The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 9 (2023) 809–814

www.elsevier.com/locate/icte

A novel person re-identification network to address low-resolution problem in smart city context

Irfan Yaqoob[a,1], Muhammad Umair Hassan[b,*,1], Dongmei Niu[a], Xiuyang Zhao[a], Ibrahim A. Hameed[b], Saeed-Ul Hassan[c]

[a] School of Information Science and Engineering, University of Jinan, Jinan, China
[b] Department of ICT and Natural Sciences, Norwegian University of Science and Technology, Ålesund, Norway
[c] Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, United Kingdom

Received 11 May 2022; received in revised form 23 July 2022; accepted 27 July 2022
Available online 3 August 2022

Abstract

We argue that accurate person re-identification is a vital problem for urban public monitoring systems in the smart city context. Since images captured from different cameras have arbitrary resolutions resulting in resolution mismatch, this work proposes a model that takes arbitrary images and converts them to a pre-defined fixed resolution. The model then passes the images to a super-resolution network, producing high-resolution images. We employ a feedback network to generate more realistic super-resolution images, which are fed to the re-identification network to acquire a unique descriptor to disclose the person's identity. We outperformed in all measures against other state-of-the-art methods.

Keywords: Person re-identification; High resolution; Smart cities; Deep learning

1. Introduction

Person re-identification (PReID) is a critical task in computer vision that aims to re-identify people across the network of different camera viewpoints. Significant breakthroughs have been achieved due to its importance in smart city applications such as computational forensics, video surveillance [1,2], and urban safety monitoring [3]. Automatically re-identifying people in different smart city cameras is one of the key challenges of today [4] and needs to be addressed by proposing novel solutions. For smart city video surveillance issues, researchers made full use of information and communication technologies, e.g., cloud computing, Internet of things, artificial intelligence, and mobile internet technologies, to support and process a massive amount of information [5–7]. The main objective of PReID is to reveal a person's identity captured across the various cameras. However, due to different resolution, occlusion, light intensity or viewpoint changes, and even variations in human pose, PReID has a challenging task for real-world applications.

To overcome the resolution mismatch problem, the super-resolution (SR) approaches have recently been adopted [8]. Many previous models are designed to assume that images captured by the cameras are of the same resolution, but this assumption is not valid in real life, as images captured from different cameras have an arbitrary resolution. Resolution mismatch problems occur when a high-resolution (HR) probe image is matched to a low-resolution (LR) gallery image, or vice versa, affecting PReID performance. The primary goal of such a technique is to retrieve the missing appearance information so that LR gallery images and HR probe images should be similar and considered equally in the re-identification network. For instance, Wang et al. [9] proposed Cascaded Super-Resolution Generative Adversarial Network (CSR-GAN) for scale-adaptive low-resolution PReID, which employs various GANs in series resulting in a great number of parameters. A large capability network contains huge room resources and suffers from the overfitting problem. Besides this, the CSR-GAN model has many other issues too. The

* Corresponding author.
E-mail addresses: irfanyaqoob373@gmail.com (I. Yaqoob), muhammad.u.hassan@ntnu.no (M.U. Hassan), dniu_ujn@hotmail.com (D. Niu), zhaoxy@ujn.edu.cn (X. Zhao), ibib@ntnu.no (I.A. Hameed), s.ul-hassan@mmu.ac.uk (S.-U. Hassan).
[1] The first two authors have equal contribution.
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2022.07.010

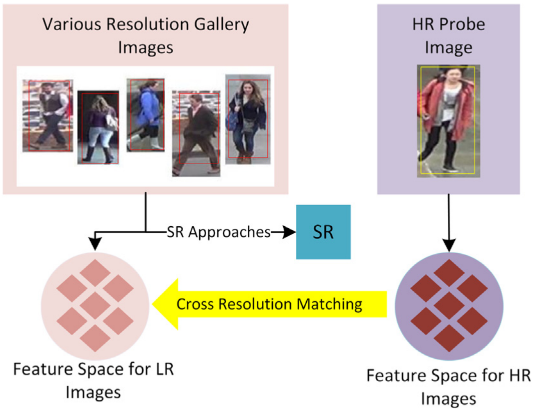

Fig. 1. Various LR gallery images need to match with HR target images. This leads to the cross resolution mismatch problem in PReID.

training of CSR-GAN is unstable; the model parameter oscillates, destabilizes, non-convergence occurs, and the generator collapses, which generates a limited variety of samples, and is extremely sensitive to hyper-parameters. To overcome the problem of LR person re-identification and improve PReID, we proposed a unique HR approach that performs better PReID as a joint learning optimal solution technique. Fig. 1 illustrates the low-resolution and cross-resolution PReID. A resolution mismatch is created when there is a discrepancy between the HR query images and LR gallery images and vice versa.

This work aims to address low-resolution PReID by utilizing the image SR strategy and PReID technique in a combined framework. Using the Alias IMAGE library, we scaled the LR gallery images to pre-defined fixed resolution images with a variational width (W) and height (H) for each dataset. As we aim to address the LR problem, the LR images are noisy and unclear and have blurred and coarse edges, so they are unable to refine the low-level information like color, pixel intensity, and edges that provide the baseline for high-level features, particularly after executing a down-sampling operation on the already existing LR images. To make the network perform better and extract unique patterns and refine the low-level features, we employ the services of super-resolution feedback network (SRFBN) [10] to regenerate HR counterparts of LR images. This model employs a feedback mechanism and provides a strong reconstruction ability. A dense skip connection helps to extract high-level information. Our proposed network utilizes a single SR network as compared to [11] resulting in a fewer number of parameters and time and memory complexity. To make the network more stable and lightweight, we apply ResNet50 network [12]. Random erasing [13], warm-up learning rate, and label smoothing [14] hyper-parameters are employed to increase the feature extraction capabilities of ResNet50.

We performed several experiments on different benchmarks and achieved superior results. We achieved 85.3%, 74.6%, 57% rank-1 accuracy, on DukeMTMC-reID [15], VR-Market1501 [16] and VR-MSMT17 [17] datasets, respectively. This paper is organized as follows. Section 1 introduces our approach to address low-resolution mismatch problem in smart cities context. We have described our proposed PReID method in Section 2. The evaluation protocols used by our network are available in Section 3. The experimental evaluations are illustrated in Section 4 while the concluding remarks are given in Section 5.

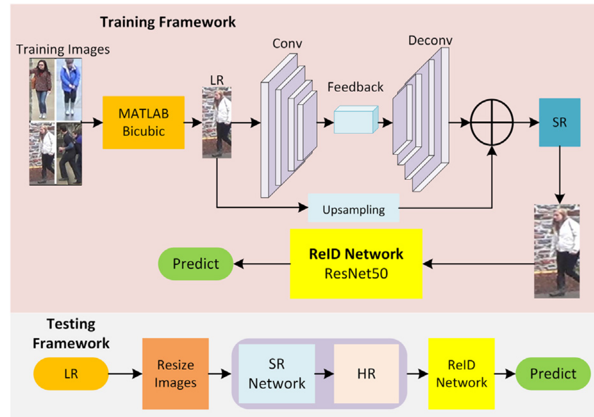

Fig. 2. An illustration of the proposed network consists of resizing the images, SR Network, and ReID network. The SR Network takes fixed sized LR images and regenerates them in the HR form, and these HR images are directly fed to the ReID network. The upper portion of the diagram shows the network's training, and the lower part illustrates the network's testing.

2. Person re-identification network

The proposed method is divided into three modules. First, a filter name alias from the IMAGE library is used to resize the arbitrary resolution of person images to a fixed resolution. Second, the super-resolution network consisting of SRFBN [10] is used to regenerate the LR images into super or high-resolution images. Third, our SR model's final HR image is fed to the ReID model to extract and learn unique features and patterns to predict the final output. The complete illustration of our network is provided in Fig. 2.

Generally, most HR-based PReID works assume that all person images have sufficiently HR, and they resize the images to a pre-defined uniform scale before re-identification. It is also common that the resolution of person images varies a lot, which occurs due to variations in the camera deployment settings and person–camera distance. In our proposed work, we used a convolution-based filter to handle the various resolution of gallery images, which allows our network to handle the various resolution in the training process. For this, we employ a scale adaptive module to allow our network to handle different resolutions to make it to a fixed size. The resolution sizes are mentioned in Table 1 for each dataset we used for our experimentation. The images of different resolutions are taken and resized to a pre-defined resolution, varying for each dataset. This strategy is utilized to save the computation power, as the network processes the images only at one scale. This enhances the overall efficacy of the network with less computation power.

Table 1
Information about the datasets used in our proposed work.

Benchmark	Identities			Images			LR Scale
	Train	Test	Total	Train	Test	Total	W ×H
DukeMTMC-reID	702	702+408	1812	16522	17661	36441	[30, 70]
Market1501	751	750	1501	12937	19733	32217	[8, 32]
MSMT17	1042	2246+807	4101	32964	68239	126411	[32, 128]

2.1. Super resolution network

A super-resolution feedback network [10] is used to improve the contextual information of the images so that regenerated images look more realistic and natural. To achieve the feedback capacity of hidden features against each iteration t from 1 to T, a recurrent neural network (RNN) is used.

The iteration t takes place in the sub-network in three blocks: (i) low-resolution features extraction block (LRFEB), (ii) feedback block (FB), and (iii) reconstruction block (RB). Each block shares weights across time. The primary objective of weight sharing is to recover the residual image by giving the input of LR image I_{LR} in the sub-network of the given framework. Our proposed network leverage upon Convolutional $Conv(s,n)$ and deconvolutional layers $Deconv(s,n)$, while S represents the filter size and the number of filters in the network are denoted by N. The convolutional features extraction blocks are of $Conv(3,4\,m)$ and $Conv(3,m)$, where m is the filter base. LR feature extraction input is I_{LR} and F^t_{in} retains the information of LR image, such as Eq. (1):

$$F^t_{in} = f_{LRFB}(I_{LR}) \tag{1}$$

f_{LRFB} is the operation for the previous block. For input F^t_{in} in FB, the F^1_{in} is the initial hidden state for F^o_{out}. The hidden state for the tth iteration in FB is received from the preceding iteration F^{t-1}_{out} by using a feedback connection in shallow features F^t_{in}. For FB, F^t_{out} represents the output. The below Eq. (2) is for FB:

$$F^t_{out} = f_{FB}(F^{t-1}_{out}, F^t_{in}) \tag{2}$$

The reconstruction and upscaling of LR features is performed by $Deconv(k,m)$ F^o_{out} whereas $(Conv(3, c_{out})$ outputs a residual image I^t_{res} using the following derived formulation:

$$I^t_{res} = f_{RB}(F^t_{out}) \tag{3}$$

Where f_{RB} represents the reconstruction block. For the tth iteration, we acquired the images using Eq. (4):

$$I^t_{SR} = I^t_{res} + f_{UP}(I_{LR}) \tag{4}$$

For upsampling kernel, F_{UP} operation summed up the tth iteration, and we got the final T and SR images in the network are obtained through $(I^1_{SR}, I^2_{SR}, I^T_{SR})$.

In FB, the tth iteration receives the information F^{t-1}_{out} helps to assemble the low-level representations F^t_{in}, after this process it passes more powerful high-level representations and give the yield F^t_{out} which can be used as next iteration for the reconstruction of the block. In FB, it encloses the groups of G projections sequentially with a dense layer by skipping the linking between them. Every projection group can project HR features to LR ones; mostly, it contains an up-sample operation and a down-sample operation. We get the refined information which is shown below:

$$L^t_0 = C_0([F^{t-1}_{out}, F^t_{in}]) \tag{5}$$

Whereas it states the initial compression C_0 and $[F^{t-}_{out}, F^t_{in}]$ shows the sequence of F^{t-1}_{out} and F^t_{in}. For HR H^t_g and LR L^t_g features maps the gth projection groups in the FB at tth iteration. Whereas H^t_g can be derived by the following:

$$H^t_g = C^{\uparrow}_g([L^t_0, L^t_1, \ldots, L^t_{g-1}]) \tag{6}$$

The $C^{\uparrow}_g$ refers to up-sample operation by using $Deconv(k,m)$ obtained as follows:

$$L^t_g = C^{\downarrow}_g([H^t_0, H^t_1, \ldots, H^t_g]) \tag{7}$$

where $C^{\downarrow}_g$ refers to the down-sampling operation.

To optimize the network, the L1 loss is used. For the HR images, the target T is placed inside the network for multiple outputs $I^1_{HR}, I^2_{HR}, \ldots, I^T_{HR}$, which is similar to single degradation of the model. In a complex model, they are ordered as the difficulty of the task for T iteration to impose a prospectus. We formulated the loss function in the network by the following equation:

$$L(\theta) = \frac{1}{T}\sum_{t=1}^{T} W^T \| I^t_{HR} - I^t_{SR} \| 1 \tag{8}$$

Whereas θ represents the parameters of the network, and the constant factor W^T determines the substance of yield at tth iteration.

2.2. Re-identification network

For ReID, we utilized the services of ResNet-50 as a standard baseline [12] with dropout and Kaiming weights initialization technique. We utilized the pre-trained weights of Image-Net as the initialization process of the proposed network, and the hyperparameters are used to boost the efficiency and performance of the ReID network.

2.2.1. Random erasing

In order to ensure that the network gives full attention to the whole image, random erasing is used, as proposed by Zhang et al. [13]. It randomly chooses a patch in an image and replaces it with a mean of pixel values ranging from 0 to 255 or random values (see Fig. 3). We use image and object-aware random erasing that pays attention to the image background and the object.

Fig. 3. An illustration of random erasing augmentation by randomly selecting the patches on the image.

2.2.2. Warm-up learning rate

Learning rate assumes a vigorous role in the presentation of a ReID model. It improves the overall performance of the network. Most standard techniques apply a huge and consistent learning rate to initially train the network. A warm-up approach is used to improve the efficiency and overall performance of the network. The learning rate $lr(t)$ at epoch t is calculated below:

$$lr(t) = 3.5 \times 10^{-5} \times \frac{t}{10} \tag{9}$$

where t represents learning rate decayed at epoch intervals.

2.2.3. Label smoothing

In order to prevent the ReID model from overfitting problems, the label smoothing (LS) is introduced by Szegedy et al. [14]. This technique is used to prevent overfitting for a classification job. Label smoothing increases the efficiency of the network.

$$(ID) = \sum_{t-1}^{N} -q_i \log(p_i) \begin{cases} q_i = 0, y \neq i \\ q_i = 1, y = i \end{cases} \tag{10}$$

2.2.4. Last stride

For the ReID network's better performance, higher spatial resolution is employed to improve the size of the feature space in the network. We also utilize the last spatial operation in our ReID network; this slightly increases the computation time but improves significantly.

3. Evaluation protocols

We used DukeMTMC-reID, VR-Market1501, and VR-MSMT17 datasets to evaluate our proposed work. Details against each dataset, e.g., total numbers of identities and images used for training and testing purposes, are available in Table 1.

3.1. DukeMTMC-reID

DukeMTMC-reID [15] is derived from DukeMTMC dataset for image-based ReID network which is designed for Market1501 dataset. It comprises of 16,522 training images of 702 identities, 2,228 query images of the other 702 identities and 17,661 gallery images.

3.2. VR-MSMT17

This dataset is extracted from 15 cameras, and consists of many real environmental problem images [18]. It is not developed for the low-resolution PReID problem; therefore, it is recreated into VR-MSMT17 (Various Resolution MSMT17). Down-sampled images have a size of [32, 128], having 96 different resolutions. 1042 identities are kept for training, and 2246+807 identities are used for testing purposes.

3.3. VR-market1501

VR-Market1501 [16] is developed in front of a supermarket at Tsinghua University, China. Market1501 is redesigned into VR-Market1501 (Various Resolution Market1501) like MSMT17. The width and height of all down-sampled images are [8, 32], comprising 24 different resolutions distinctly. 751 and 710 identities are used for training and testing purposes, respectively.

3.4. Implementation details

We utilized frequent performance metrics, cumulative matching characteristic (CMC top-K), and mean average precision (mAP) as evaluation metrics. The activation function of ReLU is employed, excluding the last layer in each sub-network, after all the convolution and deconvolution layers. The stride value is 2, and the padding value is 4. We assume the final result as SR from the last iteration until we specifically inspect each output for each iteration. The initial rate is set to 0.0001, and for every 40th epoch, it is increased by 0.5. In our experiments, we employed a total of 200 epochs. Training is done on RGB channels. Augmentation is executed on a training dataset with random horizontal flip and a 90-degree rotation. A pre-trained ReID network on ImageNet is trained with HR images gained from a SR network. For optimization of network parameters, an Adam optimizer is employed. The training process is done at 32 GB RAM, GPU NVIDIA 1080Ti. Our PReID model is developed using the PyTorch library.

4. Results and discussions

This section is divided into two modules. We exhibit the state-of-the-art performance of our method both qualitatively and quantitatively. The first module shows the efficiency of our SR network for the image creation challenge, while the second part will illustrate the competency of our ReID network on three benchmarks.

4.1. Super resolution network evaluation

Our SR network effectively generates HR images. Furthermore, it creates more realistic and natural images. Our SR network images are prominent and preserve all the essential information like background and clothing color. Moreover, it provides better and sufficient visibility, which helps for feature extraction. Fig. 4 illustrates the effectiveness of our proposed SR model. It can be seen from Fig. 4 that the LR images are successfully transformed to HR images by using our proposed approach.

Fig. 4. Some subjective results of our proposed model. Left side is of LR images, right side is the SR images generated by our model. All 6 images are selected from DukeMTMC-reID benchmark.

Table 2
The quantitative results of DukeMTMC-reID.

Method	Rank 1%	Rank 5%
SING [11]	65.2	80.1
CR-GAN [19]	77.2	88.1
CamStyle [20]	78.6	80.1
FD-GAN [21]	81.4	82
Proposed Model	**85.3**	**92.4**

4.2. Re-identification network evaluation

The comparison of our work is performed with state-of-the-art ReID techniques, such as SING [11], CSR-GAN [9], CR-GAN [19], CamStyle [20], FD-GAN [21], Densenet121 [22], SE-ResNet50 [23], and RIPR [17]. Our proposed method outperformed on all three benchmarks. The experiments are performed on three VR-MSMT17, VR-Market1501, and DukeMTMC-reID datasets. In comparison to the existing techniques, our model got distinguished results on DukeMTMC-reID. Our model achieved 85.3% rank 1 accuracy and 92.4% rank 5 accuracy, which in comparison to FD-GAN [21] 3.9% rank 1 and 10.4% rank 5 accuracy improvement. Table 2 illustrates the comparison of our model with other models at the DukeMTMC-reID benchmark. Bold and underlined are the top results.

On VR-Market1501, our model secured 74.6% rank 1 and 88.3% rank 5 accuracies. On VR-MSMT17, we yield 57.0% and 74.7% rank 1 and rank 5 accuracies, respectively. In comparison to RIPR [17], our model achieved 7.7%, 3.6%, 1.5% and 2% improved rank 1 and rank 5 accuracy on VR-Market1501 and VR-MSMT17 respectively. Table 3 depicts the supremacy of our network on the VR-Market1501 and VR-MSMT17 datasets.

Table 4 shows the experimental results of Rank 1% and mAP that our proposed model attained on all three DukeMTMC-reID, VR-MSMT17, VR-Market1501 benchmarks.

Table 3
The quantitative results of VR-Market1501 and VR-MSMT17 are compared.

Method	VR-Market1501		VR-MSMT17	
	Rank 1%	Rank 5%	Rank 1%	Rank 5%
DenseNet121 [22]	60	78.8	51.2	67.4
SE-ResNet50 [23]	58.2	78.6	52.3	68.9
SING [11]	60.5	81.8	52.1	68.3
CSR-GAN [9]	59.8	81.3	51.9	67.5
RIPR [17]	66.9	84.7	55.5	72.4
Proposed Model	**74.6**	**88.3**	**57.0**	**74.7**

Table 4
The Rank 1% and mAP score of our model achieved on all three benchmarks.

Benchmark	Rank 1%	mAP
DukeMTMC-reID [15]	85.3	71.9
VR-Market1501 [16]	76.6	47.9
VR-MSMT17 [18]	57	39.8

5. Conclusion

In this paper, we proposed a novel approach to solve the problem of low-resolution PReID, which is a solution to a resolution mismatch issue in smart city cameras. We resized the images of various resolutions and transformed them to a fixed resolution by using the Alias IMAGE library. Later, we passed these images to the SR model to regenerate the more realistic and natural images. These SR images are fed to the ReID model, which extracts the unique features that will ultimately help predict the person's identity. We tested our network's performance on three datasets, i.e., DukeMTMC-reID, VR-Market1501, and VR-MSMT17, and achieved better results than other state-of-the-art algorithms. We will focus on further improving this method and introducing more scalability in the future.

CRediT authorship contribution statement

Irfan Yaqoob: Conceptualization, Methodology, Data curation, Writing – original draft, Software. **Muhammad Umair Hassan:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Funding acquisition. **Dongmei Niu:** Supervision, Validation. **Xiuyang Zhao:** Supervision, Validation. **Ibrahim A. Hameed:** Investigation, Funding acquisition. **Saeed-Ul Hassan:** Writing – review & editing, Validation.

Declaration of competing 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.

Acknowledgments

We thank the anonymous reviewers for their critiques and comments that helped us improve our manuscript. We also acknowledge the support of Mao et al. [17], who provided us VR-MSMT17 dataset. Additionally, we are grateful to the Norwegian University of Science and Technology (NTNU), Norway, for providing the open access funding.

References

[1] K. Pawar, V. Attar, Deep learning based detection and localization of road accidents from traffic surveillance videos, ICT Express (2021).

[2] M. Shorfuzzaman, M.S. Hossain, M.F. Alhamid, Towards the sustainable development of smart cities through mass video surveillance: A response to the COVID-19 pandemic, Sustainable Cities Soc. 64 (2021) 102582.

[3] H. Nodehi, A. Shahbahrami, Multi-metric re-identification for online multi-person tracking, IEEE Trans. Circuits Syst. Video Technol. 32 (1) (2021) 147–159.

[4] C. Li, J. Li, Y. Xie, J. Nie, T. Yang, Z. Lu, Multi-camera joint spatial self-organization for intelligent interconnection surveillance, Eng. Appl. Artif. Intell. 107 (2022) 104533.

[5] J. Yang, B. Jiang, H. Song, A distributed image-retrieval method in multi-camera system of smart city based on cloud computing, Future Gener. Comput. Syst. 81 (2018) 244–251.

[6] S.-Y. Han, H.-W. Lee, Deep reinforcement learning based edge computing for video processing, ICT Express (2022).

[7] C. Zhang, P. Chen, T. Lei, Y. Wu, H. Meng, What-where-when attention network for video-based person re-identification, Neurocomputing 468 (2022) 33–47.

[8] K. Han, Y. Huang, C. Song, L. Wang, T. Tan, Adaptive super-resolution for person re-identification with low-resolution images, Pattern Recognit. 114 (2021) 107682.

[9] Z. Wang, M. Ye, F. Yang, X. Bai, S. Satoh, Cascaded SR-GAN for scale-adaptive low resolution person re-identification., in: IJCAI, Vol. 1, (2) 2018, p. 4.

[10] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, W. Wu, Feedback network for image super-resolution, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3867–3876.

[11] J. Jiao, W.-S. Zheng, A. Wu, X. Zhu, S. Gong, Deep low-resolution person re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, (1) 2018.

[12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[13] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, Y. Fu, Residual dense network for image restoration, IEEE Trans. Pattern Anal. Mach. Intell. 43 (7) (2020) 2480–2495.

[14] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[15] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, Performance measures and a data set for multi-target, multi-camera tracking, in: European Conference on Computer Vision, Springer, 2016, pp. 17–35.

[16] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, Q. Tian, Scalable person re-identification: A benchmark, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1116–1124.

[17] S. Mao, S. Zhang, M. Yang, Resolution-invariant person re-identification, 2019, arXiv preprint arXiv:1906.09748.

[18] Y. Ding, H. Fan, M. Xu, Y. Yang, Adaptive exploration for unsupervised person re-identification, ACM Trans. Multimedia Comput. Commun. Appl. (TOMM) 16 (1) (2020) 1–19.

[19] Y.-J. Li, Y.-C. Chen, Y.-Y. Lin, Y.-C.F. Wang, Cross-resolution adversarial dual network for person re-identification and beyond, 2020, arXiv preprint arXiv:2002.09274.

[20] Z. Zhong, L. Zheng, Z. Zheng, S. Li, Y. Yang, Camera style adaptation for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5157–5166.

[21] Y. Ge, Z. Li, H. Zhao, G. Yin, S. Yi, X. Wang, H. Li, Fd-gan: Pose-guided feature distilling gan for robust person re-identification, 2018, arXiv preprint arXiv:1810.02936.

[22] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[23] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

Irfan Yaqoob is currently a Ph.D. student at the Clarkson University. He obtained his master's diploma from the University of Jinan. His research interests include Image Processing, Virtual Reality, and Computer Graphics.

Muhammad Umair Hassan is a Ph.D. candidate at the Norwegian University of Science and Technology. He obtained his master's from the University of Jinan, P.R. China, and a bachelor's from the University of the Punjab, Pakistan. His research interests include Smart Cities, Computer Vision, Computer Graphics and Deep Learning. He has a soundtrack of publications in multidisciplinary areas, including journals and conference publications.

Dongmei Niu is a lecturer in School of Information Science and Engineering, University of Jinan. She received her B.S. in software engineering and Ph.D. in computer application technology from Shandong University in 2009 and 2016, respectively. From 2012 to 2014, she studied in UC Davis as a joint-training Ph.D. student.

Xiuyang Zhao is a professor in School of Information Science and Engineering, University of Jinan. He received his B.S. in material science, M.D. and Ph.D. in computational material science from Shandong University, in 1998, 2000 and 2006, respectively. His research interests include computer vision and CAGD.

Ibrahim A. Hameed has a Ph.D. in AI from Korea University, South Korea and Ph.D. in field robotics from Aarhus University, Denmark. He is a Professor and Deputy Head of Research and Innovation within NTNU, IEEE senior member, elected chair of the IEEE Computational Intelligence Society (CIS) Norway section, Founder and Head of Social Robots Lab in Ålesund. His current research interests include Artificial Intelligence, Machine Learning, Optimization, and Robotics.

Saeed-Ul Hassan is the Director of AI Lab at the Information Technology University (ITU), and a faculty member at the Manchester Metropolitan University. Dr. Saeed's research interests lie within the areas of Data Science, Machine Learning, Contextual Scientometrics, Altmetrics Bibliometric Tools for Evidence-based Research Policy Formulation, Information Retrieval, and Text Mining. He has published more than 50 papers in reputed international journals and conference proceedings including highly reputed venues like TKDE, WWW, JCDL, Journal of Infometrics, and Scientometrics. Dr. Saeed is also the recipient of the James A. Linen III Memorial Award in recognition of his outstanding academic performance. More recently, he has been awarded Eugene Garfield Honorable Mention Award for Innovation in Citation Analysis by Clarivate Analytics, Thomson Reuters.