

Article

Person Re-Identification Based on Contour Information Embedding

Hao Chen , Yan Zhao * and Shigang Wang

College of Communication Engineering, Jilin University, Changchun 130012, China

* Correspondence: zhao_y@jlu.edu.cn

Abstract: Person re-identification (Re-ID) plays an important role in the search for missing people and the tracking of suspects. Person re-identification based on deep learning has made great progress in recent years, and the application of the pedestrian contour feature has also received attention. In the study, we found that pedestrian contour feature is not enough in the representation of CNN. On this basis, in order to improve the recognition performance of Re-ID network, we propose a contour information extraction module (CIEM) and a contour information embedding method, so that the network can focus on more contour information. Our method is competitive in experimental data; the mAP of the dataset Market1501 reached 83.8% and Rank-1 reached 95.1%. The mAP of the DukeMTMC-reID dataset reached 73.5% and Rank-1 reached 86.8%. The experimental results show that adding contour information to the network can improve the recognition rate, and good contour features play an important role in Re-ID research.

Keywords: person re-identification; pedestrian contour; contour information extraction

1. Introduction

Person re-identification (Re-ID), also known as cross-camera pedestrian tracking, aims to identify the same person image under different cameras [1]. Nowadays, face recognition has been widely used in all areas of life, and the research of face recognition has also developed rapidly [2–4]. However, in many cases, the camera cannot accurately capture a clear face, so person re-identification technology can play an important role [5].

The initial Re-ID relies on manual features such as color and texture [6,7]. Although it has a certain recognition effect, it still has a certain gap with human recognition. Nowadays, most of Re-ID is based on deep learning, on which researchers are committed to solving various practical problems encountered. Some research is aimed at Re-ID itself, such as solving posture misalignment, image occlusion [8,9], etc. Some research is aimed at deep learning networks [10], such as using attention mechanism, network improvement, small sample recognition [11], etc.

It is found that contour information also plays an important role in image recognition. Some works [12] deeply understand the expression of CNN's in-depth visual features. Experiments on Imagenet show that CNN-based deep learning models prefer texture-based features to shape-based features [13,14]. In addition, the embedded hybrid model based on texture and contour has been proven to improve the performance of image classification and object detection. We extract the feature map of the ResNet-50 network, which is the most commonly used in Re-ID research, and visualize it (see Figure 1). It can be intuitively found that the CNN network is missing in the extraction of contour information. In the past two years, many researchers began to use contour information as an auxiliary item to participate in Re-ID research, and achieved some success.

Based on the above existing facts, we believe that adding contour information to the CNN network can effectively help them learn more and more robust pedestrian information, and can improve the final recognition effect. In view of the end-to-end secrecy of CNN



Citation: Chen, H.; Zhao, Y.; Wang, S. Person Re-Identification Based on Contour Information Embedding. *Sensors* **2023**, *23*, 774. <https://doi.org/10.3390/s23020774>

Academic Editors: Guangtao Zhai, Xionghuo Min, Menghan Hu and Wei Zhou

Received: 14 December 2022

Revised: 7 January 2023

Accepted: 8 January 2023

Published: 10 January 2023



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

network, one cannot intervene with a human's preconceived knowledge. Therefore, we build contour information extraction module with the help of attention thought, so that the CNN network can pay more attention to some contour information instead of losing it in the process of multi-layer convolution. Essentially, the contributions of this paper are as follows:

- We verify the lack of contour information in CNN network in person re-identification research.
- We propose a contour information extraction module, which can make the network pay more attention to the contour information in the pedestrian image without the intervention of human experience.
- The experimental results show that our method has a good effect on Market1501 and DukeMTMC-reID datasets.

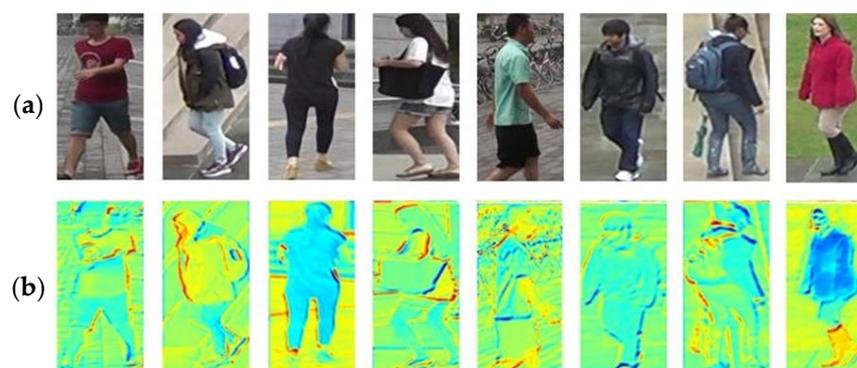


Figure 1. Mapping between original pedestrian images and feature maps (a) pedestrian images in common datasets. (b) Shallow features of RGB images in ResNet-50 network.

2. Related Works

2.1. Person Re-Identification Based on Attention Mechanism

In recent years, the attention mechanism has been widely used [15–18]. Its purpose is to enable the network to learn more important things. In Re-ID research, it can effectively solve the problem of posture misalignment, pedestrian image deviation, and partial occlusion. For example, Liu et al. [19] proposed an attention model into the Re-ID task for the first time to dynamically generate attention features to locate different local areas. Zhao et al. [20] proposed an attention model based on CNN, which uses the similarity information of paired human images to learn the part of the body used for matching. Wang et al. [21] proposed an attention model combining hard attention and soft attention, which can simultaneously learn multi-scale local and pixel level feature maps in an end-to-end manner. Sun et al. [22] proposed a local to global multi-scale attention network (LGMANet), which makes full use of context information and spatial attention information to further improve the recognition ability of the birth network. Zhang et al. [23] proposed an effective relationship-aware global attention (RGA) module, which can also be applied to scene segmentation and image segmentation tasks.

2.2. Person Re-Identification Based on Contour Information

In recent years, some researchers have applied pedestrian contour to Re-ID research; for example, Chen et al. [24] first attempted to utilize contour explicitly in deep Re-ID models, and proposed contour guidance, which greatly proves the application prospect of pedestrian contour. Yang et al. [25] believed that on the basis that the change of person's clothing is not strong (strong refers to the gap between winter clothing and summer clothing), the pedestrian contour also has the ability to distinguish personal characteristics. Based on this, the pedestrian contour map is used as the input of feature extraction for cross-dressing person Re-ID, and good results are achieved. Based on the research of Yang et al., Chen et al. [26] combined with the attention mechanism, used the pedestrian

contour to carry out the research on cross dressing Re-ID research. Recently released research [27] proposed a multi-scale appearance and contour deep infomax (MAC-DIM) to maximize mutual information between pedestrian color image features and pedestrian contour features, utilizing contour feature learning as regularization to mine more effective shape-aware feature representation from color images.

3. Methods

In this section, we will show the proposed contour information extraction module (CIEM) (as shown in Figure 2), and give a detailed description of the Re-ID method based on contour information embedding (as shown in Figure 3).

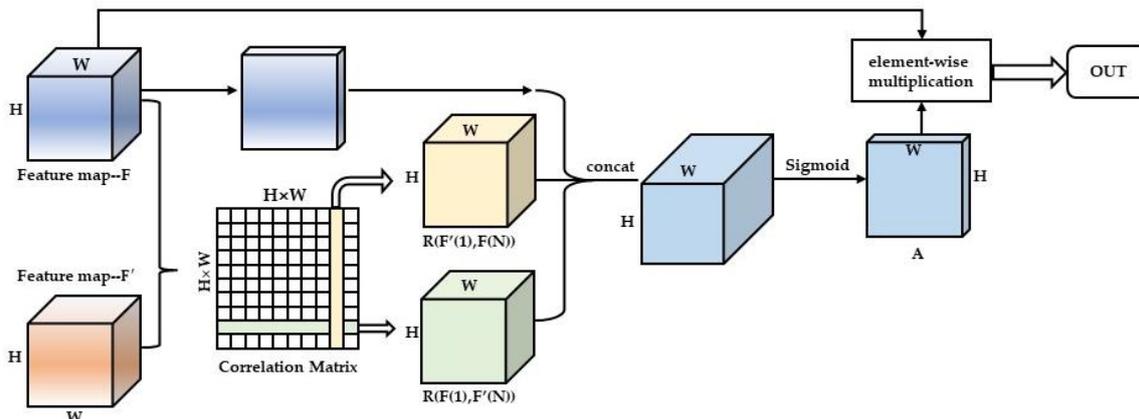


Figure 2. Structure diagram of contour information extraction module. Dimension raising and dimension reducing in the figure are omitted.

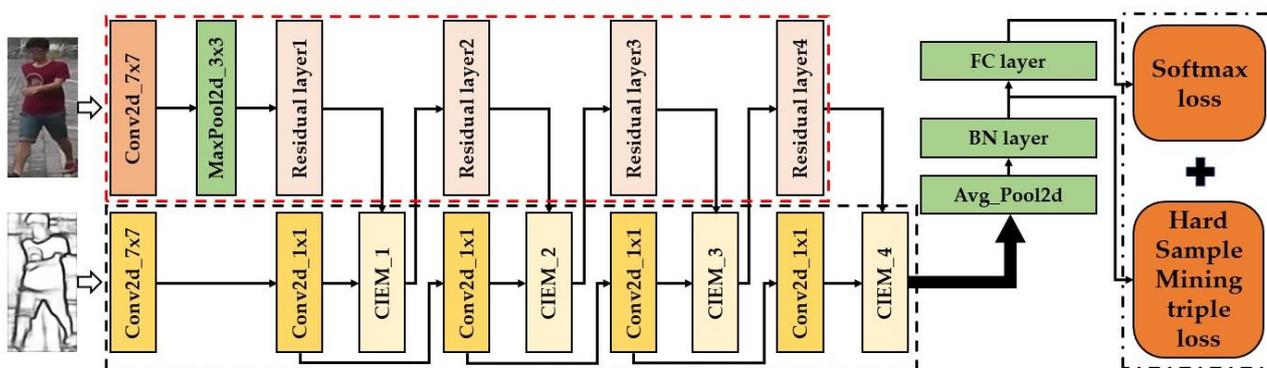


Figure 3. The architecture diagram of the method proposed in this paper.

3.1. Contour Information Extraction Module

The pedestrian contour contains relevant features that are beneficial to learning, but the general CNN network will ignore them to some extent. In order to enable the network to pay more attention to the information contained in the pedestrian contour, we adopt the idea of attention and propose the contour information extraction module. Next, we will introduce the details of the contour information extraction module and how to use it, as shown in Figure 2.

For the pedestrian image in the dataset, when using ResNet-50 for convolution, its output feature map behind a residual layer is set as F , and its size is $C \times H \times W$, where C is the number of channels and $H \times W$ is the space size. For the pedestrian contour branch, we reduce the dimension of the pedestrian contour map corresponding to the pedestrian image through convolution layer to obtain the contour feature map F' with the same size as the feature map F , and its size is $1 \times H \times W$, where 1 is the channel number, $H \times W$ is the space size, and F and F' are the inputs of the contour information extraction module.

Step 1, reduce the channel number of F to $C/8$, and increase the channel number of F' to $C/8$ to reduce the network computation.

Step 2, each feature point on the reduced dimension feature graph F is sorted into a one-dimensional vector with the length of $N = H \times W$. For convenience, we assign labels from 1 to N to each feature point, recorded as $f_i \in \mathbb{R}^{\frac{C}{8} \times N}$, where $i = 1, \dots, N$. Similarly, the characteristic points obtained from F' are recorded as $f_j \in \mathbb{R}^{\frac{C}{8} \times N}$, where $j = 1, \dots, N$.

Step 3, we calculate the correlation matrix R_{ij} between feature map F and contour feature map F' , the calculation formula is as follows:

$$R_{ij} = (f_i)^T (f_j) \quad (1)$$

where $R_{ij} \in \mathbb{R}^{N \times N}$.

Step 4, split the correlation matrix obtained in the previous step. The dimension of the correlation matrix is $N \times N$, so the transverse feature vector in the matrix represents the correlation between a feature point in F and the contour feature F' , which we record as $R(F(1), F'(N)) \in \mathbb{R}^{N \times H \times W}$. Similarly, the longitudinal feature vector in the matrix represents the correlation between a contour feature point in F' and feature F , which we record as $R(F'(1), F(N)) \in \mathbb{R}^{N \times H \times W}$.

Step 5, we concatenate $R(F(1), F'(N))$ and $R(F'(1), F(N))$ together to obtain the contour information diagram relative to the feature map F . We concatenate this relationship graph with the reduced dimension feature graph F , and then reduce the dimension of the concatenated channel number to 1. After the sigmoid function, we obtain the contour correlation weight matrix $A \in \mathbb{R}^{1 \times H \times W}$.

Step 6, the output of contour information extraction module is as follows:

$$F_{out} = A * F \quad (2)$$

where F is the feature graph before dimension reduction in step 1, “*” represents element-wise multiplication, that is, multiplication of elements at corresponding positions in tensor. $F_{out} \in \mathbb{R}^{C \times H \times W}$.

3.2. Overall Architecture

In the previous section, we introduced the contour information extraction module in detail. Next, we will introduce the Re-ID method based on contour information embedding as a whole. The overall framework we proposed is shown in Figure 3.

The whole frame is divided into two branches, with two inputs respectively. The input of the main branch is the RGB pedestrian original image in the dataset, and the input of the contour branch is the pedestrian contour image corresponding to the RGB original image.

In the main branch, we use ResNet-50 as the backbone network, which is generally divided into five parts, namely, a convolution layer conv1 and four residual layers.

The contour branch is mainly composed of five convolution layers and four contour information extraction modules. A convolution layer with the same conv1 parameter as the convolution layer is placed at the front of the contour branch to perform preliminary dimensionality reduction for the contour map. We add a contour information extraction module at the output position of each residual layer of ResNet-50, and place a convolution layer before the contour information extraction module. The input dimension of the convolution layer is 1, the output dimension is 1, the kernel_size = 1, and the step size is 2, which is used to reduce the dimension of the contour feature map.

The training method used in our experiment is the same as that used in most Re-ID studies. The loss function used in the experiment is softmax loss and the hard sample mining (trihard loss) [28], which we express as L_{ID} and L_T , respectively. After backbone receives the characteristic map, it receives the characteristic vector through an average pooling layer. After the vector passes through the BN layer, it calculates the loss function L_T , and its calculation formula is:

$$L_T = \frac{1}{P \times K} \sum_{a \in batch} (\max_{p \in A} d_{a,p} - \min_{n \in B} d_{a,n} + \alpha) \quad (3)$$

where P represents the number of person ID in a batch, randomly selecting K pictures for each person ID. a represents the anchor point, p represents the positive sample, and n represents the negative sample. A is the set of positive samples and B is the set of negative samples. $\max d_{a,p}$ represents the most difficult positive sample, $\max d_{a,n}$ represents the most difficult negative sample. α means margin and is set to 0.3.

The loss function L_{ID} is calculated by the eigenvector obtained from the BN layer output through the linear layer, the calculation formula is:

$$L_{ID} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i}^T f_i}}{\sum_{k=1}^C e^{W_k^T f_i}} \quad (4)$$

where f_i is the i -th feature, W_k corresponds to a weight vector for class k , y_i is the corresponding class label, C is the number of classes in training dataset, and the size of the mini-batch in the training process is N .

Therefore, the final loss function of this method is:

$$L = L_{ID} + L_T \quad (5)$$

4. Experiments

In this chapter, we will prove the effectiveness of the proposed method from the experimental results. Therefore, we have designed a series of ablation experiments. The proposed model will be tested on Market1501 and DukeMTMC-reID datasets to verify the universality of the method in this paper. Comparing our method with the advanced methods in Re-ID research in recent years, our method still has certain competitiveness.

4.1. Datasets and Implementation Details

We selected two datasets that are most commonly used in Re-ID research for experiments. The Market1501 dataset [29] contains 1501 different pedestrian IDs, 751 pedestrian IDs in the training set and 750 pedestrian IDs in the test set, with a total of 32,217 images. The DukeMTMC-reID dataset [30] contains 1812 different pedestrian IDs, 702 pedestrian IDs in the training set and 1110 pedestrian IDs in the test set. Among them, we mainly refer to the ablation experiment with Market1501 and DukeMTMC-reID datasets. In the experiment, we use the RCF model to extract the contour of the dataset and build the pedestrian contour dataset.

Before network training, we will perform data enhancement operations on RGB original images and pedestrian contour images, including random clipping, horizontal flipping, and other common image enhancement operations, and in order to unify different datasets, we adjust the input images to 256×128 pixels. In the training process, we use the Adam optimizer to set the learning rate to 8×10^{-4} , the weight decay rate to 5×10^{-4} , the training cycle to 600, and the batch size to 32. After the training, we did not use methods such as re-ranking to optimize the sorting. In the testing phase, we used the cumulative matching characteristics (CMC) [31] of Rank 1 and mean average precision (mAP) [29] to evaluate the performance, like most of the research on Re-ID.

4.2. Ablation Study

4.2.1. Necessity of Contour Information

In two different baseline models, we directly use the pedestrian contour map as the feature map to embed it in the middle layer of the network, and name it the contour embedding method (CEM). These two different baseline models are named baseline1 and baseline2. Baseline1 is a weak baseline model, its backbone network is ResNet-50, and the dimension reduction operation of the last residual layer is reserved. Use the pre-trained parameters on ImageNet before network training. Baseline2 is a strong baseline model. Its backbone network is ResNet-50, which removes the dimension reduction operation of the last residual layer. Before training, the network uses pre-trained parameters that are more suitable for Re-ID research. The specific operation is to embed the pedestrian contour

map into the output position of the four residual layers of ResNet-50 in the way of element level addition after dimension reduction through the convolution layer, so as to verify that the CNN network ignores the contour information in the process of extracting image features. Table 1 shows the experimental data of our weaker baseline1, baseline1-CEM, more powerful baseline2, and baseline2-CEM, we can get the following observations:

1. In our research on Re-ID, the CNN network lacks in the extraction of contour features and the expression of contour information. This can be seen from the comparison between baseline1 and baseline1+CEM. Although the final recognition rate is not very high, the effect of adding the contour map is clear. For example, on the Market1501 dataset, baseline1 added with CEM is 0.8% higher on the mAP and 1.3% higher on the Rank-1 than the original;
2. For the powerful baseline 2, perhaps due to the optimization of network pre-trained parameters, the CEM method cannot improve the final recognition rate of the network, and the method of directly using the contour map cannot effectively make the CNN network pay attention to more contour information, it will even reduce the original recognition performance. In order to make the network pay attention to the contour information on the strong baseline, the contour information extraction module is proposed.

Table 1. The model performance (%) after adding the contour map as a feature supplement on two different baseline models.

Methods	Market1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
Baseline1	74.6	88.5	61.0	76.9
Baseline1-CEM	75.4	89.8	64.4	79.9
Baseline2	82.1	93.8	72.0	84.6
Baseline2-CEM	81.8	93.4	71.7	84.2

4.2.2. Select the Position to Add CIEM

From the characteristics of the existing convolutional neural network, the edge, contour and other feature information contained in an image are all shallow feature expressions, and the visualization results of the feature maps of each layer of CNN network are also the same. Therefore, the following experiments are required to verify where to put the contour information extraction module and how to use it. We will use the contour information extraction module at the output positions of the four residual layers of ResNet-50, and name these four positions L1-L4. The experimental results are shown in Table 2, and we can receive the following observations.

The method of using contour information extraction module is quite different from the speculation before the experiment. From the experimental data, when we only add the contour information extraction module after the first three residual layers, the improvement of the experimental results is not obvious. Taking the Market1501 dataset as an example, the final rank 1 of the contour information extraction module used at L1 and L2 output locations is 94.3%, and the final rank 1 of the contour information extraction module used at L1, L2, and L3 output locations is 94.5%; the recognition effect of these two methods is better than that of the baseline model, but the improvement is not large, and the results are similar. When the contour information extraction module is used after the four residual layers, the final recognition effect of the network is 83.8% on rank1 and 95.1% on mAP, which is significantly improved compared with the former two, and also exceeds the baseline model we use. Therefore, we obtain the final model architecture of this article.

Table 2. Model performance (%) after adding contour information extraction module at the output position of different residual layers.

Methods	Market1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
L1+CIEM	81.7	93.9	72.2	85.2
L1,L2+CIEM	82.0	94.3	72.6	86.0
L1,L2,L3+CIEM	82.4	94.5	73.4	86.7
L1,L2,L3,L4+CIEM	83.8	95.1	73.5	86.8

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

We compare our method with the more classical methods in Re-ID research and some more advanced methods proposed in recent years. Table 3 shows the performance of these methods on two commonly used datasets.

Table 3. Performance (%) comparisons with the state-of-the-art on Market1501 and DukeMTMC-reID.

Methods	Market1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
Baseline	82.1	93.8	72.0	84.6
SVDNet [32]	62.1	82.3	56.8	76.7
DaRe [33]	74.2	88.5	63.0	79.1
HA-CNN [34]	75.7	91.2	63.8	80.5
DuATM [35]	76.6	91.4	64.6	81.8
Part-Aligned [36]	79.6	91.7	69.3	84.4
PCB [8]	77.4	92.3	66.1	81.8
PCB+RPP [8]	81.6	93.8	69.2	83.3
BoT [37]	85.9	94.5	76.4	86.4
DG-Net [38]	86.0	94.8	74.8	86.6
HPM [39]	82.7	94.2	74.3	86.6
CamStyle(RK) [40]	71.5	89.5	57.6	78.3
DaRe(RK) [33]	82.0	88.3	74.5	80.4
Ours	83.8	95.1	73.5	86.8

First of all, compared with the baseline model, our method can still improve the final recognition effect on a very powerful baseline model; taking Market1501 dataset as an example, our method is 1.7% higher than the baseline model in mAP and 1.3% higher than the baseline model in Rank-1. Moreover, compared with classical Re-ID algorithms, such as SVDNet, PCB, etc., our method has shown a strong competitive advantage, and has absolute advantages in mAP and Rank-1, two commonly used indicators. Compared with some more advanced methods proposed in recent years, our method has its own advantages to some extent. In the relevant experiments on dataset Market1501, our method has a certain advantage in the indicator Rank-1, for example, compared with BoT, DGNet and other methods, our method still has an advantage of about 0.5%. In terms of the indicator mAP, our method is still comparable to most, but there is still a certain gap compared with DG-Net. On dataset DukeMTMC-reID, the results presented by this method are still the same. Our method has certain advantages over other methods in terms of evaluation index Rank-1, but it is not satisfactory in terms of mAP, which is also the direction of our next research and improvement. For the rest, our method does not use the reranking technology,

but compared with other methods using the reranking technology, such as cam and dare, our method still has certain advantages.

5. Conclusions

In this paper, we propose a Re-ID method based on contour information embedding. With the idea of attention mechanism and the relationship between CNN feature map and contour map, the contour information extraction module is constructed. We use ResNet-50, which is the most commonly used in Re-ID research, as the backbone network, and use the contour information extraction module in its residual layer output position, so that the network can pay more attention to the contour information in the process of feature extraction. Our method is a breakthrough which attempts to use contour information, and it can still achieve very good recognition effect on a very powerful baseline network. The use of contour information is not limited to this, and we hope to have more research on pedestrian contour in the field of pedestrian recognition.

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

Funding: This work is supported by the National Natural Science Foundation of China (No.61631009, No.61771220), and the National Key R&D Program of China (No.2017YFB1002900, No.2017YFB0404800).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Ye, M.; Shen, J.; Lin, G.; Xiang, T.; Hoi, S. Deep Learning for Person Re-identification: A Survey and Outlook. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 2872–2893. [[CrossRef](#)] [[PubMed](#)]
2. Zhang, L.; Li, W.; Yu, L.; Sun, L.; Ning, X. GmFace: An explicit function for face image representation. *Displays* **2021**, *68*, 102022. [[CrossRef](#)]
3. Niu, C.; Nan, F.; Wang, X. A super resolution frontal face generation model based on 3DDFA and CBAM. *Displays* **2021**, *69*, 102043. [[CrossRef](#)]
4. Ruan, L.H.; Han, Y.X.; Sun, J.R.; Chen, Q.C.; Li, J.Q. Facial expression recognition in facial occlusion scenarios: A path selection multi-network. *Displays* **2022**, *74*, 102245. [[CrossRef](#)]
5. Chen, K.W.; Lai, C.C.; Lee, P.J.; Chen, C.S.; Huang, Y.P. Adaptive Learning for Target Tracking and True Linking Discovering Across Multiple Non-Overlapping Cameras. *IEEE Trans. Multimed.* **2011**, *13*, 625–638. [[CrossRef](#)]
6. Khamis, S.; Kuo, C.H.; Singh, V.K.; Shet, V.D.; Davis, L.S. Joint Learning for Attribute-Consistent Person Re-Identification. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 134–146.
7. Yang, X.; Wang, M.; Tao, D. Person Re-Identification with Metric Learning Using Privileged Information. *IEEE Trans. Image Process.* **2018**, *27*, 791–805. [[CrossRef](#)] [[PubMed](#)]
8. Sun, Y.; Zheng, L.; Yang, Y.; Tian, Q.; Wang, S. Beyond Part Models: Person Retrieval with Refined Part Pooling. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 501–518.
9. Zhao, H.; Tian, M.; Sun, S.; Jing, S.; Tang, X. Spindle Net: Person Re-identification with Human Body Region Guided Feature Decomposition and Fusion. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 907–915.
10. Zhu, J.; Yang, H.; Wang, J.; Zhang, W. Description-based person search with multi-grained matching networks. *Displays* **2021**, *69*, 102039. [[CrossRef](#)]
11. Guo, N.; Di, K.; Liu, H.; Wang, Y.; Qiao, J. A metric-based meta-learning approach combined attention mechanism and ensemble learning for few-shot learning. *Disp. Technol. Appl.* **2021**, *70*, 102065. [[CrossRef](#)]
12. Geirhos, R.; Rubisch, P.; Michaelis, C.; Bethge, M.; Wichmann, F.A.; Brendel, W. ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. In Proceedings of the International Conference on Learning Representation, New Orleans, LA, USA, 6–9 May 2019.

13. Jiang, Z.; Yuan, Y.; Wang, Q. Contour-aware network for semantic segmentation via adaptive depth. *Neurocomputing* **2018**, *284*, 27–35. [[CrossRef](#)]
14. Wei, X.L.; Hu, B.Y.; Gao, T.S.; Wang, J.; Deng, B. Multi-scale convolutional neural network for texture recognition. *Displays* **2022**, *75*, 102324. [[CrossRef](#)]
15. Ru, H.A.; Lei, M.A.; Jh, B.; Xc, C. T-GAN: A deep learning framework for prediction of temporal complex networks with adaptive graph convolution and attention mechanism. *Displays* **2021**, *68*, 102023.
16. Chen, Q.; Fan, J.; Chen, W. An improved image enhancement framework based on multiple attention mechanism. *Displays* **2021**, *70*, 102091. [[CrossRef](#)]
17. Xu, R.; Kang, X.; Li, C.; Chen, H.; Ming, A. DCT-FANet: DCT based frequency attention network for single image super-resolution. *Displays* **2022**, *74*, 102220. [[CrossRef](#)]
18. Wang, J.; Li, Y.H.; Pan, Y.W.; Yao, T.; Tang, J.H.; Mei, T. Contextual and selective attention networks for image captioning. *Displays* **2022**, *65*, 222103. [[CrossRef](#)]
19. 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](#)]
20. Zhao, L.; Xi, L.; Zhuang, Y.; Wang, J. Deeply-Learned Part-Aligned Representations for Person Re-identification. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3239–3248.
21. Wang, C.; Zhang, Q.; Huang, C.; Liu, W.; Wang, X. Mancs: A Multi-task Attentional Network with Curriculum Sampling for Person Re-Identification. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 384–400.
22. Sun, L.; Liu, J.; Zhu, Y.; Jiang, Z. Local to Global with Multi-Scale Attention Network for Person Re-Identification. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, China, 22–25 September 2019; pp. 2254–2258.
23. Zhang, Z.; Lan, C.; Zeng, W.; Jin, X.; Chen, Z. Relation-Aware Global Attention for Person Re-Identification. In Proceedings of the Computer Vision and Pattern Recognition IEEE, Seattle, WA, USA, 13–19 June 2020; pp. 3183–3192.
24. Chen, J.; Yang, Q.; Meng, J.; Zheng, W.S.; Lai, J.H. Contour-Guided Person Re-identification. In Proceedings of the Pattern Recognition and Computer Vision (PRCV), Xi'an, China, 8–11 November 2019; pp. 296–307.
25. Yang, Q.; Wu, A.; Zheng, W.S. Person Re-identification by Contour Sketch under Moderate Clothing Change. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 2029–2046. [[CrossRef](#)] [[PubMed](#)]
26. Chen, K.; Shi, L.; Pan, Z.; Wang, J.; Zhan, X. Clothes-Independent Identity Feature Learning for Long-Term Person Re-identification. In Proceedings of the International Conference on Image and Graphics, Haikou, China, 6–8 August 2021; pp. 397–409.
27. Chen, J.; Zheng, W.S.; Yang, Q.; Meng, J.; Hong, R.; Tian, Q. Deep Shape-aware Person Re-identification for Overcoming Moderate Clothing Changes. *IEEE Trans. Multimed.* **2021**, *24*, 4285–4300. [[CrossRef](#)]
28. Hermans, A.; Beyer, L.; Leibe, B. Defense of the Triplet Loss for Person Re-Identification. *arXiv* **2017**, arXiv:1703.07737.
29. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable Person Re-identification: A Benchmark. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1116–1124.
30. Ristani, E.; Solera, F.; Zou, R.S.; Cucchiara, R.; Tomasi, C. Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. In Proceedings of the European Conference on Computer Vision Springer, Amsterdam, The Netherlands, 11–14 October 2016; pp. 17–35.
31. Bolle, R.M.; Connell, J.H.; Pankanti, S.; Ratha, N.K.; Senior, A.W. The relation between the ROC curve and the CMC. In Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies IEEE, Buffalo, NY, USA, 17–18 October 2005; pp. 15–20.
32. Sun, Y.; Zheng, L.; Deng, W.; Wang, S. SVDNet for pedestrian retrieval. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3800–3808.
33. Wang, Y.; Wang, L.; You, Y.; Zou, X.; Chen, V.; Li, S.; Huang, G.; Hariharan, B.; Weinberger, K.Q. Resource aware person re-identification across multiple resolutions. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8042–8051.
34. Li, W.; Zhu, X.; Gong, S. Harmonious attention network for person re-identification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2285–2294.
35. Si, J.; Zhang, H.; Li, C.G.; Kuen, J.; Kong, X.; Kot, A.C.; Wang, G. Dual Attention Matching Network for Context-Aware Feature Sequence based Person Re-Identification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5363–5372.
36. Suh, Y.; Wang, J.; Tang, S.; Mei, T.; Lee, K.M. Part-aligned bilinear representations for person re-identification. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 402–419.
37. Hao, L. Bags of Tricks and A Strong Baseline for Deep Person Re-identification. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019; pp. 1487–1495.
38. 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. 2133–2142.

39. Fu, Y.; Wei, Y.; Zhou, Y.; Shi, H.; Huang, G.; Wang, X.; Yao, Z.; Huang, T. Horizontal pyramid matching for person re-identification. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 8295–8302.
40. Zhong, Z.; Liang, Z.; Zheng, Z.; Li, S.; Yi, Y. Camera style adaptation for person re-identification. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5157–5166.

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.