Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-16986-3

1236: EXPLAINABLE ARTIFICIAL INTELLIGENCE SOLUTIONS FOR IN-THE-WILD HUMAN BEHAVIOR ANALYSIS

Explainable graph-attention based person re-identification in outdoor conditions

Nayan Kumar Subhashis Behera[1] · Pankaj Kumar Sa[1] · Sambit Bakshi[1] · Umberto Bilotti[2]

Received: 21 May 2023 / Revised: 2 August 2023 / Accepted: 8 September 2023

Abstract
Person re-identification is the process of recognizing an individual across multiple camera views. It is essential for an extensive range of applications related to security and biometrics. We propose a shift in perspective for the ongoing re-identification studies. Present graph-based person re-identification methods need to explain the importance of graph attention and convolution techniques. However, our proposed method focuses on a less intrusive and explainable approach to attention selection and graph convolution methods. The proposed multi-channel framework utilizes visual features and attribute labels to represent each person uniquely. We applied large-scale benchmark datasets, such as MSMT17, DukeMTMC, CUHK03, and Market-1501.

Keywords Visual surveillance · Person re-identification · Explainable person re-identification · Graph-attention network · Graph convolutional network

Abbreviations

BN : Batch Normalization
CNN : Convolutional Neural Network
CMC : Cumulative Matching Characteristics
GPS : Graph Person Signature
GCN : Graph Convolution Network
GAT : Graph Attention
PGAN : Person Graph Attention Network
NLA : Non Local Attention

✉ Sambit Bakshi
sambitbaksi@gmail.com

Extended author information available on the last page of the article

1 Introduction

As the name refers, re-identifying individuals from non-overlapping cameras is person re-identification (PRId). It's an essential and ongoing research area in computer vision. PRId aims to identify the target probe again in surveillance systems with multiple cameras environment. In applications involving video surveillance, such as looking for missing individuals or suspects, then security and smart cities, and even it has a lot of potential. This area of computer vision has been getting more attention due to its principal uses, which are strongly tied to the safety and security of the public.

An average person's re-identification pipeline sends a target individual to the PRId system as a probe. It attempts to discover the matched ID recordings by searching through a gallery of known ones. Most of the PRId techniques currently use a framework for person classification that tries to identify the label of the probe image using a classifier trained on training data. However, some elements are making the work of PRId somewhat challenging. First, different data sources, including photos from security cameras or mobile devices, might impact image resolution. Second, changes in lighting and human positioning. Finally, poor detection/tracking is mostly brought on by occlusions and background clutter, which significantly alter how individuals look and make it more difficult to recognize them.

Traditional PRId models only concentrate on matching manually cropped video or image samples from various cameras. These techniques take into account distance metrics depending on particular properties. In the traditional model, it is necessary to identify or note the front pedestrian in the scene accurately. Otherwise, incorrect recognition or annotation will lead to false person recognition. To overcome such limitations, some contemporary methods [1, 3, 20, 39, 40] present the PRId setting in this field. Which mainly focuses on handling two tasks (person re-identification and pedestrian detection) simultaneously inside a single framework. But the problem is in the real world, and people are more prone to roam in groups. Other nearby individuals who appear in the same scene, even when they are going alone, also contain crucial context information. As a result, it might be challenging to tell between persons with identical attire, especially when have to look through a huge gallery set.

However, advanced deep learning led to a significant performance increase for supervised PRId [43]. With the introduction of Convolutional Neural Network (CNN) [21], researchers create several deep learning models to extract global characteristics from full pedestrian images and learn deep features to increase the effectiveness of PRId approaches. The majority of person re-ID techniques use CNN to extract distinguishing characteristics from pedestrian photos. Then they use metric learning or a classification setup to improve the deep model. But they don't provide the link between the pedestrian photos; instead, they merely use each image when learning features. Some part-based approaches [27, 32] is presented to extract local characteristics from various body sections to my local information. With regard to this, the direct partition approach is typically used to divide feature maps or pedestrian photos into many horizontal grids. Some researchers use methods of position estimation or human parsing technique [5, 9] to process pedestrian image data to discover precise and significant body components.

But after the introduction of Graph Convolutional Network (GCN) by Bruna et al. [2]. We gained the ability to understand the relationships between nodes in the network, which garnered great interest in the problem of Person re-identification.

Several recent research integrated GCN into the person re-ID pipeline. Based on the nature of connection learning, this research can be loosely split into two types. The first category of approaches uses classification loss to understand the link between various pedestrian photos and uses a single feature as the graph node. The methods in the second category interpret paired features as nodes in a graph and use verification losses to identify relationships between pairs of pedestrian photos to calculate robust node similarity. Since node and loss functions of both approaches differ, they could comprehend the association from many different aspects, encouraging us to discover more thorough connections between photos of pedestrians.

Many methods and approaches have been proposed including metric learning [6, 24], GAN-based [10, 23], attribute-based [12, 33], and spatial-temporal-based methods [36] in the solving and creating powerful model for person re-identification.

Our proposed method Person Graph Attention Network (PGAN), has the following key contributions.

- This experiment considers convolutional body parts and attributes as separate nodes to generate a unique topological structure to represent a Person's Graph (PG) using GCN.
- The proposed approach integrates an attention-based approach to evaluate the importance of different features by estimating the distance between the nodes.
- The proposed approach provides a detailed explanation of different non-intrusive attributes of a person, concentrating primarily on their attire and accessories.

2 Related works

Due to the fast development of dispersed and multi-camera systems, the difficulty of person re-identification has become necessary as a study topic. Traditional techniques make appropriate metric learning or attempt to learn powerful feature representations to increase the performance of PRId. This section includes a quick summary of some efforts on person re-identification techniques.

2.1 Attention-based PRId

In recent years, visual attention has been widely used to learn visual representations for many tasks, including object recognition, picture captioning, classification, and person ReID. Most existing methods for PRId generally concentrate on discriminative distance metric learning, including ranking by pairwise constraints, supervised learning of identity-discriminative information, and deep learning. These techniques assume that person photos are perfectly aligned, which is primarily false if there are inaccurate bounding boxes for shifting human positions. To overcome this limitation, Liu et al. [16] proposed an attention model that continuously and dynamically extracts distinct aspects from global whole-body images. The allocation of available resources can be skewed in favor of the input that contains the most information by using attention as a strategy. The attention model tries to mimic the way humans instinctively focus on what interests them.

Additionally, Huang et al. [7] use human parsing to extract precise information from several semantic regions of the body rather than considering the entire human body mask. The problem of extracting discriminative and robust features invariant to background clutter is solved by a method named MGCAM [31] proposed by Chunfeng et al.. Fan et al. [41] formulated a discriminative framework to extract robust human representation features simultaneously from whole and local body part images. Yu et al. [44] proposed a framework called AMMN to leverage the attention mechanism by emphasizing features other than the attention area.

2.2 Graph based PRId

With the emergence of deep learning, CNN were used to address PRId [21], and it has quickly dominated this sector. The high success of CNNs was due to their ability to handle data in Euclidean space. However, it failed to process complex connections and interdependency between objects in non-Euclidean space. Bruna et al. [2] proposed GCN for processing complex graph data and were inspired by CNNs. Spectral-based and non-spectral GCNs can be broadly divided into two categories based on the convolutional process. The spatial relations in the spatial-based GCN define the convolution operation on the graph. Where-else spectral-based GCN has a base of graph signal processing and spectral graph theory.

Shen et al. [30] proposed SGGNN model to learn similarity estimates by constructing a graph with various probe-gallery pairs and discovering pairwise relationships between them. Mazari et al. [22] formulated MLGCN model to address and resolve the issue of multi-label image recognition with label dependencies. Ye et al. [42] suggested an approach for PRId via dynamic graph matching to consider the anchor, negative, and positive data for constructing a graph structure. However, graph convolutions couldn't be used since this approach wasn't developed within a deep learning framework. Jiang et al. demonstrated a graph convolutional approach named PH-GCN [8] for PRId. The convolutional graph highlights the associations among various body parts. The PH-GCN method gives a broad solution that integrates local, global, and structural feature learning in a single end-to-end network. Nguyen et al. GPS [25] proposed to use the person attribute data and its corresponding body part to encode the visual person signature in a single, cohesive framework.

2.3 Graph-attention-based PRId

Graph Neural Networks (GNNs) [28] represents computer vision tasks into graph-based solutions. Bruna et al. [2] introduced Graph Convolution Networks (GCN) to aggregate computer vision tasks into the graph data. However, the GCN methods are based on graph convolutions or their approximations. Velickovic et al. [35] integrates masked self-attention layers to formulate Graph Attention Networks (GAT) and overcome the shortcoming of the GCN. The primary focus of GAT is to consider the most relevant parts, and it does not require any computationally intensive matrix operation or prior graph structure knowledge. The GAT stack layers in which nodes can attend over their neighborhood's feature, thus allowing for implicitly specifying various weights to different nodes in a neighborhood.

In this paper, we propose a graph attention based person re-identification method, which we call Person Graph Attention Network (PGAN). We evaluate our framework on four widely used datasets, including Market1501, DukeMTMC, CUHK03, and MSMT17. The results of the experiments show that our method is capable of significantly superior performance than the prior state-of-the-art.

3 Person-graph attention network

The proposed method Person-Graph Attention Network (PGAN), addresses the PRId task using crucial mutual characteristics that are undermined by traditional approaches. We describe the intake and outcome of PGAN and then concentrate on assembling graph attention.

3.1 Proposed architecture

Figure 1 describes the detailed architecture of the proposed Person Graph Attention Network (PGAN). Initially, a batch of images was provided as the input to the backbone CNN to extract the features. The extracted feature set is represented as $G = \{g_1, g_2, g_3, ..., g_N\}$, $g_i \in X^M$, where N is the length of the feature set, and M is the dimension of a single feature. Each feature set represents a node in the complete graph. Each edge depicts the similarity between the connected nodes. The PGAN's output signature is denoted by the symbol (G'). The proposed PGAN optimizes the constructed graph without losing any classification characteristics. Feature optimization is necessary for the graph-based approaches as more than thousands of images are present in the gallery sets and the probe sets of the benchmark datasets. Therefore it is practically impossible to represent the entire feature set as graphs.

PGAN is a multi-task and multi-branch PRId approach, which consists of four major phases. Firstly we extract the body features using the ResNet50 as the backbone model of the proposed architecture. Secondly, the lookup word embedding is used to extract the attributes of each person. The next phase is to construct complete graph called Person Graph (PG), from the extracted body features and the attribute information. The correlation matrix represents a complete graph based on the body parts and attribute nodes.

The Person Graph (PG) is constructed from the body features and attributes,as is represented as $G = (N, X)$, here $X = \{x_1, x_2, x_3, ...x_K\}$ is the set of nodes, K is the length of feature set and $E = (v_i, v_j)$ represents the connecting edges between nodes. A body feature or attribute is represented by a node in the full graph (G). Each edge is associated with weights represents the distance between two connected nodes. In this experiment, we use the Mahalanobis distance to compute the associated weight of the edge $e_i j$ between nodes n_i and n_j as represented in the (1).

$$e_{ij} = \sqrt{(n_i - \mu)\Sigma^{-1}(n_j - \mu)} \tag{1}$$

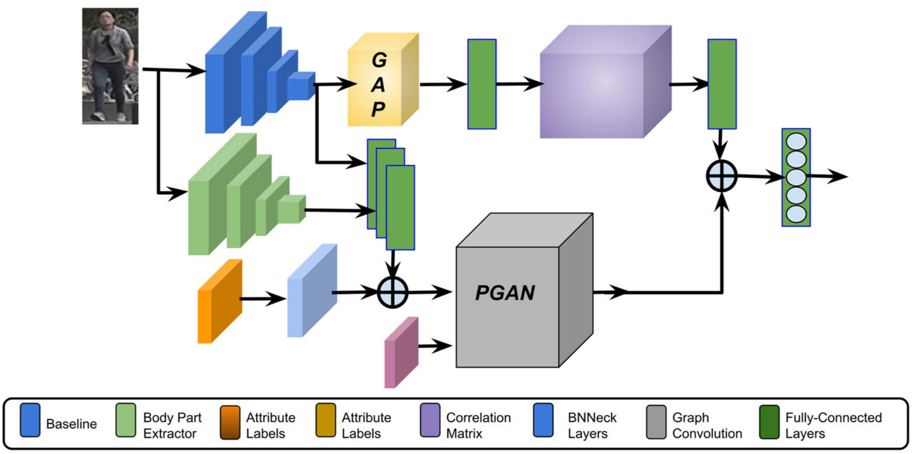

Fig. 1 Architecture of the proposed Person Graph Attention Network (PGAN)

3.2 Implementation

This experiment uses the ResNet-50 pre-trained on ImageNet as the backbone network. The proposed architecture customized the classifier module by removing the Batch-Normalization (BN) layer. Initially, each input image is padded 10 pixels with zero values, and then each input image is resized into 256×128. We use both horizontal and vertical flipping for data augmentation. Attribute labels are transformed into 284×300 word embedding. We incorporate Non-Local Attention (NLA) into each ResNet50 block to improve the discriminating ability of the backbone. Note that NA denotes the number of attributes. Market1501 [49] dataset has 30 attributes, and DukeMTMC-ReID [50] has 23 attributes.

4 Experiments

This section furnishes the particulars of the datasets, evaluation matrices, and the comparison of the results with the state-of-the-art methods.

4.1 Datasets

There are many datasets available for PRId in outdoor environment such as Market-1501 [49], DukeMTMC-reID [50], CUHK03 [14], MSMT17 [37], ETHZ Pedestrian [29], and i-LIDS 2008 [34]. In this paper, we consider four significant datasets such as Market-1501 [49] , CUHK03 [14], DukeMTMC-reID [50] and MSMT17 [37], as those datasets have a large number of objects and are considered large-scale PRId datasets as shown in Fig. 2. However,

Fig. 2 Sample pedestrian image from the used datasets such as Market-1501 [49], CUHK03 [14], and Duke MTMC-reID [50]

the network configuration of the proposed architecture is compatible with both the small and large benchmark datasets. The selected datasets have different identities, and there are no similar images. However, all those datasets are captured in outdoor environments.

Market1501 [37] This dataset contains 1,501 identities and 32,668 images in total. There are 19,732 images of 750 identities in the test set and 12,936 images of 751 in the train set.

DukeMTMC Re-id [50] The DukeMTMC re-ID dataset has eight different viewpoints. It has 1,404 identities and 34,183 images in total. The training and test set consists of 17,661 and 16,522 images, respectively. Both the test and the test sets have an equal number of identities.

CUHK03 [14] The CUHK03 dataset has eight viewpoints, with two cameras capturing each identity. It has 1,360 different identities and 13,164 images in total. This dataset presents major PRId challenges, including background clutter, occlusions, camera settings, image resolutions, blurring effects, viewpoints, poses, and complex lighting variations. Table 1 provides the attribute details of each PRId dataset used in the experiments.

MSMT17 [37] The MSMT17 is a multi-time and multi-scene PRId dataset. It has 126,441 images and 4,101 identities in total. This dataset is captured using 3 indoor and 12 outdoor cameras. It has a large number of identities extracted from 180 hours of videos.

4.2 Evaluation metrics

In this article, we have used two metrics such as mAP and rank-n to evaluate the proposed model.

4.2.1 Mean Average Precision (mAP)

The *mAP* is used to evaluate all the individuals and calculated as the *mean of average precision (mAP)* mentioned in (4). The *precision* accuracy is calculated for all the person images for

Table 1 Description of three datasets used

	Market-1501 [49]	DukeMTMC-reID [50]	MSMT17 [37]
Images	13164	36411	32668
Identities	1360	1812	1501
Camera	10	8	6
Distractors	0	17661	2793+500K
Label	DPM	Hand	Hand+DPM
Challenges	VV, OCC	VV, IV, DE, BC, OCC	IV, VV, CS, OCC
Frame Size		1920×1080	128×64
Frame Rate	Vary	60	Vary
Multishot	SS	MS+ SS	MS
Walking Conditions	NW	NW, FW	NW, Cyclicing
Modal	RGB	RGB	RGB
Scene	Outdoor	Outdoor	Outdoor

Table 2 Comparison results on Market-1501 [49], DukeMTMC-reID [50] and CUHK03 [14]

Year	Method name	Market-1501 [49]				DukeMTMC-Re-ID [50]				CUHK03 [14]			
		R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10	mAP
2018	MGCAM [31]	83.79	74.33	-	-	50.14	50.14	-	-	-	-	-	-
2021	GPS [25]	95.2	98.4	99.1	87.8	88.2	95.2	96.7	78.7	-	-	-	-
2019	PH-GCN [8]	98.5	79	93.5	97.4	94.8	70.7	85	92.7	88.1	61.5	64.9	81.8
2019	PH-GCN+reRanking [8]	97.8	91.1	94.3	97.1	95.9	85.7	88.8	93.8	87.8	73.3	70.6	81.9
2022	GS [15]	91.6	-	-	75.5	-	-	-	-	-	-	-	-
2021	HLGAT [47]	97.5	98.9	-	93.4	92.7	96.5	-	87.3	83.5	92.6	-	80.6
2021	HLGAT+reRanking [47]	98	99	-	97.5	94.7	96.7	-	94.4	90.6	94.4	-	89.9
2018	SGGNN [30]	92.3	96.1	97.4	82.8	81.1	88.4	91.2	68.2	95.3	99.1	99.6	94.3
2019	UGA [38]	87.2	-	-	70.3	75	-	-	53.3	56.5	-	-	68.2
2021	PH-GCN [8]	93.7	97.8	98.6	81.4	85.2	92.5	94.9	72.5	62	61.8	58.5	65.6
2021	PH-GCN+reranking [8]	94.6	96.9	97.7	92.1	89.3	93.9	96.1	86.8	74.1	66.5	70.3	71.7
2019	TBN [13]	93.2	-	-	83	85.5	-	-	73	72.3	-	-	68.5
2019	TBN+reranking [13]	95.4	-	-	91.3	89.2	-	-	85.1	79.5	-	-	80.7
2021	IPES-GCN [17]	96.2	-	-	94.3	90.4	-	-	88.3	-	-	-	-
2020	RRGCCAN [4]	89.9	-	-	75.2	86	-	-	77.7	-	-	-	-
2021	GCN [11]	88.65	95.75	-	74.15	80.38	90.84	-	64.32	-	-	-	-
2021	PGCN [48]	98	-	-	94.8	91.1	-	-	85.2	86.7	-	-	83.6
2021	PGCN+AGW [48]	98.2	-	-	95.1	91.6	-	-	85.7	86.9	-	-	83.9
2021	HRNet [19]	96.7	-	-	91.3	91.9	-	-	83.7	78.4	-	-	81.1
2022	PAGCN [45]	94.4	-	-	87.3	86.7	-	-	78	75.1	-	-	71.6
2022	AAGCN-MNA [26]	94.2	98.1	98.8	89.2	86.4	93.9	96	77.9	-	-	-	-
2022	AAGCN-NLM [26]	95.7	98.7	99.4	93.1	91.2	95.6	96.9	86	-	-	-	-
2022	GCR [46]	96	-	-	94.7	91.5	-	-	89.6	-	-	-	-
2021	PrGCN [18]	94.2	97.4	98.2	86.6	84.7	91.4	93.8	74.6	73.4	67.9	68.4	73.5
2023	**PGAN (proposed)**	**94.7**	**88.6**	**95.7**	**82.6**	**86.7**	78.6	84.5	87.6	85.4	76.5	74.3	76.5

a particular identity described in (2). AP is the average of the accuracy rates (precision) described in (3).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$AP = \frac{\sum Precision}{C_i} \tag{3}$$

$$mAP = \frac{\sum_{j=1}^{K} AP_i}{N} \tag{4}$$

True positive (TP) is the number of identities correctly predicted as positive. *False positive (FP)* is the number of identities incorrectly predicted as positive. C_i is the number of images in a particular identity i, and N is the total number of identities.

4.2.2 Cumulative Matching Characteristics (CMC)

The most common evaluation metrics for PRId methods are Cumulative Matching Characteristics (CMC) curves. The algorithm ranks all the galley images according to their distances in ascending order for query images. The CMC rank-n accuracy is defined in (5), and the final CMC is the average of all the accuracies.

$$Acc_k = \begin{cases} 1: & \text{if } \texttt{top}\ (k) \text{ ranked} \\ 0: & \text{otherwise} \end{cases} \tag{5}$$

4.3 Comparison to the state-of-the-art

The performance of the PGAN approach is compared to that of state-of-the-art methods. Table 2 details the performance comparison on CUHK03 [14], DukeMTMC-reID [50], and Market-1501 [49]. Table 3 details the performance comparison on MSMT17 [37].

Table 3 Comparison results on MSMT17 [37]

	MSMT17 [37]			
	R-1	R-5	R-10	mAP
GS (2022) [15]	79.2	-	-	50.9
HLGAT (2021) [47]	87.2	93	-	73.2
HLGAT+reRanking (2021) [47]	91.9	94.6	-	87.1
UGA (2019) [38]	49.5	-	-	21.7
IPES-GCN (2022) [17]	80.8	-	-	69.4
RRGCCAN (2020) [4]	74.8	-	-	53
PGCN (2021) [48]	87.7	-	-	72.7
PGCN+AGW (2021) [48]	88.1	-	-	73.2
HRNet (2021) [19]	82	-	-	60.6
PAGCN (2022) [45]	80	-	-	55.2
PGAN (proposed)	83.6	93.6	94.6	72.5

5 Conclusion

This article presents a systematic approach to recognizing a person using graph-convolution and graph-attention methods. It considers both body parts and attributes as different nodes to construct the graph. The effectiveness of PGAN is verified on four benchmark datasets. The fusion of attribute cues with body parts generates a plausible signature representing a person graph. PRId researchers can further integrate a co-matching method to utilize the abundant video information and reduce false matchings to use advanced supervised learning methods on video benchmarks.

Acknowledgements This research is partially supported by a project titled “Deep learning applications for computer vision task” funded by NITROAA with support of Lenovo P920 workstation and NVIDIA Corporation with support of NVIDIA Titan V and Quadro RTX 8000 GPUs.

Availability of data Data will be made available on legit request.

Declarations

Conflicts of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics Approval Not applicable.

Consent to participation All authors has given their consent to participate in this research.

Consent to publication All authors has given their consent to publish this research in this publication.

References

1. Batool E, Gillani S, Naz S, Bukhari M, Maqsood M, Yeo SS, Rho S (2023) Posnet: a hybrid deep learning model for efficient person re-identification. J Supercomput, pp 1–29
2. Bruna J, Zaremba W, Szlam A, LeCun Y (2013) Spectral Networks and Locally Connected Networks on Graphs. arXiv:1312.6203
3. Bukhari M, Yasmin S, Naz S, Maqsood M, Rew J, Rho S (2023) Language and vision based person re-identification for surveillance systems using deep learning with lip layers. Image Vis Comput, 132(104):658
4. Chen X, Zheng L, Zhao C, Wang Q, Li M (2020) RRGCCAN: Re-Ranking via Graph Convolution Channel Attention Network for Person Re-Identification. IEEE Access, 8:131,352–131,360 10.1109/access.2020.3009653
5. Guo J, Yuan Y, Huang L, Zhang C, Yao JG, Han K (2019) Beyond human parts: Dual part-aligned representations for person re-identification. In: Proceedings of the ICCV, pp 3642–3651. https://doi.org/10.1109/iccv.2019.00374
6. Hermans A, Beyer L, Leibe B (2017) In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737 https://doi.org/10.1109/cvprw50498.2020.00185
7. Huang H, Chen X, Huang K (2020) Human Parsing Based Alignment With Multi-Task Learning For Occluded Person Re-Identification. In: Proceedings of the ICME, pp 1–6. IEEE. https://doi.org/10.1109/icme46284.2020.9102789
8. Jiang B, Wang X, Zheng A, Tang J, Luo B (2021) PH-GCN: Person retrieval with part-based hierarchical graph convolutional network. IEEE Trans Multimed, 24:3218–3228. https://doi.org/10.1109/tmm.2021.3095789
9. Kalayeh MM, Basaran E, Gökmen M, Kamasak ME, Shah M (2018) Human Semantic Parsing for Person Re-identification. In: Proc IEEE conf comput vis pattern recognit, pp 1062–1071. https://doi.org/10.1109/cvpr.2018.00117

10. Karmakar A, Mishra D (2021) Pose Invariant Person Re-Identification using Robust Pose-transformation GAN. arXiv:2105.00930. https://doi.org/10.48550/arXiv.2105.00930
11. Kim G, Shu DW, Kwon J (2021) Robust person re-identification via graph convolution networks. Multimedia Tools Appl, 80(19):29,129–29,138. https://doi.org/10.1007/s11042-021-11127-0
12. Layne R, Hospedales TM, Gong S, Mary Q (2012) Person re-identification by attributes. In: Procedings of the BMVC, 2(8). https://doi.org/10.5244/c.26.24
13. Li H, Yang M, Lai Z, Zheng W, Yu Z (2019) Pedestrian re-Identification Based on Tree Branch Network with Local and Global Learning. In: Proceedings of the ICME, pp 694–699. https://doi.org/10.1109/icme.2019.00125
14. Li W, Zhao R, Xiao T, Wang X (2014) DeepReID: Deep filter pairing neural network for person re-identification. In: Proceedings of the CVPR, pp 152–159. https://doi.org/10.1109/cvpr.2014.27
15. Liao S, Shao L (2022) Graph sampling based deep metric learning for generalizable person re-identification. In: Proceedings of the CVPR, pp 7359–7368. https://doi.org/10.1109/cvpr52688.2022.00721
16. Liu C, Gong S, Loy CC, Lin X (2012) Person Re-identification: What Features Are Important? In: Proceedings of the ECCV, pp 391–401. Springer. https://doi.org/10.1007/978-3-642-33863-2_39
17. Liu C, Yang H, Zhu J, Zhou Q, Zheng S (2022) Modeling context appearance changes for person re-identification via IPES-GCN. Neurocomputing 483:210–220. https://doi.org/10.1016/j.neucom.2022.02.001
18. Liu H, Xiao Z, Fan B, Zeng H, Zhang Y, Jiang G (2021) PrGCN: Probability prediction with graph convolutional network for person re-identification. Neurocomputing 423:57–70. https://doi.org/10.1016/j.neucom.2020.10.019
19. Liu S, Huang W, Zhang Z (2021) Learning Hybrid Relationships for Person Re-identification. Proceedings of the AAAI 35:2172–2179. https://doi.org/10.1609/aaai.v35i3.16315
20. Maqsood M, Yasmin S, Gillani S, Bukhari M, Rho S, Yeo SS (2023) An efficient deep learning-assisted person re-identification solution for intelligent video surveillance in smart cities. Front Comput Sci, 17(4):174,329
21. Matsukawa T, Suzuki E (2016) Person re-identification using CNN features learned from combination of attributes. In: Proceedings of the ICPR, pp 2428–2433. IEEE. https://doi.org/10.1109/icpr.2016.7900000
22. Mazari A, Sahbi H (2019) MLGCN: Multi-Laplacian Graph Convolutional Networks for Human Action Recognition. In: Proceedings of the BMVC
23. Miao J, Wu Y, Liu P, Ding Y, Yang (2019) Pose-Guided Feature Alignment for Occluded Person Re-Identification. In: Proceedings of the ICCV, pp 542–551. https://doi.org/10.1109/iccv.2019.00063
24. Nguyen BX, Nguyen BD, Carneiro G, Tjiputra E, Tran QD, Do TT (2020) Deep metric learning meets deep clustering: An novel unsupervised approach for feature embedding. arXiv preprint arXiv:2009.04091
25. Nguyen BX, Nguyen BD, Do T, Tjiputra E, Tran QD, Nguyen A (2021) Graph-based Person Signature for Person Re-Identifications. In: Proceedings of the CVPR, pp 3492–3501. https://doi.org/10.1109/cvprw53098.2021.00388
26. Pan H, Bai Y, He Z, Zhang C (2022) AAGCN: Adjacency-aware Graph Convolutional Network for person re-identification. Knowl-Based Syst 236(107):300. https://doi.org/10.1016/j.knosys.2021.107300
27. Quan R, Dong X, Wu Y, Zhu L, Yang Y (2019) Auto-ReID: Searching for a Part-Aware ConvNet for Person Re-Identification. In: Proceedings of the ICCV, pp 3750–3759. https://doi.org/10.1109/iccv.2019.00385
28. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G (2008) The graph neural network model. Trans Neural Netw 20(1):61–80. https://doi.org/10.1109/TNN.2008.2005605
29. Schwartz WR, Davis LS (2009) Learning discriminative appearance-based models using partial least squares. In: Proceedings of the SIGCGIP, pp 322–329. https://doi.org/10.1109/sibgrapi.2009.42
30. Shen Y, Li H, Yi S, Chen D, Wang X (2018) Person Re-identification with Deep Similarity-Guided Graph Neural Network. In: Proceedings of the ECCV, pp 486–504. https://doi.org/10.1007/978-3-030-01267-0_30
31. Song C, Huang Y, Ouyang W, Wang L (2018) Mask-guided contrastive attention model for person re-identification. In: Proceedings of the CVPR
32. Sun Y, Zheng L, Li Y, Yang Y, Tian Q, Wang S (2019) Learning part-based convolutional features for person re-identification. Trans Pattern Anal & Mach Intell 43(3):902–917. https://doi.org/10.1109/TPAMI.2019.2938523
33. Taherkhani F, Dabouei A, Soleymani S, Dawson J, Nasrabadi NM (2021) Attribute guided sparse tensor-based model for person re-identification. arXiv:2108.04352
34. Team IL (2006) Imagery library for intelligent detection systems (i-LIDS). In: Proceedings of the IETCCS, pp 75–80. https://doi.org/10.1109/ccst.2006.313432

35. Veličković P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y (2017) Graph attention networks. arXiv:1710.10903
36. Wang G, Lai J, Huang P, Xie X (2019) Spatial-temporal person re-identification. In: Proc AAAI Conf Artif Intell, 33:8933–8940. https://doi.org/10.48550/arXiv.1812.03282
37. Wei L, Zhang S, Gao W, Tian Q (2018) Person transfer gan to bridge domain gap for person re-identification. In: Proceedings of the CVPR, pp 79–88. https://doi.org/10.1109/cvpr.2018.00016
38. Wu J, Yang Y, Liu H, Liao S, Lei Z, Li SZ (2019) Unsupervised Graph Association for Person Re-Identification. In: Proceedings of the ICCV, pp 8321–8330. https://doi.org/10.1109/iccv.2019.00841
39. Xiao T, Li S, Wang B, Lin L, Wang X (2017) Joint Detection and Identification Feature Learning for Person Search. In: Proceedings of the CVPR, pp 3415–3424. https://doi.org/10.1109/cvpr.2017.360
40. Xu Y, Ma B, Huang R, Lin L (2014) Person Search in a Scene by Jointly Modeling People Commonness and Person Uniqueness. In: Proceedings of the AICM, pp 937–940. https://doi.org/10.1145/2647868.2654965
41. Yang F, Yan K, Lu S, Jia H, Xie X, Gao W (2019) Attention driven person re-identification. Pattern Recognit 86:143–155. https://doi.org/10.1016/j.patcog.2018.08.015
42. Ye M, Li J, Ma AJ, Zheng L, Yuen PC (2019) Dynamic Graph Co-Matching for Unsupervised Video-Based Person Re-Identification. Trans Image Process 28(6):2976–2990. https://doi.org/10.1109/tip.2019.2893066
43. Ye M, Shen J, Lin G, Xiang T, Shao L, Hoi SC (2021) Deep Learning for Person Re-Identification: A Survey and Outlook. Trans Pattern Anal & Mach Intell 44(6):2872–2893. https://doi.org/10.1109/TPAMI.2021.3054775
44. Yu Y, Yang S, Hu H, Chen D (2022) Attention-Guided Multi-Clue Mining Network for Person Re-identification. Neural Processing Letters pp 1–14. https://doi.org/10.1007/s11063-022-10757-1
45. Zhang J, Ainam JP, Song W, Zhao Lh, Wang X, Li H (2022) Learning global and local features using graph neural networks for person re-identification. Signal Processing: Image Communication p 116744. https://doi.org/10.1016/j.image.2022.116744
46. Zhang Y, Qian Q, Liu C, Chen W, Wang F, Li H, Jin R (2022) Graph convolution for re-ranking in person re-identification. In: Proceedings of the ICASSP, pp 2704–2708. IEEE
47. Zhang Z, Zhang H, Liu S (2021) Person Re-identification using Heterogeneous Local Graph Attention Networks. In: Proceedings of the CVPR, pp 12,136–12,145. https://doi.org/10.1109/cvpr46437.2021.01196
48. Zhang Z, Zhang H, Liu S, Xie Y, Durrani TS (2021) Part-guided graph convolution networks for person re-identification. Pattern Recognition 120(108):155. https://doi.org/10.1016/j.patcog.2021.108155
49. Zheng L, Shen L, Tian L, Wang S, Wang J, Tian Q (2015) Scalable person re-identification: A benchmark. In: Proceedings of the ICCV, pp 1116–1124. https://doi.org/10.1109/iccv.2015.133
50. Zheng Z, Zheng L, Yang Y (2017) Unlabeled samples generated by GAN improve the person re-identification baseline in Vitro. In: Proceedings of the ICCV, pp 3754–3762. https://doi.org/10.1109/iccv.2017.405

Authors and Affiliations

Nayan Kumar Subhashis Behera[1] · **Pankaj Kumar Sa**[1] · **Sambit Bakshi**[1] · **Umberto Bilotti**[2]

Nayan Kumar Subhashis Behera
nksbehera@gmail.com

Pankaj Kumar Sa
pankajksa@nitrkl.ac.in

Umberto Bilotti
ubilotti@unisa.it

1 Department of Computer Science and Engineering, National Institute of Technology Rourkela, 769008 Odisha, India

2 Department of Computer Science, University of Salerno, Salerno, Fisciano, Italy