Rank-1. The mAP and Rank-1 of AML baseline is 6.9% and 4.5% higher than that of BagTricks. In addition, the evaluating indicator of AML baseline is 3.5% and 1.7% higher than AMSL baseline. For the results using different loss functions, the combination of ID,

triplet, and AMS loss in AML baseline is 1.4% in mAP higher than that of ID, triplet, and circle loss, and 1.1% in Rank-1 higher than that of ID, triplet, and contrastive loss.

Table 3. Comparison of advanced baselines and different loss functions on three datasets.

Baseline	Loss	Market-1501		DukeMTMC-ReID		CUHK-03	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
AWTL [27]	Trip	75.7	89.5	63.4	79.8	-	-
GP [12]	ID	78.8	91.7	68.8	83.4	-	-
PCB [13]	ID	81.6	93.8	69.2	83.3	57.5	63.7
AMSL [28]	ID + AMSL	88.1	95.1	78.2	89.4	75.6	78.4
BagTricks [6]	ID + Tri + Center	88.4	94.8	74.8	86.6	56.6	58.8
	ID + Circle	86.8	94.5	79.7	88.9	75.2	78.0
	ID + Tri + Circle	88.4	95.1	80.3	89.3	77.2	80.1
	ID + Tri + Contrastive	87.9	94.4	78.4	90.0	81.8	78.4
Ours	ID + Tri + AMSL	89.5	95.7	81.7	91.1	82.3	85.6

On CUHK-03: The improvement in AML baseline on this small dataset is the most obvious. The AML baseline reaches 82.3% on mAP and 85.6% on Rank-1. The mAP and Rank-1 of AML baseline is 25.7% and 26.8% higher than that of BagTricks. In addition, the mAP and Rank-1 of AML baseline are 6.7% and 7.2% higher than AMSL baseline. For the results using different loss functions, the combination of ID, triplet, and AMS loss in AML baseline is 0.5% in mAP higher than that of ID, triplet, and contrastive loss, and 5.5% in Rank-1 higher than that of ID, triplet, and circle loss.

The above experiments confirm that AML baseline is a high-precision model. Therefore, AML baseline performs better than advanced baseline algorithms have in recent years.

4.4. Compared with the State-of-the-Art

In addition, we also compare AML baseline with the other advanced Re-ID algorithms. Tables 4–6 show the specific results. In this paper, AML baseline was compared with the most advanced (SOTA) methods: AGW [14], CAM [3], ABD-Net [5], BagTricks [6], AANet [1], RAG-SC [11], SGSC [31], GLWR [2], MGN [32], GD-Net [33], IANet [34], Pyramid [8], SONA [7], SCAL [35], Auto-ReID+ [36], Ms-Mb [37], TransReID [38], CAL [39], and PAN [40]. Admittedly, these papers do not take the skill of reranking. The backbone of most of the above algorithms is ResNet-50. However, the backbone network of Pyramid and TransReID are ResNet-101 and ViT [41].

Table 4. Comparison of advanced baselines on Market-1501.

Methods	Rank-1	mAP
AANet [1] (CVPR2019)	93.9	83.4
IANet [34] (CVPR2019)	94.4	83.1
PAN [40] (IEEE TCSVT 2019)	91.45	87.44
SONA [7] (CVPR2019)	95.6	88.8
AGW [14] (arXiv2020)	95.1	87.8
ABD-Net [5] (CVPR2019)	95.6	88.3
MGN [32] (ACM2018)	95.7	86.9
CAM [3] (CVPR2019)	94.7	84.5
GD-Net [33] (CVPR2019)	94.8	86.0
BagTricks [6] (CVPR2019)	94.5	85.9
GLWR [2] (IEEE Access 2020)	95.5	88.5
Pyramid [8] (CVPR2019)	95.7	88.2
SGSC [31] (four stages) (CVPR2020)	95.7	88.5
CAL [39] (ICCV2021)	95.5	89.5
TransReID [38] (ICCV2021)	95.2	88.9
AML baseline (ours)	95.7	89.5

Table 5. Comparison of advanced baselines on DukeMTMC-ReID.

Methods	Rank-1	mAP
AANet [1] (CVPR2019)	87.7	74.3
IANet [34] (CVPR2019)	83.1	73.4
PAN [40] (IEEE TCSVT 2019)	75.94	66.74
SONA [7] (CVPR2019)	89.5	78.3
AGW [14] (arXiv2020)	89.0	79.6
ABD-Net [5] (CVPR2019)	89.0	78.6
MGN [32] (ACM2018)	88.7	78.4
CAM [3] (CVPR2019)	85.8	72.9
GD-Net [33] (CVPR2019)	86.6	74.8
BagTricks [6] (CVPR2019)	86.4	76.4
SCAL (spatial) [35] (ICCV2019)	89.0	79.6
SCAL (channel) [35] (ICCV2019)	88.9	79.1
GLWR [2] (IEEE Access 2020)	90.7	81.4
Pyramid [8] (CVPR2019)	89.0	79.0
Auto-ReID+ [36] (Neurocomputing2021)	90.1	80.1
SGSC [31] (four stages) (CVPR2020)	91.0	79.0
CAL [39] (ICCV2021)	90.0	80.5
TransReID [40] (ICCV2021)	90.7	82.0
AML baseline (ours)	91.1	81.7

Table 6. Comparison of advanced baselines on CUHK-03.

Methods	Rank-1	mAP
PAN [40] (IEEE TCSVT 2019)	41.9	43.8
SONA [7] (CVPR2019)	79.9	77.3
AGW [14] (arXiv2020)	63.6	62.0
RAG-SC [11] (CVPR2020))	79.6	74.5
MGN [32] (ACM2018)	68.0	66.0
CAM [3] (CVPR2019)	66.6	64.2
BagTricks [6] (CVPR2019)	58.8	56.6
SCAL (channel) [35] (ICCV2019)	71.1	68.6
GLWR [2] (IEEE Access 2020)	82.3	78.9
Pyramid [8] (CVPR2019)	78.9	74.8
Auto-ReID+ [36] (Neurocomputing2021)	78.1	74.2
Ms-Mb [37] (Neurocomputing2020)	75.4	72.9
SGSC [31] (four stages) (CVPR2020)	84.7	81.0
AML baseline (ours)	85.6	82.3

Results on the Market-1501 dataset: The comparison results of algorithms are shown in Table 4. This paper shows the specific data of the algorithm comparison in Table 4. In these methods, what needs to be pointed out is that Pyramid uses a more powerful global feature algorithm to represent the complex features of the trunk and local. SGCS extracts the potential features of different stages by a cascade strategy and integrates the features of each stage for final representation. TransReID processes the partitioned image by using Transformer [42] to realize feature learning of large-scale and long-distance spatial structures and proposes a pure Transformer framework for the object Re-ID task. However, AML baseline, without any improvement in the network part, is more competitive than many SOTA methods, whether they are based on CNN or Transformer. AML baseline achieves the same performance as the frontier algorithm. In addition, it should be noted that the results of AML baselines on the lightweight network ResNet-50 also achieve the performance of the Pyramid algorithm.

Results on DukeMTMC-ReID dataset: This paper shows the comparison results in Table 5. This dataset has various samples that contain many different identities of pedestrians, and AML baseline achieves a high evaluation index. Without any improvement in the network part, the Rank-1 index of AML baseline is higher than many current advanced algorithms. In addition, the results of this paper on the lightweight network are 2.7 percentage points higher than the performance of the Pyramid algorithm on the mAP.

Results on CUHK-03 dataset: The comparison results of algorithms are shown in Table 6. Compared with the above two datasets, CUHK-03 is a more challenging small dataset because it has fewer samples and serious occlusion problems. However, AML baselines can still extract pedestrian features to accurately retrieve the target pedestrian. AML baselines exceeded Pyramid by 6.7% and 7.5% on these two indicators. Furthermore, the performance of AML baseline is better than that of the high-performance SGSN algorithm.

Specifically, Pyramid achieved 95.7 % Rank-1 accuracy and 88.2% mAP accuracy on Market-1501. SGSN achieved 91.0 % Rank-1 accuracy on DukeMTMC-ReID. In addition, SGSN obtained 84.7% Rank-1 accuracy and 81.0% mAP accuracy on CUHK-03, as shown in Table 6. However, the results of AML baselines clearly reach or exceed these algorithms. Especially on the CUHK-03 dataset, AML baselines in Rank-1 and mAP exceed SGSN by 0.9% and 1.3%. Through comprehensive analysis of these experimental results, AML baseline has the best performance. The mAP values of AML baseline on the three datasets are 89.5%, 81.7%, and 82.3%, respectively, which are higher than those of the above suboptimal networks. This baseline reaches a new SOTA on these three common public datasets.

4.5. Ablation Experiments

In this section, we analyze the effectiveness of components of the AML baseline by ablative experiments. The specific results are as Table 7 shows.

Table 7. Ablation experiments of AML baseline.

Methods	Market-1501		DukeMTMC-ReID		CUHK-03	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
B	93.7	83.9	85.6	74.8	60.2	55.0
+GeM	94.1	84.5	86.8	75.4	62.1	56.4
+Triplet	95.1	87.9	89.9	80.0	63.8	62.7
+AMSL	95.7	89.5	91.1	81.7	85.6	82.3

BagTricks of cross-entropy loss function is used as our comparison algorithm. Then, this paper adds the generalized mean pooling, triplet, and AMSL loss functions to the baseline. Generalized mean pooling can eliminate the redundant information of input samples and also extract the characteristics of specific regions of samples. This paper finds that these two loss functions can optimize the baseline model effectively. As described earlier in this manuscript, the network trained by this multivariate loss strategy is robust. AML baselines are prominent on the three datasets. Our baseline has the highest accuracy improvement on the CUHK-03 (detected) dataset. Finally, it can be found that AML baselines combined with all modules have high accuracy.

5. Visualization Results

This manuscript shows the visualization results of BagTricks [6] in Figure 1. Due to the occlusion and similar appearance of target pedestrians, there are many errors in the retrieval results of this algorithm. Figure 7 shows the visualization of sorting results for AML baselines (same input samples). Although the input samples have problems with occlusion and similar appearance, the AML model can still accurately retrieve images with the same identity as the input sample. This also confirms that AML baselines can alleviate occlusion and similar appearance problems to some extent.

Figure 7. The visualization of sorting results for AML baseline. (**a**) Sample 1; (**b**) Sample 2; (**c**) Sample 3.

6. Conclusions

In this manuscript, we designed an uncomplicated but effective Re-ID baseline. AML baseline reached 82.3% mAP value and 85.6% Rank-1 value on CUHK-03. These two indexes are higher than the current popular advanced baseline algorithm. At the same time, the AML baseline model has low complexity and fast convergence speed. Existing baselines generally use ID and triplet loss training models. However, we notice that triplet loss will disrupt the original information within the sample. AML adopts adaptive sample mining loss based on ID and triplet loss, and this adaptive multivariate loss strategy can realize the spatial structure of samples. AML baseline model has high accuracy on both large and small datasets, which effectively alleviates occlusion and appearance similarity problems in the retrieval process. Finally, we hope our baseline will help Re-ID task research.

Author Contributions: Conceptualization and methodology, Z.H.; software, Z.H. and Y.L.; validation, Y.L. and L.W.; formal analysis, L.W. and A.D.; data curation and writing original draft preparation, A.D.; writing—review and editing, S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation of China under grant number U1903213, and the Tianshan Innovation Team of Xinjiang Uygur Autonomous Region grant number 2020D14044.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We evaluated our network on three large public datasets: Market-1501, DukeMTMC-ReID, and CUHK-03 (https://github.com/NEU-Gou/awesome-reid-dataset, accessed on 24 May 2022).

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

References

1. Tay, C.-P.; Roy, S.; Yap, K.-H. Aanet: Attribute attention network for person re-identifications. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 7127–7136. [CrossRef]
2. Gong, Y.; Wang, L.; Li, Y.; Du, A. A discriminative person re-identification model with global-local attention and adaptive weighted rank list loss. *IEEE Access* **2020**, *8*, 203700–203711. [CrossRef]

3. 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 1389–1398. [CrossRef]
4. Chen, B.; Deng, W.; Hu, J. Mixed high-order attention network for person re-identification. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 371–381. [CrossRef]
5. 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 (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 8350–8360. [CrossRef]
6. Luo, H.; Jiang, W.; Gu, Y.; Liu, F.; Liao, X.; Lai, S.; Gu, J. A strong baseline and batch normalization neck for deep person re-identification. *IEEE Trans. Multimed.* **2019**, *22*, 2597–2609. [CrossRef]
7. Xia, B.N.; Gong, Y.; Zhang, Y.; Poellabauer, C. Second-order non-local attention networks for person re-identification. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 3759–3768. [CrossRef]
8. Zheng, F.; Deng, C.; Sun, X.; Jiang, X.; Guo, X.; Yu, Z.; Huang, F.; Ji, R. Pyramidal person re-identification via multi-loss dynamic training. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 8506–8514. [CrossRef]
9. Alemu, L.T.; Pelillo, M.; Shah, M. Deep constrained dominant sets for person re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9855–9864. [CrossRef]
10. Zhang, Z.; Lan, C.; Zeng, W.; Chen, Z. Densely semantically aligned person re-identification. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 667–676. [CrossRef]
11. Hang, Z.; Lan, C.; Zeng, W.; Jin, X.; Chen, Z. Relation-aware global attention for person re-identification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 3183–3192. [CrossRef]
12. Xiong, F.; Xiao, Y.; Cao, Z.; Gong, K.; Fang, Z.; Zhou, J.T. Good practices on building effective CNN baseline model for person re-identification. In Proceedings of the Tenth International Conference on Graphics and Image Processing (ICGIP 2018), Chengdu, China, 12–14 December 2018; pp. 1–2. [CrossRef]
13. 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; Volume 11208, pp. 480–496. [CrossRef]
14. Ye, M.; Shen, J.; Lin, G.; Xiang, T.; Shao, L.; Hoi, S.C.H. Deep learning for person re-identification: A survey and outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 2872–2893. [CrossRef] [PubMed]
15. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803. [CrossRef]
16. 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. [CrossRef]
17. Liu, H.; Feng, J.; Qi, M.; Jiang, J.; Yan, S. End-to-end comparative attention networks for person re-identification. *IEEE Trans. Image Process.* **2017**, *26*, 3492–3506. [CrossRef] [PubMed]
18. Wu, C.-Y.; Manmatha, R.; Smola, A.J.; Krahenbuhl, P. Sampling matters in deep embedding learning. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October·2017; pp. 2859–2867. [CrossRef]
19. Luo, H.; Gu, Y.; Liao, X.; Lai, S.; Jiang, W. Bag of tricks and a strong baseline for deep person re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 1487–1495. [CrossRef]
20. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random erasing data augmentation. In Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 13001–13008. [CrossRef]
21. Zheng, Z.; Zheng, L.; Yang, Y. A discriminatively learned CNN embedding for person reidentification. *ACM Trans. Multimed. Comput. Commun. Appl.* **2018**, *14*, 1–20. [CrossRef]
22. Fan, X.; Jiang, W.; Luo, H.; Fei, M. Spherereid: Deep hypersphere manifold embedding for person re-identification. *J. Vis. Commun. Image Represent.* **2019**, *60*, 51–58. [CrossRef]
23. 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, 7–13 December 2015; pp. 1116–1124. [CrossRef]
24. Ristani, E.; Solera, F.; Zou, R.; Cucchiara, R.; Tomasi, C. Performance measures and a data set for multi-target, multicamera tracking. In Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 17–35. [CrossRef]
25. 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, Columbus, OH, USA, 23–28 June 2014; pp. 152–159. [CrossRef]
26. Fan, D.; Wang, L.; Cheng, S.; Li, Y. Dual branch attention network for person re-identification. *Sensors* **2021**, *21*, 5839. [CrossRef] [PubMed]

27. Ristani, E.; Tomasi, C. Features for multi-target multi-camera tracking and 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. 6036–6046. [CrossRef]
28. Gong, Y.; Wang, L.; Cheng, S.; Li, Y. A strong baseline based on adaptive mining sample loss for person re-identification. In Proceedings of the CAAI International Conference on Artificial Intelligence 2021, Hangzhou, China, 5–6 June 2021; pp. 469–480. [CrossRef]
29. Sun, Y.; Cheng, C.; Zhang, Y.; Zhang, C.; Zheng, L.; Wang, Z.; Wei, Y. Circle loss: A unified perspective of pair similarity optimization. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 6397–6406. [CrossRef]
30. Hadsell, R.; Chopra, S.; LeCun, Y. Dimensionality reduction by learning an invariant mapping. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006; pp. 1735–1742. [CrossRef]
31. Chen, X.; Fu, C.; Zhao, Y.; Zheng, F.; Song, J.; Ji, R.; Yang, Y. Salience-guided cascaded suppression network for person re-identification. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 3297–3307. [CrossRef]
32. Wang, G.; Yuan, Y.; Li, J.; Ge, S.; Zhou, X. Receptive multi-granularity representation for person re-identification. *IEEE Trans. Image Process* **2020**, *29*, 6096–6109. [CrossRef] [PubMed]
33. Zheng, Z.; Yang, X.; Yu, Z.; Zheng, L.; Yang, Y.; Kautz, J. Joint discriminative and generative learning for person re-identification. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2138–2147. [CrossRef]
34. Hou, R.; Ma, B.; Chang, H.; Gu, X.; Shan, S.; Chen, X. Interaction-and-aggregation network for person re-identification. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9317–9326. [CrossRef]
35. Chen, G.; Lin, C.; Ren, L.; Lu, J.; Zhou, J. Self-critical attention learning for person re-identification. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9637–9646. [CrossRef]
36. Gu, H.; Fu, G.; Li, J.; Zhu, J. Auto-ReID+: Searching for a multi-branch ConvNet for person re-identification. *Neurocomputing* **2021**, *435*, 53–66. [CrossRef]
37. Jiao, S.; Pan, Z.; Hu, G.; Shen, Q.; Du, L.; Chen, Y.; Wang, J. Multi-scale and multi-branch feature representation for person re-identification—ScienceDirect. *Neurocomputing* **2020**, *414*, 120–130. [CrossRef]
38. He, S.; Luo, H.; Wang, P.; Wang, F.; Li, H.; Jiang, W. TransReID: Transformer-based object re-identification. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 11–17 October 2021; pp. 14993–15002. [CrossRef]
39. Rao, Y.; Chen, G.; Lu, J.; Zhou, J. Counterfactual attention learning for fine-grained visual categorization and re-identification. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 11–17 October 2021; pp. 1005–1014. [CrossRef]
40. Zheng, Z.; Zheng, L.; Yang, Y. Pedestrian alignment network for large-scale person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 3037–3045. [CrossRef]
41. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 $\times$ 16 words: Transformers for image recognition at scale. *arXiv* **2021**, arXiv:2010.11929. [CrossRef]
42. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Neural Inf. Process. Syst.* **2017**, 6000–6010. [CrossRef]