IFS/ACM Transactions on Machine Learning VOL. 1, NO. 1 (OCTOBER 2024) PP. 19-24

Multi-branch Collaboration Based Person Re-identification

Shoulin Yin
College of Information and Communication Engineering
Harbin Engineering University
Harbin, China

Asif Ali Laghari
School of Computer Science
Sindh Madressatul Islam University
Karachi, Pakistan

Editor: Edwin Engin Yaz

Abstract

As computer vision algorithms have advanced rapidly, the widespread deployment of video surveillance systems has enhanced traffic safety and propelled the growth of intelligent highway systems. However, the intricate nature of real-world scenarios, particularly the presence of occlusions, introduces noise that can cause the loss of critical feature information for identified individuals or objects. This poses significant challenges to current person re-identification algorithms. In response, this study introduces an innovative person re-identification approach that leverages a hybrid network architecture. The method performs feature extraction across four collaborative branches: a local branch, a global branch, a global contrast pooling branch, and an associative branch. This comprehensive approach yields a robust and diverse representation of person features, addressing the limitations posed by occlusion and noise in the environment. The neural network presented in this study is versatile and can be integrated with various backbone architectures. Empirical evaluations demonstrate that our proposed algorithm outperforms state-of-the-art techniques, while deep ablation studies substantiate the efficacy of the network's design. This suggests that the architecture contributes significantly to the performance gains observed.

Keywords: person re-identification, hybrid network, feature extraction, branch collaboration

1 Introduction

With the considerable enhancement in the standards of highway construction, video surveillance systems have been extensively deployed across various critical locations prone to safety incidents, such as stations, tunnels, bridges, interchanges, steep slopes, and service areas. These systems are instrumental in ensuring the safety of highway construction and operations. Notably, within the realm of highway construction and operational safety, video surveillance systems encompass a comprehensive process that includes person detection, tracking, and re-identification, which are pivotal for maintaining safety. In particular, person re-identification stands out as a complex issue in video surveillance, focusing on locating and recognizing individuals across different camera views. Over the past decade, numer-

ous algorithms have been developed by researchers, drawing on principles from pattern recognition and machine learning to address this challenge.

Person re-identification has garnered widespread interest from both academic researchers and industry practitioners, emerging as a focal point and a challenging area within the realms of artificial intelligence and computer vision in recent years. Person re-identification fundamentally employs computer vision techniques to ascertain the presence of a specific person within an image or video footage. The challenge of this technology is primarily centered around two aspects: feature extraction and the assessment of distance (or similarity). The complexity in feature extraction stems from the fact that the operating environment for video surveillance systems is typically uncontrolled, leading to feature representation being subject to a multitude of factors inherent to real-world scenarios. These include, but are not limited to, fluctuating lighting conditions, partial occlusions, varying camera angles, shifting postures of individuals, and alterations in attire.

Partial occlusion of individuals is a frequent occurrence in the construction and operational phases of highway systems. During construction, for instance, workers may be obscured by vehicles and machinery. Similarly, in operational settings, tollbooths, safety barriers, vegetation, or signage can obstruct views of people. Such occlusions result in a reduced visible area of the target person in imagery, causing a loss of detail and, compounded by noise from the occluding objects, can significantly impair the precision of person detection and re-identification tasks. The majority of current person re-identification algorithms are not equipped to handle occlusion effectively, making the resolution of partial occlusion a critical barrier to the practical application of these algorithms within highway surveillance systems.

Feature extraction serves as the cornerstone of person re-identification, with the caliber of feature representation profoundly influencing identification accuracy. Regardless of the specific feature extraction technique employed, its core objective is to encode the intrinsic characteristics present within images or video sequences into vector form, following defined protocols. Consequently, determining how to measure the similarity among features within a feature space–essentially, the distance between analogous feature vectors–is pivotal for accurately re-identifying individuals across different views. Conventional algorithms for person re-identification predominantly rely on manual feature extraction methods, which are typically categorized based on their focus on spatial, temporal, or a combination of spatio-temporal features.

Nevertheless, the advent of deep learning techniques, particularly the convolutional neural network (CNN), has marked a significant shift in the field of person re-identification. These methods have been swiftly adopted and expanded due to their robust generalization and feature representation capabilities. Their primary distinction from traditional, manually-driven feature extraction techniques is the approach of learning features directly from extensive datasets. Features are progressively extracted from raw pixel levels to more abstract levels through neural network architectures that mimic the human brain, thereby yielding deep features that are well-suited for identification tasks.

However, because of the inherent limitations of CNNs, including information loss due to pooling layers and convergence issues stemming from gradient descent, there has been a surge in research on person re-identification leveraging generative adversarial networks (GANs), following the introduction of GANs. Studies based on GANs or their enhanced

variants aim to capture the intrinsic properties of real data through the GAN framework, emulate the true data distribution, and enhance identification performance via the interplay of identification and generation. However, given the challenges associated with GANs, such as training instability and mode collapse, effectively addressing the impact of occlusions and similar factors on person re-identification remains a cutting-edge topic of investigation.

Over the past few years, algorithms based on the Transformer architecture have made their way from the domain of natural language processing into computer vision, with some researchers demonstrating their efficacy in image recognition tasks. While Transformer-based models are capable of addressing the issue of global information loss that is common in CNN models, they require substantial datasets for training and tend to have longer training periods. These models prioritize the capture of global features and often neglect the relationships and interactions with local features, which are essential for person re-identification, particularly in scenarios where occlusions and other scene factors are present.

Accordingly, this study introduces a person re-identification network that operates on the principle of multi-branch collaboration. The network performs feature extraction across four collaboration branches: a local branch, a global branch, a global contrast pooling branch, and an associative branch, thereby acquiring a robust and diverse capability for expressing person features. The architecture presented in this paper is versatile and can be integrated with various backbone networks. For the purposes of this research, the lightweight OSNet (Pan et al. (2023)) is utilized as the backbone network to construct new networks, which have achieved state-of-the-art results on several benchmark person re-identification datasets.

2 Model

The paper introduces a person re-identification network that operates through the collaboration of multiple branches. For an input image with dimensions $H \times W \times C$, where H represents the height, W is the width, and $C = 3$ indicates the color channels, the network is designed with a series of convolutional and transitional layers. Subsequently, it employs four collaborative branches to extract features, encompassing a local branch for capturing fine details, a global branch for holistic perceptions, a global contrast pooling branch for enhancing discriminative capabilities, and an associative branch for relational learning. These branches are sequentially engaged to acquire a comprehensive set of features that are both abundant and distinctive.

2.1 Collaboration of Multiple Branches

The local branch is dedicated to local feature extraction. Within this branch, the feature map is segmented into four equal horizontal strips, and local features are captured through the application of average pooling with a kernel size of $1 \times 1 \times C$. It's important to highlight that the amalgamation of these four local features into a column vector results in an ID prediction loss, with each local feature set being refined by the ID prediction loss. The concatenated features form the output:

$$f = [f_1^T, f_2^T, \cdots, f_4^T]^T. \tag{1}$$

where f_1, f_2, f_3 and f_4 are four column vectors.

Given data set $(x_i, y_i), i = 1, 2, \cdots, N$. ID predicted loss can be computed via:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log_2(\frac{exp((W^{y_i})^T f^i + b_{y_i})}{\sum_j exp((W^j)^T f^j + b_j)}). \tag{2}$$

where W^{y_i} and W^j are the y_i-th and j-th columns of the weight matrix W. N is the number of data. f^i and f^j are characterized by columns i and j. b_{y_i} and b_j are biases. Local branches provide effective and differentiated information for networks.

The global branch performs GeM pooling directly after convolution layers:

$$GeM(f_k = [f_0, f_1, \cdots, f_n]) = [\frac{1}{n}\sum_{i=1}^{n}(f_i)^{p_k}]^{\frac{1}{p_k}}. \tag{3}$$

where f_k is the GeM operator. If $p_k \to \infty$, GeM corresponds to maximum pooling, and if $p_k \to 1$, GeM corresponds to average pooling.

Through average pooling and maximum pooling, global contrast pooling (GCP) branch obtains local contrast features. Given the feature map input, GCP divides it into six horizontal grids, and outputs 256-dimensional feature vectors (Teng et al. (2023); Yin et al. (2023)).

Considering that local features only contain local information but do not the relationships between them, the associative branch associate image parts with the remaining parts, via computing six horizontal grid features like GCP.

2.2 Loss Function

Person re-identification task aims to minimizing the intra-class distance (maximizing the intra-class similarity) and maximizing the inter-class distance (minimizing the inter-class similarity). Based on the cosine similarity, the intra-class similarity needs to tend to 1 as much as possible, and the inter-class similarity needs to tend to 0 as much as possible. So the loss function of this paper is the binary cross-entropy loss between predicted and true labels.

3 Experiment

3.1 Experimental Setting

Market-1501 (Zheng et al. (2015)) and DukeMTMC-reID (Ma et al. (2018)) are used in the experiments. Market-1501 includes 32668 images of 1501 people from 6 cameras. DukeMTMC-reID includes 36411 images of 1404 people from 8 cameras. The mean Average Precision (mAP) which indicates the mean of the average accuracy of the model on all classes in the data set, is utilized as the evaluation criteria. The comparison methods includes HOReID (Wang et al. (2020)), PGFA (Miao et al. (2019)),TCSDO (Zhuo et al. (2019)), and Pose Transfer (Liu et al. (2018)).

3.2 Experimental Results

As shown in Table 1, the experimental results of the proposed model outperforms comparison methods on the two data sets, which demonstrates the superior of the proposed

Table 1: Experimental results (mAP) of different methods on the two datasets/%

Methods	mAP on Market-1501	mAP on DukeMTMC-reID
HOReID	85.0	75.7
PGFA	76.9	65.6
TCSDO	73.3	60.7
Pose Transfer	69.0	48.2
Proposed	**88.3**	**80.4**

multi-branch collaboration. Through the collaboration of features from four branches: a local branch, a global branch, a global contrast pooling branch, and an associative branch, the task of person re-identification is solved effectively.

4 Conclusion

In this study, we introduce a person re-identification network that operates on the principle of multi-branch collaboration, featuring a four-branch structure comprising a local branch, a global branch, a global contrast pooling branch, and an associative branch. This design enables the acquisition of more varied and higher resolution features. Experimental results validate the efficacy of this four-branch framework. The proposed network achieves state-of-the-art outcomes on two prominent person re-identification datasets. Currently, the enhancement of person re-identification performance through multi-branch cooperation is in need of theoretical validation, necessitating further verification and analysis.

Acknowledgments and Disclosure of Funding

The authors declare that there is no conflict of interest and no funding.

References

Jinxian Liu, Bingbing Ni, Yichao Yan, Peng Zhou, Shuo Cheng, and Jianguo Hu. Pose transferrable person re-identification. In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 4099–4108, Salt Lake City, UT, USA, 2018. Computer Vision Foundation / IEEE Computer Society.

Liqian Ma, Qianru Sun, Stamatios Georgoulis, Luc Van Gool, Bernt Schiele, and Mario Fritz. Disentangled person image generation. In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 99–108, Salt Lake City, UT, USA, 2018. Computer Vision Foundation / IEEE Computer Society.

Jiaxu Miao, Yu Wu, Ping Liu, Yuhang Ding, and Yi Yang. Pose-guided feature alignment for occluded person re-identification. In *2019 IEEE/CVF International Conference on Computer Vision*, pages 542–551, Seoul, Korea (South), 2019. IEEE.

Keyu Pan, Yishi Zhao, Tao Wang, and Shihong Yao. Msnet: A lightweight multi-scale deep learning network for pedestrian re-identification. *Signal, Image and Video Processing*, 17 (6):3091–3098, 2023.

Lin Teng, Yulong Qiao, Muhammad Shafiq, Gautam Srivastava, Abdul Rehman Javed, Thippa Reddy Gadekallu, and Shoulin Yin. Flpk-bisenet: Federated learning based on priori knowledge and bilateral segmentation network for image edge extraction. *IEEE Transactions on Network and Service Management*, 20(2):1529–1542, 2023.

Guan'an Wang, Shuo Yang, Huanyu Liu, Zhicheng Wang, Yang Yang, Shuliang Wang, Gang Yu, Erjin Zhou, and Jian Sun. High-order information matters: Learning relation and topology for occluded person re-identification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6448–6457, Seattle, WA, USA, 2020. Computer Vision Foundation / IEEE.

Shoulin Yin, Liguo Wang, Muhammad Shafiq, Lin Teng, Asif Ali Laghari, and Muhammad Faizan Khan. G2grad-camrl: An object detection and interpretation model based on gradient-weighted class activation mapping and reinforcement learning in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:3583–3598, 2023.

Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision*, pages 1116–1124, Santiago, Chile, 2015. IEEE Computer Society.

Jiaxuan Zhuo, Jianhuang Lai, and Peijia Chen. A novel teacher-student learning framework for occluded person re-identification. *arXiv preprint arXiv: 1907.03253*, 2019.