Rethinking Illumination for Person Re-Identification: A Unified View

Suncheng Xiang[1]*, Guanjie You[2], Leqi Li[1], Mengyuan Guan[1], Ting Liu[1], Dahong Qian[1], Yuzhuo Fu[1]
[1]Shanghai Jiao Tong University, Shanghai, China
[2]National University of Defense Technology, Changsha, China
{xiangsuncheng17, gemini.my, louisa_liu, dahong.qian, yzfu}@sjtu.edu.cn
{ygjssxz, charles_iee}@163.com

Abstract

As a fundamental problem in video surveillance, person re-identification (re-ID) contributes a lot to the development of modern metro city. Recently, learning from synthetic data on re-ID task, which benefits from the popularity of synthetic data engine, has achieved remarkable performance in both supervised and unsupervised manner. However, previous researches mainly lay emphasis on employing synthetic data to achieve the state-of-the-art performance with a strong backbone, while neglects to perform quantitative studies on how visual factors affect re-ID system. To facilitate the research in this field, firstly, we manually construct a large-scale synthetic dataset named SynPerson, which has diversified human characters and distinguished attributes with accurate annotations. Secondly, we quantitatively analyze the influence of illumination on re-ID system. To our best knowledge, this is the first attempt to explicitly dissect person re-ID from the aspect of illumination on synthetic dataset. Comprehensive experiments help us have a deeper understanding of the fundamental problems in person re-ID. Furthermore, we will release SynPerson to the community, as part of efforts to alleviate the shortage of large-scale pedestrian dataset of future works[1].

1. Introduction

Person re-identification aims at matching and returning a specified probe person from a large-scale gallery set collected by different cameras at a different time, which has attracted lots of interests and attention in both academia and industry. Encouraged by the remarkable success of deep learning methods [5, 6] and the availability of re-ID datasets [20], re-ID research community has achieved significant progress during the past few years. Actually, current researches mainly concentrate on achieving satisfactory performance with large-scale datasets at the price of large amount of accurate annotations obtained by expensive human labor, while neglect to conduct comprehensive experiments to quantitatively assess the influence of attributes on person re-ID accuracy, or perform the dataset-level analysis in related image retrieval tasks. Therefore, it remains largely unknown how these factors (*e.g.* illumination) affect the re-ID system quantitatively.

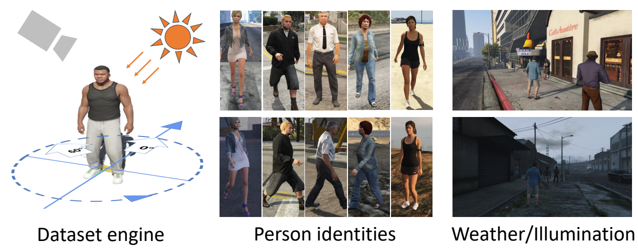


Figure 1. The procedure for building the dataset. We collect data with synthetic game engine and annotate them for every single pedestrian by analytic visual factors, *e.g.* viewpoint, weather and illumination.

Another challenge we observe is that, there has been an increasing concern over data safety and ethical issues, *e.g.* DukeMTMC-reID [7] has been taken down due to privacy problem. Some European countries already passed privacy-protecting laws [11] to prohibit the acquisition of personal data without authorization, which makes collection of large-scale datasets extremely difficult. To address this issue, many successful re-ID approaches [1, 2, 10, 12] have been proposed to take advantage of game engine to construct large-scale synthetic datasets, which can be used to pre-train or fine-tune CNN networks. In essence, it helps to provide more complete and better initialization parameters for potentially promoting the development of re-ID task. However, pre-existing synthetic datasets have limited identities and are lack of diversity, some of them might have a biased and fixed distribution of environments, which cannot satisfy the need of dissecting person re-ID from the perspective of

*Corresponding author.
[1]Our SynPerson dataset is publicly available at https://github.com/JeremyXSC/SynPerson.

Table 1. Comparison of real-world and synthetic Re-ID datasets. "#Wea" and "#Illum" denote whether the dataset has Weather and Illumination labels.

dataset		#identity	#box	#Wea	#Illum
Real	Market-1501 [20]	1,501	32,668	✘	✘
	CUHK03 [5]	1,467	14,096	✘	✘
	MSMT17 [14]	1,404	36,411	✘	✘
Synthetic	SOMAset [2]	50	100,000	✘	✘
	SyRI [1]	100	1,680,000	✘	✔
	GPR [17]	754	443,352	✔	✘
	PersonX [10]	1,266	273,456	✘	✘
	RandPerson [12]	8,000	1,801,816	✘	✘
	Unreal [19]	3,000	120,000	✘	✘
	SynPerson	**5,345**	**1,122,450**	✔	✔

illumination.

To remedy the above problems, we start from two aspects, namely data and methodology. From the data aspect, we manually construct a large-scale and diverse **Syn**thetic **Person** re-ID dataset named **SynPerson**, which is both densely annotated and visually coherent with real world. The procedure for building the dataset with game engine is demonstrated in Fig. 1. Such procedure can greatly reduce the labor of annotating multiple-attributes datasets without security and privacy concerns. Compared with existing datasets, our SynPerson has several advantages: 1) free collection and annotation; 2) larger data volume; 3) more diversified scenes and 4) high resolution. It supports future research in not only algorithm design, but also scientific discoveries how environmental factors affect the system.

From the methodology aspect, we propose to dissect a person re-ID system by quantitatively understanding the role of image illumination. To achieve this goal, three questions are considered in this paper: *(1) How does the illumination of the training set influence the retrieval? (2) How does the query illumination influence the retrieval? (3) How does true match illumination in the gallery affect retrieval?* To find the answers to these questions, we quantitatively analyze the influence of illumination with SynPerson dataset on a standard re-ID system. Typically, the goal of this work is not to achieve the-state-of-the-art performance on re-ID task, but to dissect the person re-ID system from the perspective of illumination. To this end, we just select a simple but effective baseline for re-ID evaluation. During the experiments, both the control group and experimental group are carefully designed, so as to obtain convincing scientific conclusions. To the best of our knowledge, there is no work in the existing literatures that comprehensively study the impacts of dataset illumination on re-ID system. The empirical results from our research are consistent with our intuition on the real-world application.

As a consequence, this paper makes two contributions to the community. (1) We introduce a large-scale and diverse synthetic dataset SynPerson, which contains 1,122,450 images of 5,345 manually designed identities and editable visual attributes. (2) Based on it, we dissect a person re-ID system by quantitatively understanding the role of illumination attributes. Conclusions derived from extensive experiments provide re-ID community with some useful guidance and insights for dataset building and future practical usage, helping us to have a deeper understanding of the fundamental problems in person re-ID.

2. Related Work

2.1. Learning from Synthetic Dataset

It is known that manual labelling is generally time-consuming and labour-intensive for each new (target) domain. Transfer learning [3, 15] sometimes, to some extent, works but fails to solve this problem fundamentally. More recently, leveraging synthetic data has been proved to be a useful idea to alleviate the reliance on large-scale real datasets in segmentation [8] and object detection [9]. In re-ID community, the very recent approaches [1, 2, 10, 17] incorporate this idea to further boost re-ID performance in an unsupervised manner. For example, Barbosa *et al.* [2] propose a synthetic instance dataset SOMAset, which is created by photorealistic human body generation software. Bak *et al.* [1] construct a SyRI dataset by using 100 virtual humans illuminated with multiple HDR environment maps. Xiang *et al.* [17] manually build a GPR dataset which consists of 443,352 images of 754 identities based on game engine. In addition, Sun *et al.* [10] introduce a large-scale synthetic data engine named PersonX. However, neither these synthetic datasets are intensively diversified in terms of identity, nor editable or released to the public. Moreover, current re-ID datasets lack significant diversity in the number of lighting conditions or weather modes, since they are usually limited to a relatively small number of cameras. Consequently, CNN-based models trained on these special illuminations are thus biased to the limited illumination conditions seen during training, which fails to adapt the model to unseen illuminations.

2.2. Influence of Visual Factor

Visual attribute in synthetic dataset is an important factor in image retrieval due to its high-level semantic knowledge, which could greatly bridge the gap between low-level features and high-level human recognitions. More recently, Sun *et al.* [10] try to dissect person re-ID system from the viewpoint of viewpoint. As for attribute analysis, Xiang *et al.* [16] quantitatively analyze the influence of different environment attributes on re-ID system. Additionally, Zeng *et al.* [18] also propose an illumination identity disentanglement network to dispel different scales of illuminations away while preserving individuals' identity informa-

Figure 2. Illustration of SynPerson dataset. (A): Identity: some images from the SynPerson dataset with different characters in different scenes; (B): Weather: Visual exemplars of different weather annotations; (C): Illumination: Sample pedestrians in different illuminations. The illuminations with orange color indicate representative illumination of a person, denoted as the due Weak Light (23:00∼04:00), due Middle Light (06:00∼11:00) and due Strong Light (13:00∼18:00) respectively.

tion. Although these researches provide some new insights for dataset construction, other visual factors such as illumination still remain unexplored. Motivated by this, we firstly construct a large-scale synthetic dataset SynPerson based on controllable synthetic data engine, thus dissect person re-ID system from the perspective of illumination. Departing significantly from previous objectives of attribute analysis works, our work makes an early attempt in the community to quantitatively assess the influence of illumination on re-ID system, which can provide more meaningful guidance to construct a high-quality dataset for person re-ID and other related tasks.

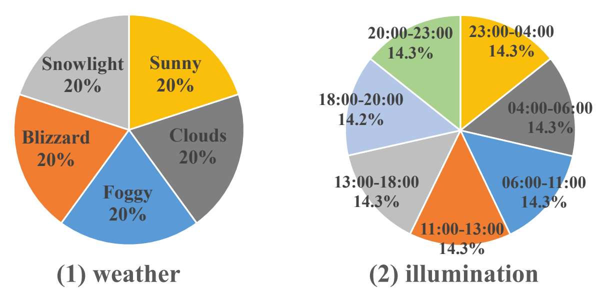


Figure 3. The statistical information of SynPerson dataset on weather and illumination distribution.

3. SynPerson Dataset

3.1. Description

Data Engine. The SynPerson dataset is built on popular game engine called the Grand Theft Auto V[2] (named "GTAV" for short). As a controllable system, it can satisfy various attributes and scene requirements. In GTAV game engine, the person models and scenes look realistic. More importantly, the values of visual variables, such as weather, illumination and viewpoint are designed to be editable, which allow the data engine to be highly flexible and extendable.

Identities. SynPerson has 5,345 hand-crafted identities including different skin colors, body forms (*e.g.*, height and weight) and hair styles. To ensure diversity, we change the human appearance and cloth styles by editing related model parameters, the clothes style in the game engine include jeans, pants, shorts, slacks, skirts, *etc*. In particular, some of these identities have a backpack, shoulder bag, glasses or hat. The motion of these characters can be set as walking, running, standing, or even having a dialogue etc. Fig. 2 (A) presents some examples of the character phototypes with customizable body parts and clothing. When not specified, all images are captured with resolution of 200×470.

3.2. Visual Factors in SynPerson

In this work, SynPerson dataset is featured by editable environmental factors such as weather, illumination, viewpoint and background. More detailed information of these factors are described below.

Weather. In our large-scale synthetic data SynPerson there are 5 different types of weather, *e.g.*, Sunny, Clouds, Foggy, Blizzard and Snowlight. Parameters like degree and intensity can be modified for each weather type. By editing the values of these terms, various kinds of weather can be created. Some exemplars of synthetic scenes in different weather distributions from the proposed SynPerson dataset

[2] http://www.dev-c.com/gtav/

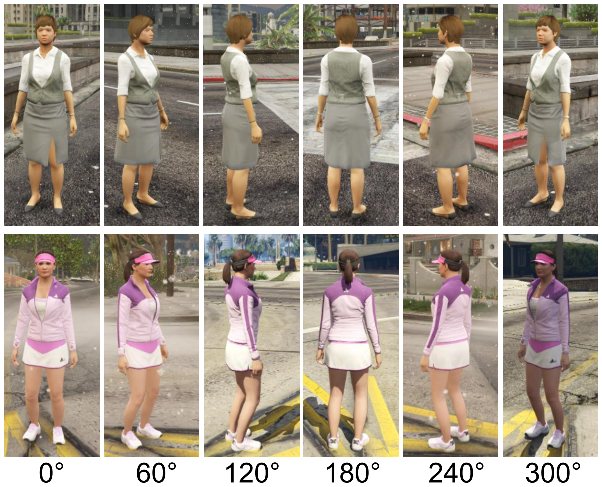


Figure 4. Sample images from the SynPerson dataset under a specified viewpoints, which is sampled every 60° from 0° to 300° (6 different viewpoints in total).

are depicted in Fig. 2 (B). Fig. 3 also illustrates the statistical information of SynPerson on weather attribute.

Illumination. Illumination can be obtained at different time in whole day. For one specific pedestrian image, we also provide its capturing time in 24 hours. In this paper, we introduce a new synthetic dataset that contains 7 illumination conditions. The exemplars of different time distribution from the proposed SynPerson dataset is illustrated in Fig. 2 (C), *e.g.*, "09∼12" denotes the time period during "9:00∼12:00" in 24 hours a day. Intuitively, we classify these illumination distributions into three categories: **Weak Light:** "20:00∼06:00"; **Middle Light:** "06:00∼11:00" & "18:00∼20:00"; **Strong Light:** "11:00∼18:00". It is worth mentioning that the subset of middle illumination have large variation due to its discontinuous time period, which will bring additional difficulties during retrieval. Detailed statistics of illumination distribution are elaborated in Fig. 3.

Viewpoint. The image examples in SynPerson dataset under specified viewpoints follow a uniform distribution. Those images are sampled during normal talking or walking. Specifically, as depicted in Fig. 4, a person image is sampled every 60° from 0° to 300° (6 different viewpoints in total). Each view has 1 image, so each person has 6 images. The entire SynPerson has 5,345 (identities) × 5 (weathers) × 7 (illuminations) × 6 (viewpoints) = 1,122,450 images.

Backgound. Currently SynPerson has several different backgrounds, including urban area and wild area. In each background, a person moves freely in arbitrary directions, exhibiting arbitrary viewpoints relative to the camera. Notably, adopting more types of scenes close to target domain seems to have a positive influence on the re-ID performance. Specifically, pedestrian of different backgrounds in SynPerson is illustrated in Fig. 8. We believe SynPerson will be a useful tool for the community and scientific analysis.

3.3. Re-ID baseline and Evaluation Protocol

Re-ID baseline. In this paper, we follow the training procedure in [6] and adopt a widely used open-source[3] as our standard baseline. Specially, we employ global features provided by vanilla backbone ResNet-50 to perform feature learning with cross-entropy loss and triplet loss [4], the batch size of training samples is set as 64. It is worth noting that we only modify the output dimension of the last fully-connected layer to the number of training identities. During testing, we extract the 2,048-dim pool-5 vector for retrieval under the commonly used Euclidean distance.

Evaluation Protocol. In order to evaluate the impact of illumination on re-ID task, we randomly divide SynPerson into training set and testing set. Similar with the previous datasets [5, 20], we set the training and testing ratio nearly close to 1:1. Finally, the training set contains 43,920 bounding boxes of 2,440 identities, and the testing set contains 52,290 bounding boxes of 2,905 identities. From the testing set, 17,430 bounding boxes are randomly selected as query images and the other 34,860 bounding boxes are used as gallery images. As for the evaluation metric, we adopt mean Average Precision (mAP) and Cumulative Matching Characteristics (CMC) at rank-1 and rank-5 for evaluation on re-ID task.

3.4. Benchmarking Validation

Aiming to validate the SynPerson is indicative of the real-world, we employ a standard re-ID backbone with different losses to perform evaluation on synthetic dataset, so that conclusion derived from SynPerson can be applicable to the real-world.

As illustrated in Fig. 5, the performance trend on standard re-ID backbone with two different loss functions (*e.g.*, cross-entropy loss and triplet loss) is similar between SynPerson and real-world dataset. On Market-1501 dataset, for example, the cross-entropy loss & triplet loss have the best accuracy, and cross-entropy loss shows a slight inferiority of mAP performance for its limited discriminative learning ability. When tested on the SynPerson dataset and its corresponding subsets with different illumination settings (*e.g.* Syn_weak, Syn_middle, Syn_strong), we can also observe the same performance trends, which suggests that conclusions derived from SynPerson is meaningful and feasible in practical application. Additionally, the re-ID accuracy on SynPerson and its subsets are relatively high compared to the real-world dataset, this is due to the high-quality of our SynPerson, which tend to have consistent backgrounds and uniform attribute distribution. So SynPerson dataset and its

[3] https://github.com/Cysu/open-reid

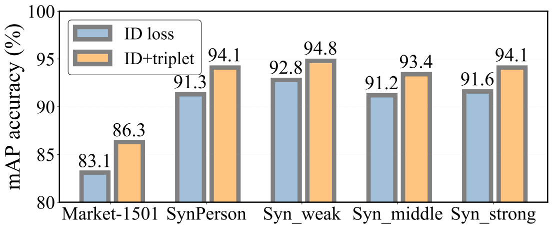


Figure 5. Performance of re-ID system on real-world datasets and SynPerson dataset with different loss functions, *i.e.* cross-entropy loss (ID loss) and triplet loss. "Syn", "Syn_weak", "Syn_middle", "Syn_strong" represent original SynPerson dataset, its corresponding subset with weak, middle and strong illumination settings respectively.

subsets are thus ideal ones for studying the impact of illumination.

4. Evaluation of Illumination

In this section, we evaluate the impacts of illumination on re-ID system. All training and testing datasets are based on our SynPerson dataset. To go even further, three questions will be investigated in the following subsections: how does the illumination in (1) the training set, (2) the query set, and (3) the gallery set affect the re-ID accuracy? For a clearer understanding, we show both Figures and Tables in this section.

4.1. How Do Illumination Distributions in the Training Set Affect Model Learning?

In this section, we will quantitatively investigate the influences of illumination attributes on re-ID system, the conclusion derived from this dataset can be of great value in practice, especially when building high-quality datasets.

Experiment Design. Initially, the subsets contain all the illumination distributions for the training and testing parts. That is, a person has 6 images under a specific illumination and weather condition. In this section, to study the influence of different illuminations in the training set, we remove other unrelated illumination conditions during training. Consequently, we design five experimental groups with different illumination settings, as shown in Table 2, where the number of images used for training is the same, which can exclude the influence of the number of pictures on the re-ID model during the training.

Results Analysis. In this section, we evaluate the model trained on subsets with different illuminations of SynPerson, then summarize the key experimental results in Table 2. From these results, we have several observations as follows:

First, person with a Weak or Strong illumination may deteriorate the model performance in some degree. For example, we can achieve a remarkable mAP accuracy of 78.0% when employing the Middle Light as training set. However, the performance of mAP accuracy will dramatically drop to 63.2% or 60.9% when using the Weak Light or Strong Light as training set respectively. We suspect that this is due to: (1) Training set trends to be diversified and have larger illumination variation when adopting Middle Light as training set; (2) There exists a great illumination gap between training set and testing set during the testing period, where testing set trends to have a uniform and middle distribution in terms of illumination.

Table 2. Evaluation performance of baseline model trained on SynPerson in terms of different kinds of illumination. "W","M" and "S" denote Weak Light, Middle Light and Strong Light, respectively.

Experiment	Bboxes	Components			SynPerson		
		W	M	S	mAP	rank-1	rank-5
Group 1	43,930	✓			63.2	93.2	98.6
Group 2	43,930		✓		78.0	96.1	99.3
Group 3	43,930			✓	60.9	95.0	98.9
Group 4	43,930	✓	✓		90.3	97.4	99.5
Group 5	43,930	✓	✓	✓	93.8	98.5	99.8

Second, using more illumination categories is always beneficial to the re-ID system. Compared with only one kind of illumination, using three kinds of illumination can significantly boost the mAP performance to 93.8%, leading by **+30.6%**, **+15.8%** and **+32.9%** improvement comparing to Weak, Middle and Strong Light respectively. This observation is intuitive because training samples in diversified illumination can always increase the generalization capability and robustness of re-ID model.

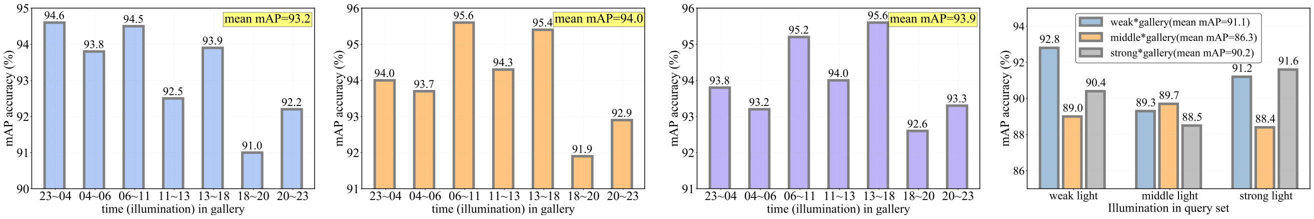
4.2. How Does Query Illumination Affect Retrieval?

In this section, we continuously study how query illumination influences the re-ID performance results.

Experiment Design. We train a model on the original training set comprised of all illumination conditions, then modify the query illumination to see its effect during testing. Specifically, the illumination of a probe image can be set to the due Weak Light (23:00$\sim$04:00), due Middle Light (06:00$\sim$11:00) and due Strong Light (13:00$\sim$18:00) respectively to indicate representative light environments. During the retrieval, the testing set can be randomly set between 00:00$\sim$24:00 all through the day.

Results Analysis. Fig. 6 (a-c) presents the results obtained by the above query and gallery images.

First, when the illumination of the true match is similar to the query, the highest re-ID accuracy can be achieved. For example, as shown in Fig. 6 (a), the maximum mAP scores of Weak Light (23:00$\sim$04:00) in queries correspond exactly to the due Weak Light (23:00$\sim$04:00) true match in the gallery, which reaches a remarkable performance of



(a) Weak light in query set (b) Middle light in query set (c) Strong light in query set (d) Impact of light disparity

Figure 6. (a-c) Impacts of query illumination on re-ID system performance with SynPerson dataset. Three illuminations are evaluated, *i.e.*, due Weak Light (23:00∼04:00), due Middle Light (06:00∼11:00) and due Strong Light (13:00∼18:00). The value in the yellow box represents the average mAP accuracy of each query illumination. (d) The impacts of illumination disparity between a query and gallery. For training, we use the original training sets which have balanced illumination distribution.

94.6%. Under the condition of same illumination, the query and true match are different only in terms of viewpoint and background. This indicates that illumination differences between two to-be-matched images cause performance degradation.

Second, queries of the due Middle Light and due Strong Light lead to a higher mAP accuracy than queries of the due Weak Light. For example, we can only obtain an mean mAP performance of **93.2%** with due Weak Light illumination, while the mAP accuracy of the due Middle Light and the due Strong Light can reach **94.0%** and **93.9%** respectively, which demonstrates that a high quality illumination condition of query set has noticeable superiority for re-ID task.

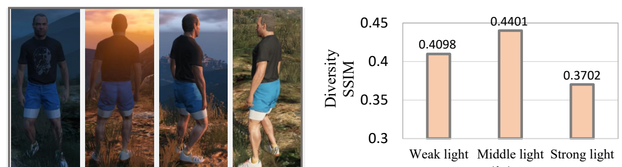
4.3. How Do True Match illumination in the Gallery Affect Retrieval?

Finally, we study how the gallery illumination distribution affects re-ID accuracy, especially when there exists the illumination disparity between the query and its true matches.

Experiment Design. SynPerson dataset has three main kinds of illumination, *i.e.* Weak Light: "20:00∼06:00"; Middle Light: "06:00∼11:00" & "18:00∼20:00"; Strong Light: "11:00∼18:00". We design several experiments to see how the true match illumination in the gallery affect retrieval performance on re-ID model.

Results Analysis. According to the results in Fig. 6 (d), we have several valuable constructions.

The major observation is that if true matches with same illumination are not in the gallery set, there will be a non-trivial performance drop. In other words, if true march in the gallery has large illumination disparity with the query, the retrieval performance will be negatively affected. For instance, as shown in the Fig. 6 (d), when the illumination of testing set is same with the query set, we can obtain a mAP performance of **92.8%**, **89.7%** and **91.6%** with weak, middle and strong illumination, respectively. However, there will be a catastrophic performance drop (nearly **3.8%**↓) when gallery is set as Middle level illumination while query is set as Weak level illumination. Same conclusion can also be drawn when query images are set as Middle Light or Strong Light separately.



(a) (b)

Figure 7. (a) Variation of SynPerson with Middle Light setting; (b) SSIM value of samples in different illumination settings.

Moreover, the accuracy decrease caused by illumination disparity between a query and its true match in gallery becomes more obvious especially when the illumination variation becomes more challenging. For example, as depicted in the Fig. 6 (d), the mAP drop of query with Weak Light on the SynPerson is almost **3**× as large as the performance decline for the query with Middle illumination (mAP **3.8%** vs. **1.2%**). This confirms that illumination gap between query and gallery is still a critical factor for boosting the re-ID performance to a new level. In essence, this conclusion, which is drawn from the massive experiments, can give the re-ID community some insights for future research.

Discussion. As depicted in Fig. 6 (d), we found an interesting phenomenon that the average performance of gallery with Middle Light (**86.3%**) is slightly inferior to the gallery with Weak Light (**91.1%**) or Strong Light (**90.2%**). To go even further, we gave an explanation about this phenomenon from two aspects:

From the *Qualitative* perspective, as shown in Fig. 7 (a), it can be easily noticed that there exists a large illumination variation among subset of Middle Light due to its discontinuous time period ("06:00∼11:00" & "18:00∼20:00"), which will undoubtedly bring additional difficulties or burdens for re-ID model during retrieval.

From the *Quantitative* perspective, we adopt Structural SIMilarity (SSIM) [13] to measure the intra-class similarity among subsets with different illuminations[2]. According to Fig. 7 (b), the dataset with Middle Light has higher scores in terms of SSIM, suggesting the larger variation of subset with Middle Light when comparing to Weak or Strong Light. To this end, the discriminability of re-ID model will be negatively affected in a scenario with a larger illumination variation, especially for the gallery set. These challenges warrant further research and consideration when deploying re-ID model in real scenarios.

5. Conclusion

In this paper, we make two contributions to the re-ID community, which takes a big step forward from developing new technologies to explore new discoveries. First, we manually construct a large-scale synthetic dataset named SynPerson, which has diversified characters and distinguished attributes. Importantly, both training and testing set are also carefully split to be indicative of the real-world. Based on it, we conduct extensive experiments to quantitatively assess the influences of *Illumination* on re-ID accuracy, constructive insights are derived to help us take a closer look at fundamental problems in person re-ID. In the future, we will further explore the influences of these visual factors on other human-related tasks, such as pose estimation and human part segmentation.

Broader Impacts It is known that re-ID technologies rely heavily on non-consensual surveillance data, which may lead to the infringement of person's privacy. Motivated by this, we carefully construct the SynPerson dataset without ethics issues, conclusion derived from experiments can help advance the research of person re-ID, thus promoting the development of smart city and security systems in the future metropolises. Furthermore, we should be cautious of the annotation procedure of large scale datasets towards the social implications. Also, note that the demographic makeup of the datasets used is not representative of the broader population.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China under Grant No.61977045 and No.81974276. The authors would like to thank the anonymous reviewers for their valuable suggestions and constructive criticisms.

References

[1] Slawomir Bak, Peter Carr, and Jean-Francois Lalonde. Domain adaptation through synthesis for unsupervised person re-identification. In *Proceedings of the European conference on computer vision (ECCV)*, pages 189–205, 2018. 1, 2

[2] Igor Barros Barbosa, Marco Cristani, Barbara Caputo, et al. Looking beyond appearances: Synthetic training data for deep cnns in re-identification. *Computer Vision and Image Understanding*, 167:50–62, 2018. 1, 2

[3] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 994–1003, 2018. 2

[4] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 4

[5] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 152–159, 2014. 1, 2, 4

[6] Hao Luo, Youzhi Gu, Xingyu Liao, et al. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1, 4

[7] Ergys Ristani, Francesco Solera, Roger Zou, et al. Performance measures and a data set for multi-target, multi-camera tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 17–35. Springer, 2016. 1

[8] Swami Sankaranarayanan, Yogesh Balaji, Arpit Jain, et al. Learning from synthetic data: Addressing domain shift for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3752–3761, 2018. 2

[9] Tobias Scheck, Roman Seidel, and Gangolf Hirtz. Learning from theodore: A synthetic omnidirectional top-view indoor dataset for deep transfer learning. In *WACV*, pages 943–952, 2020. 2

[10] Xiaoxiao Sun and Liang Zheng. Dissecting person re-identification from the viewpoint of viewpoint. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 608–617, 2019. 1, 2

[11] Paul Voigt and Axel Von dem Bussche. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10:3152676, 2017. 1

[12] Yanan Wang, Shengcai Liao, and Ling Shao. Surpassing real-world source training data: Random 3d characters for generalizable person re-identification. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 3422–3430, 2020. 1, 2

[13] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 7

[14] Longhui Wei, Shiliang Zhang, Wen Gao, et al. Person transfer gan to bridge domain gap for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 79–88, 2018. 2

[2] In this paper, a higher SSIM score means larger variation among a specific dataset, and vice versa.

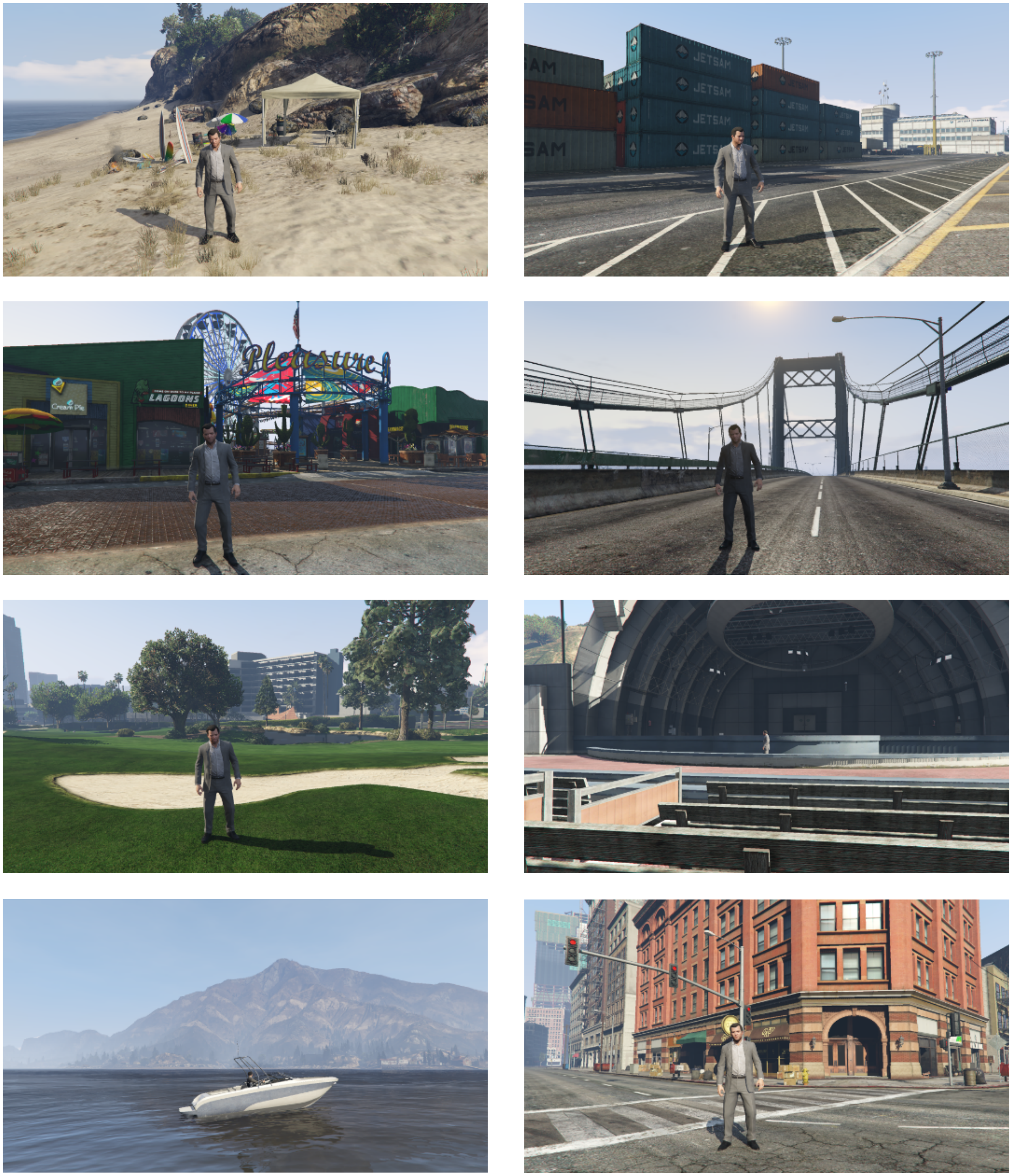


Figure 8. Image examples with same character in different backgrounds. To maintain a high diversity of generated dataset, a pedestrian model is ordered to do a random action in the real-world scenarios in each background.

[15] Suncheng Xiang, Yuzhuo Fu, and Ting Liu. Deep unsupervised progressive learning for distant domain adaptation. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 903–908. IEEE, 2019. 2

[16] Suncheng Xiang, Yuzhuo Fu, Guanjie You, et al. Taking a closer look at synthesis: Fine-grained attribute analysis for person re-identification. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3765–3769. IEEE, 2021. 2

[17] Suncheng Xiang, Yuzhuo Fu, Guanjie You, and Ting Liu. Unsupervised domain adaptation through synthesis for person re-identification. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. 2

[18] Zelong Zeng, Zhixiang Wang, Zheng Wang, Yinqiang Zheng, Yung-Yu Chuang, and Shin'ichi Satoh. Illumination-adaptive person re-identification. *IEEE Transactions on Multimedia*, 22(12):3064–3074, 2020. 2

[19] Tianyu Zhang, Lingxi Xie, Longhui Wei, et al. Unrealperson: An adaptive pipeline towards costless person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11506–11515, 2021. 2

[20] Liang Zheng, Liyue Shen, Lu Tian, et al. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1116–1124, 2015. 1, 2, 4