IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 31, 2019, accepted July 7, 2019, date of publication July 18, 2019, date of current version August 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929854

Night Person Re-Identification and a Benchmark

JIAN'AN ZHANG, YUAN YUAN, (Senior Member, IEEE), AND QI WANG, (Senior Member, IEEE)
School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China
Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China

Corresponding author: Yuan Yuan (y.yuan1.ieee@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U1864204 and Grant 61773316, in part by the State Key Program of National Natural Science Foundation of China under Grant 61632018, in part by the Natural Science Foundation of Shaanxi Province under Grant 2018KJXX-024, and in part by the Project of Special Zone for National Defense Science and Technology Innovation.

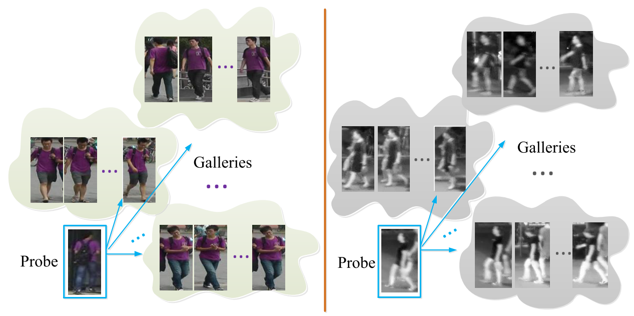
ABSTRACT Person re-identification is an important problem in computer vision fields due to its widely application. However, most of existing person re-identification methods are evaluated in daytime scenarios which is still far from real applications. In this paper, we pay attention to the night scenario person re-identification problem which most of works are not focused on. For this purpose, we contribute a large and real-scenario person re-identification dataset for night scenario named KnightReid, which aims to bridge the gap between theoretical research and practical application. To the best of our knowledge, the KnightReid dataset is the first night scenario dataset for the person re-identification which distinguishes existing works. Furthermore, by carefully examining the properties of night scenario data, we propose to combine image denoising networks with common used person re-identification networks to adapt to this kind of problem. Besides, we provide a comprehensive benchmark result that is evaluated on the dataset. The extensive experiments convince the effectiveness of the proposed model.

INDEX TERMS Night scenario person re-identification, re-identification dataset, denoising, deep neural network.

I. INTRODUCTION

Recently, person re-identification has drawn more and more attention in the computer vision field due to its important applications in many real scenarios such as video surveillance, robotics, automated driving [1], [2]. Generally, person re-identification refers to matching a given pedestrian across non-overlapping cameras in a network. In practice, a person re-identification system consists of person detector, person tracker and person matcher. Because person detection and person tracking are independent problems in computer vision, person re-identification usually focuses on the third part and it is often regarded as a retrieval problem.

Although person re-identification achieves constant improvements in recent years, it is still far from real applications. Especially, some special cases are very important in practice such as night scenario and rain scenario. However, there are less works that pay attention to. To bridge the gap between theoretical research and practical application, in this paper, we focus on the person re-identification in the night scenario for the first time. As we can expect, night scenario person re-identification is important for practical applications as nearly half a day is in night time.



(a) (b)

FIGURE 1. Illustration of person re-identification problem in the daytime scenario and the night scenario. Given a probe image, search its matches in galleries. (a) Person images in the daytime scenario, (b) person images in the night scenario.

Fig.1 shows the person re-identification problem in the daytime and night scenario. From Fig.1, it can be found the main difference between daytime scenario person re-identification and night scenario person re-identification is that the images in night scenario show less color information

The associate editor coordinating the review of this manuscript and approving it for publication was Long Wang.

while the images in daytime scenario show much color information. However, night scenario person re-identification is much more difficult than the daytime scenarios. As the appearance of a pedestrian changes dynamically under different cameras, the general person re-identification problem shows many difficulties such as low resolution, illumination changes, various poses, occlusion and background clustering. For example, as shown in Fig.1 (a), the three different gallery persons all wear purple short sleeves and light blue jeans, which looks similar to each other. Apart from the above mentioned problems, different from daytime scenario person re-identification, images of night scenario person re-identification suffer from serious noising. This makes it more difficult. As shown in Fig. 1, all the images of the three different persons wear dark color short sleeves and light color pants, which not only look similar to each other but also have serious noise.

Although night scenario person re-identification is important for both theoretical research and practical application, there are few works focusing on it. To our best knowledge, there are only two pioneer works [3], [4] that are related with but different from night scenario person re-identification. Reference [3] was the first work that focuses on infrared images based person re-identification. However, the images they used for evaluation are not captured in a night scenario. Reference [4] constructed a new dataset named SYSU-MM01 that contains both RGB images and infrared images. However, such a dataset is designed for cross-modality RGB-IR person re-identification problem, which is not designed for night scenario person re-identification. Furthermore, both the above two works have their own limits as they have not or have not enough real scenario data. So the most urgent problem is to collect enough person data in night scenario for person re-identification to satisfy the research demand.

To meet the urgent demand for night scenario data, we contribute a large real surveillance scenario person re-identification dataset in night time named KnightReid (for its pronunciation is same as night re-id) to help the re-identification research community. In fact, it is the first person re-identification dataset that specially designed for night scenario. The constructed KnightReid dataset consists of 937 identities captured in three cameras. A total of 315354 bounding boxes are cropped from video frames to constitute the whole dataset. All the bounding boxes are hand-drawn from a real monitor video surveillance system.

As the night scenario person re-identification is different from general person re-identification. Classical person re-identification methods may not be adapted to such a special case. So it is required new methods to solve this kind of problem. As an attempt, taking the noising property of night scenario images into consideration, we propose to combine a denoising network with a re-identification network to handle this kind of problem. Experimental results confirm the effectiveness of this pipeline.

In a word, the main contribution of this work can be summarized as

- A new dataset named KnightReid is constructed to meet the demand of data in night scenario and to help the person re-identification research community.
- A new deep network model that combines a denoising network and a re-identification network is designed specially for the person re-identification in night scenario.
- Extensive experiments are conducted on the proposed dataset and benchmark results are provided including hand-designed feature methods, metric learning methods and deep learning methods.

In the next section, we will review related works. And then a detailed description about the proposed KnightReid dataset is given in Section III. Next, we will present the proposed method in Section IV. Section V presents an extensive comparison with state-of-the-art algorithms. Section VI concludes the paper and discusses future works.

II. RELATED WORKS

In this section, we will briefly introduce some works with regards to our paper including person re-identification datasets, classical hand-designed feature and metric learning methods, deep learning based methods and some related image denoising methods.

A. DATASETS FOR PERSON RE-IDENTIFICATION

There are many existing datasets for person re-id. An overview of the statistic of these datasets is shown in Table 1. Among all the datasets, VIPeR [5] is the most popular dataset for it is the first person re-id dataset, but the dataset is small and there is only one image for every identity in each of the two camera views. CUHK01 [8] is proposed with more identities than VIPeR but still with two camera views. Later, PRID2011 [7] and CUHK03 [9], [16] are constructed with multi-camera views and multi-images for one identity. Recently, with the need of training for deep neural networks, large datasets with multi-camera views and multi-bounding box are proposed one after another, such as Market1501 [11], DukeMTMC4ReID [13], [17] and MSMT17 [14]. Besides, BPReid [15] dataset is a special person re-identification dataset which focus on a special kinds of data, namely, a person with bikes. However, all these large scale datasets are in daytime scenario. Different from existing datasets, the proposed Knight dataset focuses on the person images in night scenario for the first time. All the images in the KnightReid dataset are captured in night scenario. Besides, it is a large scale person re-id dataset in real surveillance environment with multi-camera views and multi-images for each identity that can be used for training deep neural networks.

B. CLASSICAL METHODS FOR PERSON RE-IDENTIFICATION

Before the rise of deep learning methods, there are two mainstream methods for person re-identification, namely,

TABLE 1. Comparing with existing datasets.

Dataset	Year	#person	#BBox	#cameras	Label Method	Image size	Multi-shot
VIPeR [5]	2007	632	1264	2	hand	128×48	No
GRID [6]	2009	1025	1275	8	hand	Varied	No
PRID2011 [7]	2011	934	24,541	2	hand	128×64	√
CUHK01 [8]	2012	971	3,884	2	hand	160×60	√
CUHk03 [9]	2014	1467	13,164	10	hand/DPM	Varied	√
iLIDS-VID [10]	2014	300	42,495	2	hand	Varied	√
Market1501 [11]	2015	1501	32,217	6	hand/DPM	128×64	√
MARS [12]	2016	1261	1,119,003	6	DPM+GMMCP	256×128	√
DukeMTMC4ReID [13]	2017	1852	46,261	8	Doppia	varied	√
SYSU-MM01 [4]	2017				hand	varied	√
MSMT17 [14]	2018	4101	126441	15	Fater RCNN	varied	√
BPReid [15]	2018	4579	200,680	6	hand	varied	√
KnighReid	**2019**	**937**	**315354**	**3**	**hand**	**varied**	√

hand-designed feature methods and metric learning methods. Feature-based approaches focus on design discriminative and robust features for pedestrians such as ELF [5], SDALF [18], LOMO [19] and GOG [20]. Among these features, ELF is the most used feature and LOMO and GOG are regarded as the state-of-the-art methods. Metric learning methods aim at designing discriminative metrics for matching and most of the methods used for person re-id are on the basis of Mahalanobis distance. A large majority of methods fall into this group. Classical methods include LMNN [21], ITML [22], LDML [23], KISSME [24], PCCA [25], LFDA [26], LADF [27], XQDA [19] and MLAPG [28], where LFDA, KISSME, XQDA and MLAPG are regarded as the state-of-the-arts. This work provides a benchmark results of both the feature based methods and metric learning methods on the proposed KnightReid dataset.

C. DEEP LEARNING METHODS FOR PERSON RE-IDENTIFICATION

Recently, deep learning methods have been successfully applied in many computer vision tasks [29], [30] and a lot of deep learning methods are developed, which makes a great success in person re-identification. Existing deep learning methods for person re-identification either regard it as a ranking problem or a classification problem. The ranking related methods take three images as input where two of them are of the identity and another is from a different identity. A triple loss or improve triple loss is often adopted, such as DRDC [31], Quadruple [32] and [33]. Classification related works usually adopt classification losses with a carefully designed special structure for person re-id, such as binary classification related methods FPNN [9], Impl [34], and Gated [35]. Multi-classes related methods are also developed such as SpindleNet [36], PSE [37], HACNN [38]. In this paper, we will evaluate how deep learning methods contribute to the night scenario person re-identification on the KnightReid dataset.

D. IMAGE DENOISING METHODS

As the images in night scenario suffer from serious noise, we will combine image denoising methods into person re-identification network to handle the night scenario person re-identification problem. In this paragraph, we will give a brief introduction to image denoising methods. Image denoising aims to recover a high-quality image from a noisy image and many methods have been proposed in the past decades. Among so many denoising methods, NLM [39] and BM3D [40] are two best-known methods, which take the local structure into consideration. Recently, with the advance of deep neural network models, many works turn to this kind of method and achieve state-of-the-art performance, such as NLCNN [41], DnCNN [42], UDNet [43], Noise2Noise [44] and N3 [45]. All these deep neural network methods focus either on model designing or on special module designing, which show much improvements compared with other methods.

III. THE KNIGHTREID DATASET

In this section, we will give a detailed description of the KnightReid dataset and introduce the difference between general person re-identification and night scenario person re-identification. Besides, we will also introduce how difficult the night scenario person re-identification is.

A. DESCRIPTION

The KnightReid dataset are collected from a real video surveillance system in a campus environment. All frames in the BPReid dataset are captured by 3 static cameras at 25 frames per second, which consists of the subset of the real surveillance system. All the 3 cameras are high definition (HD) with size 1920×1080. We labeled all the 3 cameras in continuous nighttime with average 3 hours per camera. In total, this dataset contains 937 identities which are captured at least by 2 cameras and more than 31k bounding boxes are acquired.

In fact, the KnightReid dataset can be divided into three sub-datasets according to the layout of the 3 cameras as shown in Figure 2. From this figure, it can be inferred that a person can appear in all the three camera views. So we can get three camera pairs cam1-cam2, cam1-cam3 and cam2-cam3. We will evaluate the proposed algorithm on all the three setting in experiments section to imitate real application scenarios.

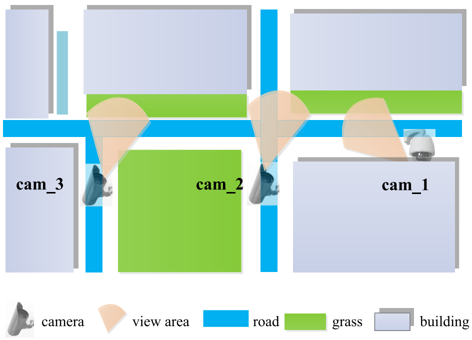


FIGURE 2. **The layout of cameras in KnightReid dataset.**

Fig. 3 shows the detailed information of the KnightReid dataset. Fig. 3 (a) shows the statics of every camera pair. The number of identities for every camera pair is 474, 176 and 639 respectively. Fig. 3 (b) shows the number of bounding boxes (images) for every camera pair. As we can find the pair cam2-cam3 has the most identities as well as bounding boxes while the pair cam1-cam3 has least. In addition, the sizes of images in the KnightReid dataset are various. We show the distribution of the width and height of images in Fig. 3(c). It can also reflect what the real application scenario is. Besides, some typical samples of the images can be found in Fig. 4.

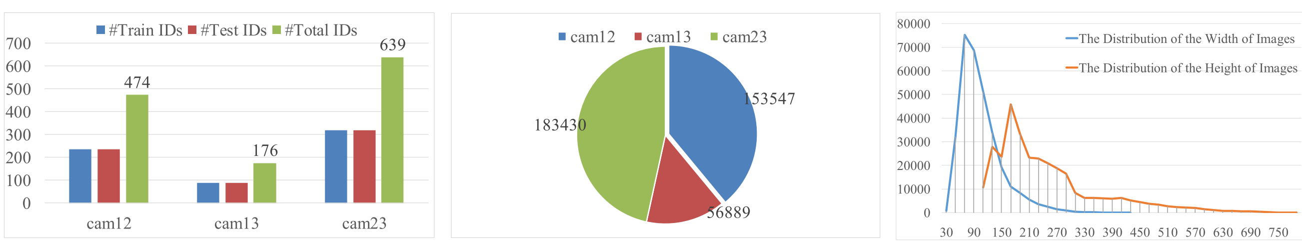
Compared with existing person re-identification datasets, the KnightReid extinguishes them in three aspects. First, it is the first night scenario dataset customized for person re-identification. To our knowledge, the KnightReid dataset is the first dataset evaluated for the practical night scenario person re-identification problem. It has 937 different identities as well as more than 31k bounding boxes, which makes it suitable for deep learning. Second, the proposed KnightReid dataset is collected from a real surveillance monitor system which make it closer to real application. As we know, existing datasets, such as Market-1501 and CUHK01, are all sampled from a temporal cameras which are not real enough for person re-identification application. While the KnightReid dataset shows much real scenario. Third, images in KnightReid dataset introduce more noise than existing datasets. As the all the images are collected in a night time, noise is inevitably introduced. Apart from low-resolution, the noise problem is the major difference compared with images in daytime scenario.

B. COMPARISON WITH GENERAL PERSON RE-IDENTIFICATION

From what we have introduced in the previous section, it is easy to find the person re-identification in night scenario is a variant of person re-identification. It differs from general person re-identification mainly in the aspect that it focuses on images in night time which often captured by infrared cameras.

Similar to general person re-identification problem, the person re-identification in night scenario also suffers from typical challenge problems that general person re-identification has such as illumination changes, occlusions, different viewpoint, complex background. Besides, the night scenario person re-identification problem also introduces some new challenges. We will demonstrate the similarity and difference compared with general person re-identification by some examples as shown in Fig. 4.

Fig. 4 (a) shows two occlusion samples which are occluded by a person and a retractable door respectively. Especially in the first row of person images, it is nearly totally occluded by another person, which makes it hard to distinguish the right person. Fig. 4 (b) shows one person of different views captured by different cameras. Different viewpoints caused the misalignment problem which is one of the main challenges for person re-identification. Fig. 4 (c) illustrates the case that different persons are of high similarity. Each row consists of two similar persons of different identities. As the night scenario images are only shown black-white color information, so the inter-class similarity problem is more serious than general person re-identification. Fig. 4 (d) displays two illuminance change samples. Different from illuminance change in general person re-identification, night scenario illuminance change often comes from street lamp and vehicle light. While the daytime scenario illuminance change comes from sunlight. So the illuminance of images in night scenario often varies dramatically as shown in (d). Fig. 4 (e) demonstrates a typical problem, namely, complex background. Compared with general person re-identification, this problem is much more serious. As we can find in (e), when the light is too dark, it hard to distinguish a person from its complex background.



(a) (b) (c)

FIGURE 3. **Statistics of the proposed KnightReid Dataset. (a) The number of identities of every camera pair as well as the number of training identities and test identities. (b) The number of images in each camera pair. (c) The distribution of the width and height of images in the whole dataset.**

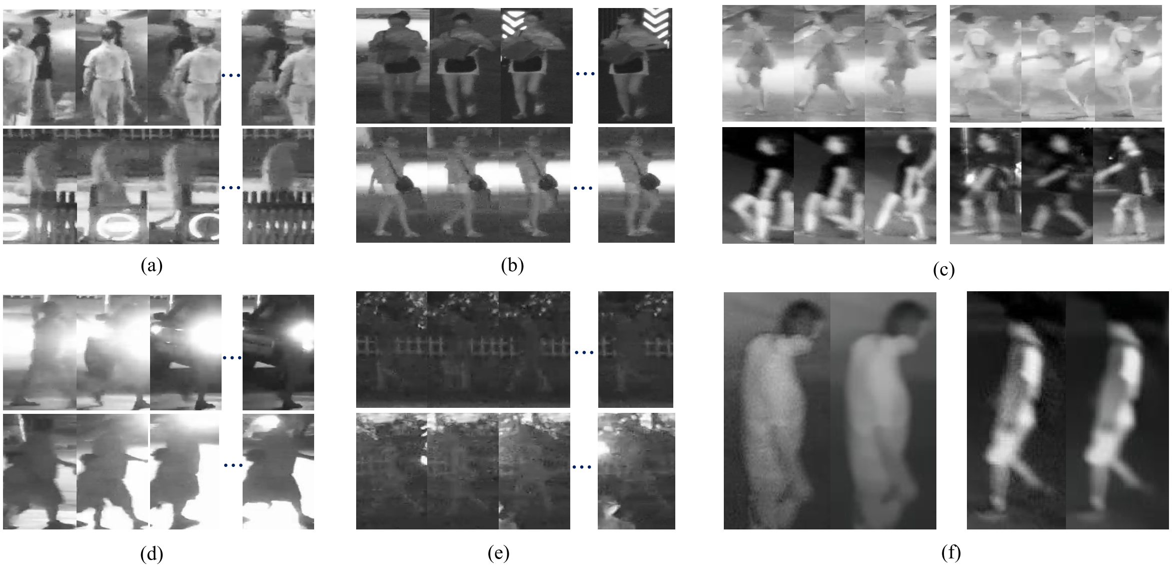


FIGURE 4. Challenges of the night scenario person re-identification. (a) Occlusions. (b) View points changes. (c) Person of highly similarities. (d) Illuminance changes. (e) Complex background. (f) Serious noise.

In fact, it is also very hard for us to label such identities as their characteristics are not evident between cameras. So such a challenge make it more difficult than general person re-identification. Fig 4(e) shows the typical image noise problem the night scenario person re-identification has. In each group, the left image is an original night scenario one while the right one is its denoising version. It is a general case for night scenario images. Although a part of images in daytime have also such a problem, but they are not as common as the images in night scenario. Hence this is a main difference compared with general person re-identification.

IV. MOTHED

In this section, we will introduce our idea for solving the night scenario person re-identification problem. Classical person re-identification methods may not be adapted to the night scenario person re-identification as it is different from general person re-identification in many aspects. So it is required to develop new methods to solve this kind of problem.

As we can see from previous sections, the main challenge of night scenario person re-identification is that the images are of low resolution and high noising. So it is natural to come out that we can improve the images' quality with some techniques such as image denoising. Note that, apart from person re-identification, the image denoising fields also benefit from deep neural networks as discussed in related work section. Namely, both the image denoising and person re-identification can be modeled as deep neural network models. So we can boost the quality of the image with a noising network before it feeds to a person re-identification network. Motivated by the jointly trained techniques and the success of neural network models for both image denoising and person re-identification problems, we propose to combine a denoising network with a re-identification network to handle night scenario person re-identification problem. We argue that it can benefit from such as structure in two folds. On one hand, we can get rid of the noise and one the other hand we can get low-level features as a deep neural network has the ability.

The whole structure of the proposed method is shown as Fig. 5. It consists of two sequential subnets which are used for denoising and feature extraction respectively. Generally, the denoising network consists of convolution modules and pooling modules in low layers and up-pooling and deconvolution modules in high layers. There are no fully connected layers as they cannot handle different image size. Apart from general convolution and pooling modules, the person re-identification networks always have some special structure that is suit for this problem such as attention modules. In fact, it can be regarded as a basic pipeline and any networks can be combined. In this work, we adopt N3 [45] network for image denoising and HACNN [38] for person re-identification as they show state-of-the-art performance for image denoising and person re-identification.

Besides, it is also important to adopt effective loss functions to train a deep neural network model. In image denoising field, PSNR (Peak Signal to Noise Ratio) is a widely accepted evaluation indicator, which can be used as a loss function. Its formulation is shown as follows,

$$\mathcal{L}_{psnr} = 10\log_{10}\left(\frac{L^2}{\sum_{ij}(O_{ij} - I_{ij})}\right). \tag{1}$$

where O_{ij} and I_{ij} are the output noised image and input image respectively.

And in person re-identification, the triplet loss is the widely used one. Define $\mathcal{L}_{tri}$ as the hard-mining triplet loss function and its formulation is

$$\mathcal{L}_{tri} = \frac{1}{N_s}\sum_{a=1}^{N_s}[\max_{y_a=y_b} d(\mathbf{f}_a, \mathbf{f_b}) - \min_{y_n\neq y_a} d(\mathbf{f}_a, \mathbf{f}_n) + \tau]_+. \tag{2}$$

FIGURE 5. The proposed pipeline network for the night scenario person re-identification.

where $\mathbf{f}_a, \mathbf{f}_b, \mathbf{f}_n$ are features extracted by a network and $\mathbf{f}_a$ and $\mathbf{f}_b$ are of the same label $y_a = y_b$ while $\mathbf{f}_a$ and $\mathbf{f}_n$ are of the different label. Eq (2) aims to push the same class features closer while pull the different class features far from the distances of same class features than a margin τ.

In order to get a effective network, we propose to jointly train the whole network with the following loss function.

$$\mathcal{L}_{total} = \lambda \mathcal{L}_{psnr} + (1 - \lambda)\mathcal{L}_{tri}, \tag{3}$$

where the total loss function $\mathcal{L}_{total}$ is the linear combination of $\mathcal{L}_{psnr}$ and $\mathcal{L}_{tri}$. And the parameter λ controls the trade-off between the two losses and in experiments we set $\lambda = 0.1$.

V. EXPERIMENTS

In this section, we will give a comprehensive evaluation on the KnightReid dataset including benchmark results tested on the dataset and the results of the proposed method.

A. SETUP

1) DATASET SETTINGS

There are two kinds of settings for multi-camera dataset in existing person re-identification works. The first one ignores the camera pair information and conducts experiments by treating the dataset as a whole part, such as experiments tested on Market1501 dataset. The second one takes the camera pair information into consideration and conducts experiments every camera pair. In fact for the KnightReid dataset, both the above two settings can be used. But, as we can find, the second setting is closer to real scenario. So we turn to this kind of setting and conducts experiments every camera pair. Namely, we will evaluate the performance of all the comparison algorithms in cam1-cam2, cam1-cam3 and cam2-cam3 settings.

2) EXPERIMENTAL SETTINGS

All the network models used in this work are implemented by PyTorch and run on a computer configured with NIDIA Tesla K80 GPU cards. For all the experiment, training data is randomly divided into mini-batches with batch size of 128. Forward propagation is performed on each mini-batch and the loss is computed. After then, back propagation is executed to compute the gradients and weights are updated with Adam optimizer. We set the initial learning rate to $3e^{-4}$ for all the three settings. We use a momentum of 0.9 and weight decay 5×10^{-3}. For the person re-identification, we utilize the pretrained model on ImageNet to initialize the weights. Due to the different size of data in three settings, so we set different training epoches for different settings. Namely, we set 150, 100 and 200 epoch for cam1-2, cam1-3 an cam2-3 respectively. Besides, for the total training loss, we set $\lambda = 0.1$ for all the three setting.

3) COMPETITORS

We compare the proposed model against some existing state-of-the-art person re-identification methods as listed in Table 2. These methods adopt different pipelines such as hand-crafted methods, hand-crafted combined with metric learning and deep learning based methods. Besides, to verify the performance of the proposed pipeline, we also take the denoising methods into consideration.

B. BENCHMARK RESULTS AND ANALYSES

In this section, we will give benchmark results of typical person re-identification methods.

1) COMPARE WITH HAND-DESIGNED FEATURES METHODS

We evaluate three most widely used hand-designed feature in person re-identification related works, namely, ELF [5], LOMO [19] and GOG [20]. The performances of the three methods in every setting are shown in Table 2. Generally speaking, all the hand-designed features perform poor compared with their performances in related daytime scenario datasets. The main reason is that the night scenario images introduce more challenges than the daytime scenario ones. Besides, hand-designed features often focus on the combination of low-level features such as color and texture, and these low-level features are not as useful as the daytime scenario case especially for the color features. It can be found that the GOG method achieves better performance than ELF and LOMO. As we known, both ELF and LOMO methods extract

TABLE 2. Comparison results on Night Dataset.

Methods	cam1-cam2					cam1-cam3					cam2-cam3				
	r1	r5	r10	r20	mAp	r1	r5	r10	r20	mAp	r1	r5	r10	r20	mAp
ELF [5]	1.2	2.4	3.7	5.7	1.6	0.8	2.3	4.8	9.0	1.9	1.0	2.6	4.0	6.5	1.1
LOMO [19]	0.4	2.5	4.8	8.4	0.8	1.1	4.5	8.6	24.9	1.7	0.4	2.3	4.3	7.8	0.7
GOG [20]	2.2	5.0	7.3	10.8	2.2	2.1	4.9	7.2	11.2	2.8	2.5	5.8	8.4	12.3	2.4
GOG+MLAPG [28]	4.5	10.2	15.1	22.1	3.5	3.1	8.7	13.5	20.5	3.9	7.7	16.6	23.0	30.9	5.9
GOG+KISSME [24]	5.6	13.1	17.9	24.4	3.4	5.2	14.4	21.1	30.7	4.3	7.4	16.7	22.7	30.5	4.3
GOG+XQDA [19]	7.4	16.2	22.9	31.4	5.1	3.3	11.4	19.1	30.4	3.1	9.5	21.2	29.1	38.7	5.6
GOG+LFDA [26]	7.0	15.4	23.2	32.3	5.2	4.4	13.4	21.2	31.8	4.1	9.5	22.5	30.8	40.6	6.6
Baseline [37]	9.6	15.5	19.7	24.6	7.1	5.5	14.2	20.8	28.9	4.0	16.1	24.9	29.8	35.4	11.7
HACNN [38]	10.9	17.8	22.1	27.4	7.9	6.1	15.5	21.7	29.3	5.1	13.1	20.5	24.7	29.6	9.1
NLM+Baseline [39]	9.1	15.7	20.1	25.2	7.4	5.4	14.5	21.2	29.1	4.1	15.9	24.5	29.7	35.6	11.6
BM3D+Baseline [40]	9.7	16.0	21.7	26.6	7.9	5.6	15.2	22.8	30.4	4.7	16.4	25.3	30.9	36.8	12.1
w/o PSNR Loss	10.5	16.5	22.7	28.1	8.1	6.1	15.8	22.2	30.6	5.2	13.2	21.8	25.6	30.7	9.6
With PSNR Loss	11.1	16.7	23.2	28.6	8.3	6.3	16.5	23.4	31.2	5.5	14.3	22.5	26.7	31.4	10.2

the histograms of different color space, while GOG methods exploit more texture information. For the images in night scenario, they share similar color information which indicates that color histogram based features are not as efficient as it in daytime scenario. So it is no surprise that the GOG method achieves the best performance among the three hand-designed methods.

2) COMPARE WITH METRIC LEARNING METHODS

Metric learning methods are another kind of classical methods for person re-identification. We evaluate four methods including LFDA [26], KISSME [24], XQDA [19] and MLAPG [28], which are regarded as state-of-the-arts. All the metric learning methods are tested based on the GOG feature as it shows better discriminant ability than ELF and LOMO. And the results evaluated on the three settings are listed in Table 2. As we can expect, with metric learning methods, the performances on all the three settings are largely enhanced. This also indicates that metric learning methods are still effective for night scenario person re-identification. However, compared with the results evaluated in other daytime scenario dataset, the performance is still poor due to the difficulty of the night scenario person re-identification. From Table 2, we can find that the LFDA and XQDA methods achieve best performance among all the four metric learning methods. We argue that this is because LFDA and XADA are all FDA-based methods which try to separate different classes to the maximum and they may adapt more to such kind of data.

3) COMPARE WITH DEEP LEARNING METHODS

Deep learning based methods are recent state-of-the-arts for person re-identification. We find it is very hard to train a deep learning model especially in cam1-cam2 and cam1-cam3 settings. It often takes about one week to train a convergent model on one setting as we have to adjust the learning rate and other parameters. Taking consideration of the limits of computation source, for this kind of methods, we only compare two classical state-of-the-art methods HACNN [38] and Resnet-50 [37] and we regard Resnet-50 as Baseline. The performance on the three settings are shown in Table 2. Generally speaking, with deep learning model, the re-identification performances are largely enhanced due to their discriminative representation ability. Compared with the Resnet-50 Baseline, the HACNN model achieves better performance in the cam1-cam2 and cam1-cam3 settings as it imposes an attention structure to the network and has the ability to capture the most import feature. However, as shown in Table 2, the HACNN method achieves lower performance than the Baseline method in the cam2-cam3 setting. We argue that this is because the HACNN method is trained from scratch while the Baseline model has a pretrained model and the cam2-cam3 setting introduces more challenges. So this leads to the degraded performance than Baseline model.

4) COMBINED WITH IMAGE DENOISING METHODS

We evaluate two widely used image denoising methods NLM [39] and BM3D [40] as pre-processing methods for Resnet-50 model. In fact, we first apply NLM or BM3D denoising method to all images in the KnightReid dataset and store them as enhanced datasets. Next, we train the Resnet-50 model on the enhanced datasets to evaluate the effect of image denoising methods. As shown in Table 2, in general with denoising methods, it improves the performance compared with the Baseline especially for the mAP evaluation. However, we have to note that it only improves a limited performance. By carefully checking the result and investigating the whole network structure, we argue that the discriminative ability of neural network account for this result. As we known, the convolution module has an ability of denoising in the low levels of a neural network. So with denoised images, the baseline shows limited improvement. But, we believe it is also effective as a pre-processing step for person re-identification due to their improvements.

C. COMPARISON OF THE PROPOSED METHOD

1) EFFECTIVENESS OF THE PROPOSED METHOD

We combine the N3 [45] denoising method and the HACNN [38] method to jointly train a person re-identification model. As shown in Table 2, the proposed method shows its effectiveness compared with the HACNN method. Generally, it improves mAP that ranges from 0.5 to 1.1 and rank-1 rate that ranges from 0.2 to 1.2. Especially, in cam2-cam3 setting, it improves more than 1.0 mAP and rank-1 rate. We argue that combining the denoising network rather than pre-prossing denoising methods, the person re-identification network can get rid of noise more effective as the networks are compatible with each other. Besides, different from pre-prossing denoising methods, the jointly learned denoising network can have the ability of capturing low-level feature for person re-identification as the convolution network does, while pre-prossing denoising methods have not. So the proposed method is effective as we expect.

2) WITH/WITHOUT PSNR LOSS

The loss function is essential for training a good deep neural network model. So we evaluate how the PNSR loss impacts the performance. As shown in Table 2, we can find easily that the proposed method is improved with the PSNR loss. Compared with the setting without PSNR, the loss with PSNR is more effective than the original triplet loss. The result is the same as we can expect as the intuition of design such a loss is to make the two sub-networks compatible with each other and yield an effective model.

VI. CONCLUSION AND FUTURE WORK

In this work, a new dataset named KnightReid is constructed for the use of research in night scenario person re-identification. This is the first person re-identification dataset that is designed for night scenario. There are 937 identities and more than 31k bounding boxes in the KnightReid dataset. Besides, by carefully observing the properties of images in the night scenario, we make an attempt to combine the N3 denoising network with the HACNN re-identification network to adapt to such a special problem. And in order to make the two kinds of network compatible, a new loss function that combines the triplet loss and the PSNR loss is proposed to jointly train the proposed pipeline. Finally, we evaluate different kinds of methods on the dataset and give a benchmark results including hand-crafted feature methods, metric learning methods, deep learning methods and combined denoising methods.

REFERENCES

[1] L. Sun, Z. Jiang, H. Song, Q. Lu, and A. Men, "Semi-coupled dictionary learning with relaxation label space transformation for video-based person re-identification," *IEEE Access*, vol. 6, pp. 12587–12597, 2018.
[2] T. Ni, Z. Ding, F. Chen, and H. Wang, "Relative distance metric leaning based on clustering centralization and projection vectors learning for person re-identification," *IEEE Access*, vol. 6, pp. 11405–11411, 2018.
[3] K. Jungling and M. Arens, "Local feature based person reidentification in infrared image sequences," in *Proc. AVSS*, Boston, MA, USA, 2010, pp. 448–455.
[4] A. Wu, W.-S. Zheng, H.-X. Yu, S. Gong, and J. Lai, "RGB-infrared cross-modality person re-identification," in *Proc. ICCV*, Venice, Italy, 2017, pp. 5380–5389.
[5] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. ECCV*, Munich, Germany, 2008, pp. 262–275.
[6] C. C. Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," in *Proc. CVPR*, Miami, FL, USA, Jun. 2009, pp. 1988–1995.
[7] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Proc. SCIA*, Berlin, Germany, 2011, pp. 91–102.
[8] W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in *Proc. ACCV*, Daejeon, South Korea, 2012, pp. 31–44.
[9] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: Deep filter pairing neural network for person re-identification," in *Proc. CVPR*, Columbus, OH, USA, Jun. 2014, pp. 152–159.
[10] T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *Proc. ECCV*, Zurich, Switzerland, 2014, pp. 688–703.
[11] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proc. ICCV*, Dec. 2015, pp. 1116–1124.
[12] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "MARS: A video benchmark for large-scale person re-identification," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 868–884.
[13] M. Gou, S. Karanam, W. Liu, O. Camps, and R. J. Radke, "DukeMTMC4ReID: A large-scale multi-camera person re-identification dataset," in *Proc. CVPR Workshops*, Honolulu, HI, USA, Jul. 2017, pp. 10–19.
[14] L. Wei, S. Zhang, W. Gao, and Q. Tian, "Person transfer GAN to bridge domain gap for person re-identification," in *Proc. CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 79–88.
[15] Y. Yuan, J. Zhang, and Q. Wang, "Bike-person re-identification: A benchmark and a comprehensive evaluation," *IEEE Access*, vol. 6, pp. 56059–56068, 2018.
[16] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 1318–1327.
[17] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline *in vitro*," in *Proc. ICCV*, Venice, Italy, 2017, pp. 3754–3762.
[18] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. CVPR*, Denver, CO, USA, Jun. 2010, pp. 2360–2367.
[19] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 2197–2206.
[20] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical Gaussian descriptor for person re-identification," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 1363–1372.
[21] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Feb. 2009.
[22] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proc. ICML*, Corvalis, OR, USA, 2007, pp. 209–216.
[23] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *Proc. ICCV*, Kyoto, Japan, 2009, pp. 498–505.
[24] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Proc. CVPR*, Providence, RI, USA, Jun. 2012, pp. 2288–2295.
[25] A. Mignon and F. Jurie, "PCCA: A new approach for distance learning from sparse pairwise constraints," in *Proc. CVPR*, Providence, RI, USA, Jun. 2012, pp. 2666–2672.
[26] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local Fisher discriminant analysis for pedestrian re-identification," in *Proc. CVPR*, Portland, OR, USA, Jun. 2013, pp. 3318–3325.
[27] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proc. CVPR*, Portland, OR, USA, Jun. 2013, pp. 3610–3617.
[28] S. Liao and S. Z. Li, "Efficient PSD constrained asymmetric metric learning for person re-identification," in *Proc. ICCV*, Santiago, Spain, 2015, pp. 3685–3693.
[29] Q. Wang, J. Gao, and Y. Yuan, "A joint convolutional neural networks and context transfer for street scenes labeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1457–1470, May 2017.
[30] Q. Wang, J. Gao, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 230–241, Jan. 2018.

[31] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognit.*, vol. 48, no. 10, pp. 2993–3003, Oct. 2015.

[32] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 403–412.

[33] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, *arXiv:1703.07737*. [Online]. Available: https://arxiv.org/abs/1703.07737

[34] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. CVPR*, Boston, MA, USA, Jun. 2015, pp. 3908–3916.

[35] R. R. Varior, M. Haloi, and G. Wang, "Gated siamese convolutional neural network architecture for human re-identification," in *Proc. ECCV*, Amsterdam, The Netherland, 2016, pp. 791–808.

[36] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang, "Spindle net: Person re-identification with human body region guided feature decomposition and fusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1077–1085.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[38] W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in *Proc. CVPR*, Salt Lake City, UT, USA, Jun. 2018, pp. 2285–2294.

[39] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, vol. 2, Jun. 2005, pp. 60–65.

[40] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image restoration by sparse 3D transform-domain collaborative filtering," *Proc. SPIE*, vol. 6812, p. 681207, Mar. 2008.

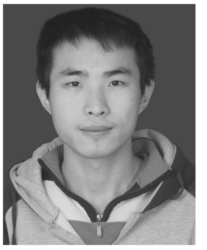
[41] S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 3587–3596.

[42] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

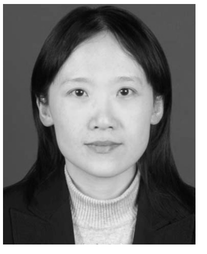
[43] S. Lefkimmiatis, "Universal denoising networks: A novel CNN architecture for image denoising," in *Proc. CVPR*, Salt Lake, UT, USA, Jun. 2018, pp. 3204–3213.

[44] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," 2018, *arXiv:1803.04189*. [Online]. Available: https://arxiv.org/abs/1803.04189

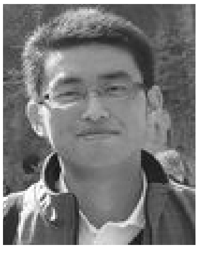
[45] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, "Non-local recurrent network for image restoration," in *Proc. NIPS*, Paris, France, 2018, pp. 1673–1682.



JIAN'AN ZHANG received the B.E. degree in information and computing science from the Ocean University of China, Qingdao, China, in 2015. He is currently pursuing the Ph.D. degree with the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.



YUAN YUAN (M'05–SM'09) is currently a Full Professor with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. She has authored or coauthored over 150 papers, including about 100 in reputable journals, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and also conference papers in CVPR, BMVC, ICIP, and ICASSP. Her current research interests include visual information processing and image/video content analysis.



QI WANG (M'15–SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science, and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.