Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Advancements in Pedestrian Re-Identification

Ziqi Ren*

Faculty of Science, Shanghai University, Shanghai, China

* Corresponding Author Email: 22210100005@m.fudan.edu.cn

Abstract. The realm of pedestrian re-identification (Re-ID) technology, a vital component in areas such as video surveillance and intelligent security, has witnessed significant advancements, particularly in the context of deep learning. This domain, despite its nascent start in the broader image processing landscape, has seen rapid evolution with the advent of robust deep learning techniques. The present article delves into a range of methodologies pertinent to pedestrian Re-ID, emphasizing both the integral loss functions in metric learning and the necessary datasets for effective Re-ID implementation. Additionally, it explores the diverse applications of pedestrian Re-ID technology, shedding light on its multifaceted utility. The discourse extends to an analysis of the impending research challenges and potential directions in this field, encapsulating the dynamic and evolving nature of pedestrian re-identification technology. In conclusion, this article offers a comprehensive overview of the current state and forward-looking insights into pedestrian Re-ID, marking a pivotal contribution to the understanding of this transformative technology.

Keywords: Pedestrianv Re-ID technology; Deep learning; Global and Local Features; attention mechanism; Pooling; Loss function.

1. Introduction

Pedestrian re-identification (Re-ID) technology, leveraging computer vision and artificial intelligence, aims to match and recognize multiple pedestrian images captured across various cameras or locations and times. In this context, the set of images for comparison is termed a "query," each constituting a "probe," while the set of candidate images is referred to as a "gallery."

$$I_g = argmax_{I_{i,j \in P_i}} s(I_q, I_{i,j}) \tag{1}$$

Among them, $s(I_1, I_2)$ represents the evaluation index for measuring the similarity between two images I_1andI_2.It should be noted that similarity measurement is very important in pedestrian re identification tasks. Different similarity measurement methods can be used in different scenarios to achieve better recognition results [1]. Compared to facial recognition, Re-identification (Re-ID) leverages the complete body features of individuals depicted in images, facilitating pedestrian comparison or retrieval in situations where identity confirmation based solely on facial information is insufficient. The significance of re-identification technology in surveillance systems has prompted extensive research and garnered widespread attention. In real-world pedestrian recognition scenarios, variations in camera angles, environmental factors, lighting conditions, and pedestrian poses result in substantial appearance discrepancies among individuals captured by monitoring systems, presenting formidable challenges to pedestrian re-identification. As a result, Re-ID has emerged as a crucial alternative technology for acquiring high-quality facial images. As shown in Figure 1.

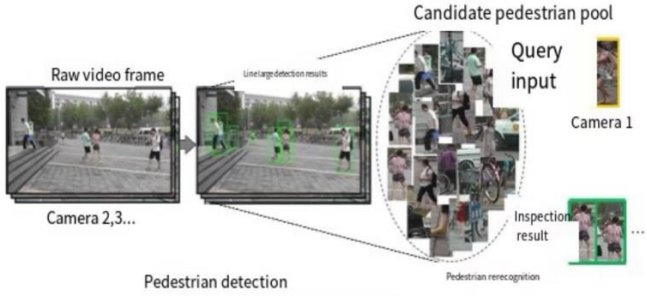

Fig. 1 Pedestrian re-recognition system (Photo/Picture credit: Original).

In the early days, the main task of pedestrian re identification was to manually extract more reasonable visual features and construct more reasonable similarity measures. However, when faced with issues such as changes in pedestrian posture, changes in image lighting angles, and camera performance, traditional methods appear to be more challenging. So deep learning techniques based on a large number of samples that can automatically learn problem features have begun to emerge and gradually become mainstream. Generally, we use deep learning to extract the feature vectors of images, and then use metric learning to distinguish the feature vectors. After that, it quantifys the differences between images and use massive annotated data to train and iterate the model, making it more applicable [2]. Then, the required loss functions and the datasets used in deep learning in recent years were introduced. Finally, this article summarizes and organizes the existing problems in the field of pedestrian re identification, and looks forward to future research directions.

2. Relevant theories

2.1. Re-ID based on representation learning

Representation learning can yield more discriminative and robust pedestrian features, including pedestrian ID features. Alternatively, it can be framed as an image verification task, wherein a pair of pedestrian images are input, and a trained network model determines whether they belong to the same pedestrian identity. This approach enhances the accuracy and robustness of Re-identification (Re-ID) [3]. Specifically, attention mechanisms can be employed to compute the importance weights of various regions in each pedestrian image. Subsequently, these weights are used to weigh the pedestrian image, resulting in a more precise and robust representation of pedestrian features [4]. To address the challenge of pedestrian re-identification in misaligned images, a coordinated attention model has been proposed [5].

2.2. Re-ID based on global or local features

In the realm of feature representation learning, global features pertain to the vector representations extracted from each pedestrian image via a network model. During the initial phase of implementing deep learning methodologies for pedestrian re-identification tasks, these global features were predominantly utilized. Additionally, local features typically originate from a certain decomposition of global features, followed by independent training of these segmented local features. This process enables the model to assimilate more precise feature information. The primary methods of decomposition include: analysis of the human form or estimation of human posture, horizontal segmentation of feature maps, and the division of feature map channels, among others [6]. As shown in Figure 2.

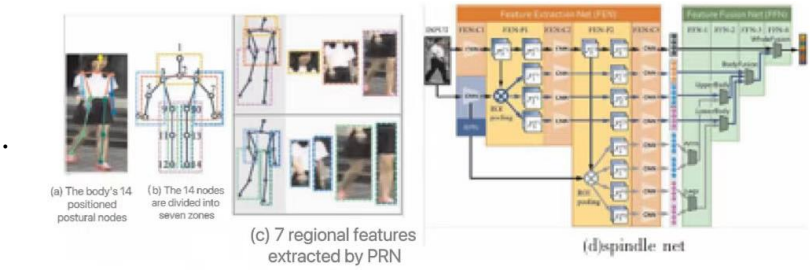

Fig. 2 Extract local features with pose points (Photo/Picture credit: Original).

In human body analysis or human pose estimation to generate local features, it is commonly used to combine global and local features to make the human pose highly matched.

2.3. Re-ID based on metric learning

Metric learning seeks to transform data from high-dimensional feature spaces into lower-dimensional metric spaces. This process commonly employs classification loss, comparison loss, and similar functions to facilitate effective embedding space learning. For instance, to enhance neighborhood similarity metrics, Luo et al. developed a lightweight reordering technique utilizing a clustering structure. Acknowledging the variance in queries, certain approaches have devised query-adaptive retrieval strategies, replacing a uniform search mechanism to bolster performance [7]. Zhou et al. introduced an innovative online local metric adaptation algorithm designed for creating specialized Markov metrics tailored to each probe during testing [8]. Notably, this method exclusively employs negative samples for metric adaptation, proving advantageous in real-world scenarios. Concurrently, the re-sorting approach can be seamlessly integrated into other high-accuracy pedestrian re-identification algorithms. This strategy not only enhances model efficacy but also represents a burgeoning research area with significant potential for future exploration.

2.4. Re-ID based on video sequences

While image-based pedestrian re-identification methods have demonstrated commendable results across numerous datasets, their limited generalization capacity is attributable to dataset size constraints. Specifically, single-frame image-based approaches struggle with generalizability [9]. Additionally, extracting information about pedestrian movements and posture changes from a single-frame image is challenging, leading to reduced model discrimination, especially in scenarios involving obstruction or significant lighting variations. To mitigate these limitations, researchers have introduced video-based pedestrian re-identification methods. These methods not only extract features from single frames but also consider inter-frame relationships [10]. Primarily, video-based pedestrian re-identification encompasses three components: extraction of single-frame image features, derivation of temporal features from video sequences, and metric learning [11]. Commonly, feature aggregation from image sequences employs pooling. However, this process can result in feature information loss. To combat this, some researchers have redefined the similarity of lengthy video sequences into similarities between multiple video segments. This approach involves filtering images within segments through collaborative attention embedding to discard redundant information, extract distinctive features, and ultimately compute the final similarity between two long video sequences based on the average similarity of the top-k video segments. Video datasets are rich in temporal information, but they also contain substantial noise. Effective noise management in image sequences, along with reducing model computational complexity, remains a pivotal aspect of this research domain.

3. Loss function

Choosing a suitable loss function is crucial for deep learning network models, especially for the training process and parameter learning process. As shown in Figure 3.

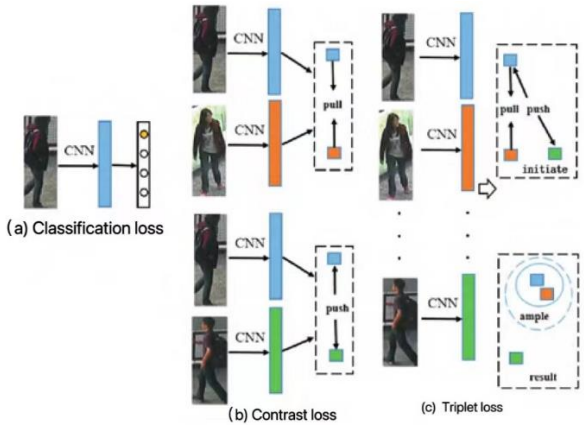

Fig. 3 Comparison of three loss functions (Photo/Picture credit: Original).

The ID of pedestrians is used as the training label for deep neural networks, treating each pedestrian as an independent category. Classification loss is widely used in deep metric learning of pedestrian re identification methods due to its advantages such as simple model training and hard sample mining. Assuing that the number of training samples in each batch is n, and given an input image $x_i$with label y_i, the ID loss is calculated through cross entropy [12].

$$L_{id}=-\frac{1}{n}\sum_{i=1}^{n} q(k)\log_{10}(p(y_i|x_i)) \tag{2}$$

Contarst loss: Comparative loss is mainly used in training twin networks, where the input is a pair of images$I_a$andI_b, which can be the same or different pedestrians. The basic idea of contrast loss is to take the feature vectors of two images as inputs and determine whether they come from the same pedestrian by calculating their distance or similarity. Its function can be expressed as:

$$L_c = yd(x_a - x_b)^2 + (1-y)[m - d(x_a - x_b)^2] \tag{3}$$

Where$x_a$and$x_b$are two images simultaneously inputted into the twin network [13].

Triple loss: In metric learning, triple loss is a commonly used loss function, and with the help of triple loss, many subsequent metric learning methods have been developed.

The commonly used expression of the triplet loss function is as follows: $L_{tri}=max(m+d_{ap}-d_{an},0)$, where m is a hyperparameter used to control the minimum interval between the distances between positive and negative samples [14].

Some scholars have designed a stronger triplet loss function by adding richer information. Hermans pointed out that keeping the network learning simple sample combinations will limit the network's generalization ability [15]. To address this problem, an improvement method for the triplet loss function is proposed.

In order to effectively improve the performance of the triplet loss function. However, the loss function generally does not limit the distance of samples, resulting in a long absolute distance between the positive sample and the target image.

4. Introduction to Datasets

Research methodologies in pedestrian re-identification have transitioned from conventional artificial feature-based approaches to deep learning techniques, a shift inseparable from the evolution of expansive datasets. Notably, the MSMT17 dataset, compiled by Peking University, stands out as a significant contribution in this domain. Comprising over 120,000 pedestrian images captured by 15

indoor and outdoor cameras, it is presently the most extensive single-frame Re-ID dataset available [16]. As shown in Figure 4.

Fig. 4 Example of MSMT17 dataset (Photo/Picture credit: Original).

The LvreID dataset, a collaborative effort between Peking University and the Microsoft Research Institute, encompasses sequence images of over 3,000 pedestrian IDs captured by 15 indoor and outdoor cameras, amounting to more than 3 million images across 14,943 sequences. This dataset's distinctiveness lies in its composition of multiple independent scenes, each capable of functioning as a standalone dataset. The division of training and testing sets is based on lighting conditions, aligning closely with real-world application scenarios.

In the realm of pedestrian re-identification models, the most prevalent evaluation metrics are Rank-n Accuracy, mAP (mean Average Precision), and the CMC (Cumulative Matching Characteristics) curve. Rank-n Accuracy measures the likelihood that an algorithm returns correct results among the top k responses in a given test set. This metric is operationalized by assessing whether the top result for a query matches the ID of the image being queried.

Mean Average Precision, denoted as mAP, is a vital evaluation index for multi-object detection and multi-label classification [17]. It involves plotting the relationship curve between precision and recall, followed by calculating the area under this curve as the mean precision. This process is repeated for all categories to ascertain their respective mean precision, culminating in the mean of these values to yield the mAP. This index not only reflects the accuracy of the model but also evaluates the ranking order it provides.

The CMC curve is an additional frequently used metric that reflects the accuracy of algorithms across different rankings. For each query image, the algorithm's returned results are ranked in descending order based on matching scores (or distance). The accuracy at various rankings is then computed from these ordered results. The CMC curve is plotted with ranking on the horizontal axis and accuracy on the vertical axis.

Table 1 compiles and summarizes the accuracy data of the most effective models on widely recognized datasets, along with other exemplary models..

Table 1. Performance optimal models and accuracy data for each dataset.

data set	SOTA	Map	Rank-l accuracy
	SCSN (2020)	84.0	86.8
	HLGAT (2021)	80.6	83.5
	RGA-SC (2020)	77.4	81.1
Market1501	HLGAT (2021)	94.3	97.5
	SGR (2021)	89.3	96.1
	B-VNet (2021)	89.2	96.0
Duke MTMC- ReID	HLGAT (2021)	87.3	92.7
	SGR (2021)	81.3	91.1
	B-VNet (2021)	80.6	90.5
MSMT17	SCSN (2020)	58.5	83.8
	RGA-SC (2020)	57.5	80.3
	CDNet (2021)	54.7	78.9
iLIDS-VID	CTL (2021)	89.7	
	BiCnet-TKs (2021)	75.1	84.6
	STMN (2021)	66.6	80.6
MARS	CTL (2021)	86.7	91.4
	BiCnet-TKS (2021)	86.0	90.2
	STMN (2021)	84.5	90.5
Duke MTMC-Video RelD	BiCnet-TKS (2021)	96.1	96.3
	STMN (2021)	95.9	97.0
	PSTA (2021)	97.4	98.3

5. Use of Re-ID

With societal progress and development, the public's demand for safety has heightened. Consequently, governments have incrementally augmented the installation of cameras in public and pivotal locations to bolster surveillance and crime prevention. These cameras amass real-time images and videos, offering vital data sources for a myriad of applications: public safety, intelligent transportation, commercial retail, medical sectors, and more. Pedestrian re-identification, employing advanced computer vision technology, automates and analyzes voluminous video and image data, leading to significant time and labor cost reductions.

This technology is critically important in personnel deployment, searches for missing individuals, and security incident investigations. In security systems, pedestrian re-identification aids police in identifying and tracking the movements of specific suspects in real-time. For instance, in traffic monitoring and signal control, pedestrian detection yields precise statistics on pedestrian counts and behavior, enabling the estimation of pedestrian flow to optimize traffic light timing. This optimization enhances traffic flow and safety.

In urban planning, pedestrian re-identification facilitates the analysis of pedestrian flow distribution, improving public facility layouts. Understanding people flow in various areas through this technology informs the planning of parks, commercial zones, transportation hubs, and more. Analysis of high foot traffic areas aids in establishing subways, shopping malls, amusement parks, and similar facilities for public use. Conversely, areas with lower pedestrian volumes can be considered for industrial developments, thereby minimizing disruptions to the majority's lifestyle.

6. Challenges

In the realm of pedestrian recognition technologies, prevalent models tend to be notably resource-intensive. Consequently, the formulation of an efficient, lightweight algorithm for robust pedestrian recognition emerges as a promising avenue for future advancements. This approach could significantly enhance processing speed without compromising accuracy.

Regarding dataset compilation, the current reliance on manually annotated datasets is both time-consuming and labor-intensive. A pivotal challenge lies in effectively bridging the domain gap inherent in virtual data, facilitating learning from such data while preserving the representational learning capacity of the models. This endeavor aims to augment the generalizability of the models, a critical factor in their practical applicability.

Furthermore, there is a noteworthy scarcity of research addressing the complexities of pedestrian re-identification in scenarios involving changes in attire. Existing methodologies, largely based on facial recognition, contextual body information, and coordinate transformation, exhibit limitations in real-world applications, often resulting in instability and errors. An exploration into alternative distinguishing features, such as gait and posture, could offer viable solutions to the challenges posed by clothing alterations in pedestrian re-identification. This uncharted research path holds the potential to significantly refine the accuracy and reliability of monitoring systems in dynamic environments.

7. Conclusion

While deep learning-based approaches to pedestrian re-identification have made significant strides, the field remains nascent with numerous unresolved challenges. This review delves into the existing methodologies for pedestrian re-identification, offering a detailed analysis aimed at imparting a thorough understanding of these techniques. Emphasis is placed on the intricate aspects of deep learning algorithms in this context, highlighting their strengths, limitations, and the complex interplay of factors that influence their efficacy. Additionally, the article explores emergent trends and potential avenues for future research, thereby contributing to the ongoing development in this domain. The objective is to equip researchers and practitioners with a nuanced perspective on deep learning applications in pedestrian re-identification, fostering further exploration and innovation in this vital area of study.

Reference

[1] Zhang Liang, "Research and System Design of Pedestrian Recognition Technology Based on Deep Learning" [D] Zhejiang University, April 10, 2023.

[2] Huang Yewen, Xu Zhicong, A Review of Pedestrian Recognition Based on Deep Learning, [D] Journal of Guangzhou University, Volume 21, Issue 2, June 20222.

[3] Research on Pedestrian Recognition Technology Based on Deep Learning [C], Wang Jingyu, Nanjing University of Posts and Telecommunications, October 20, 2023.

[4] Liu X, Zhao H, Tian M, et al.Hydraplus-net: Attentive deep features for pedestrian analysis [C]// Proceedings of the IEEE international conference on computer vision, 2017: 350-359.

[5] Li W, Zhu X, Gong S. Harmonious attention network for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 2285-2294.

[6] Research on Pedestrian Detection and Re identification Technology Based on Deep Learning[C] Gong Li, Beijing Jiaotong University, June 7, 2023.

[7] LUO C, CHEN Y, WANG N, et al. Spectral feature transformation for person re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 4976-4985.

[8] Research on Key Technologies for Pedestrian Recognition Based on Deep Learning [C] Fei Shengyu, Central South University, June 1, 2022.

[9] Kviatkovsky I, Adam A, Rivlin E. Color invariants for person re identification. IEEE Trans Pattern Anal Mach Intell, 2013, 35(7): 1622.

[10] Chen D P, Yuan Z J, Chen B D, et al. Similarity learning with spatial constraints for person re-identification // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016: 1268.

[11] ZHENG L, ZHANG H, SUN S, et al.Person re-identifi cation in the wild[C]//Proceedings of the IEEE Confer ence on Computer Vision and Pattern Recognition, 2017: 1367-1376.

[12] Chen Shijin: Research on Pedestrian Recognition Algorithms Based on Deep Learning [D] Nanchang University, May 25, 2023.

[13] HERMANS A, BEYER L, LEIBE B.In defense of the triplet loss for person re-identification[J]arXiv: 1703. 07737, 2017.

[14] Hermans A, Beyer L, Leibe B, et al. In defense of the triplet loss for person re-identification [J/OL]. Arxiv Perprint (2017-03-22) [2020-12-22]. https://arxiv.org/abs/1703.07737.

[15] Chen W H, Chen X T, Zhang J G, et al. Beyond triplet loss: A deep quadruplet network for person Re-identification // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 1320.

[16] Sun Y, Zheng L, Yang Y, et al.Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baselinse)[J].Springer, Cham, 2017.

[17] Yu S J, Li S H, Chen D P, et al. COCAS: A large-scale clothes changing person dataset for Reidentification // 2020IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 3397.