GENERATED DATA WITH SPARSE REGULARIZED MULTI-PSEUDO LABEL FOR PERSON RE-IDENTIFICATION

Liqin Huang, Member, IEEE, Qingqing Yang, Junyi Wu, Yan Huang, Qiang Wu, Senior Member, IEEE, and Jingsong Xu⋆

Abstract—Recently, Generative Adversarial Network (GAN) has been adopted to improve person re-identification (person re-ID) performance through data augmentation. However, directly leveraging generated data to train a re-ID model may easily lead to over-fitting issue on these extra data and decrease the generalisability of model to learn true ID-related features from real data. Inspired by the previous approach which assigns multi-pseudo labels on the generated data to reduce the risk of over-fitting, we propose to take sparse regularization into consideration. We attempt to further improve the performance of current re-ID models by using the unlabeled generated data. The proposed Sparse Regularized Multi-Pseudo Label (SRMpL) can effectively prevent the over-fitting issue when some larger weights are assigned to the generated data. Our experiments are carried out on two publicly available person re-ID datasets (e.g., Market-1501 and DukeMTMC-reID). Compared with existing unlabeled generated data re-ID solutions, our approach achieves competitive performance. Two classical re-ID models are used to verify our sparse regularization label on generated data, i.e., an ID-embedding network and a two-stream network.

Index Terms—Person re-identification, Generated data, Sparse pseudo label

I. Introduction

PERSON re-identification (re-ID) is an important research topic in the field of computer vision, it aims to identify the same person in the view of non-overlapping cameras [1], [2], [3], [4], [5]. Peron re-ID is a challenging task due to variations of poses, occlusions, illuminations, viewpoints, and background clutters in training datasets [6], [7], [8], [9], [10].

At present, Convolutional Neural Network (CNN) has achieved promising performance in person re-ID [11], [12], [13]. A majority of existing methods treat person re-ID as a multi-classification task during the training process and learn a discriminative model through large amount of training data. These CNN-based methods normally require a large number of labeled data for training. However, it is challenging to do data augmentation with manpower.

L. Huang, Q. Yang and J. Wu is with Fuzhou University, Fuzhou, FJ 350108 China.(e-mail: hlq@fzu.edu.cn; n181127051@fzu.edu.cn; wujunyi627@gmail.com).

Y. Huang, Q. Wu and J. Xu is with the Faculty of Engineering and Information Technology, University of Technology Sydney, NSW, Australia (e-mail: yan.huang-3@student.uts.edu.au; qiang.wu@uts.edu.au; jingsong.xu@uts.edu.au).

⋆ means the corresponding author.

This work was supported by Major Science and Technology Projects in Fujian, China (2018H0018).

Fig. 1: Image samples of the Market-1501 and Market-DCGAN. The left is real data, the right is generated data.

With the development of deep learning network, in 2014's Generative Adversarial Network (GAN) [14] was proposed to generate data. In existing person re-ID works by using GAN, it can be divided into the following two categories: 1) [15], [16] use the conditional GAN to generate labeled data that is closer to the real data. 2) [17], [18] use the unconditional GAN to generate unlabeled data with an appropriate pseudo-label strategy to improve person re-ID performance. The first category mainly aims at modifying GAN to make the generated data more real. The second category is mainly proposing better pseudo-label to the unlabeled generated data. Our work focuses on the second category and proposes a better pseudo-label strategy.

At present, a variety of methods are proposed to assign labels to the generated data [17], [18], [19], [20]. For instance, all-in-one [19] assigns a new label to the data generated by the GAN so that a new class of data can be added in the training data. Specifically, the original training set has K classes and the way of all-in-one is to treat all generated data as (K+1)-th class. But the all-in-one method simply treats the generated data as a new class without considering the relationship between generated data and real data. Unlike the all-in-one method, One-hot pseudo approach [20] assigns virtual labels to generated data without extra class to address the disadvantage of all-in-one. The one-hot pseudo label selects the maximum prediction probability of the pre-defined training classes as virtual label for the generated data. However, the one-hot pseudo label may lead to over-fitting when the a specified label is assigned to some generated data. In addition,

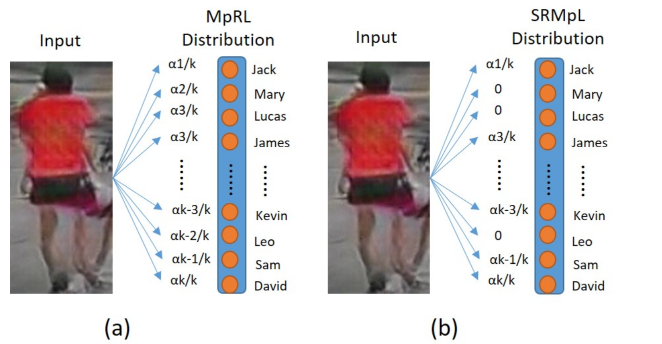

Fig. 2: The label distributions of generated data. (a) MpRL considers different weights over all pre-defined training classes. (b) the proposed SRMpL uses sparsity to prevent over-fitting issue during the training process.

[17] proposes the LSRO method. The LSRO method assigns a uniform label distribution to all generated data. However, the uniform distribution is equally applied the contributions of all pre-defined classes to each generated image, ignoring the contributions of different pre-defined classes. To handle this issue, [18] presents Multi-Pseudo Regularized Label (MpRL) which dynamically assigns different pseudo-label to generated data.

Unlike LSRO, MpRL argues that generated data cannot be considered equally to all the pre-defined training classes. Therefore, MpRL takes different contributions of each pre-defined class to the generated data into account. The generator of GAN generates images based on the feature information (e.g., color and texture) of the real data. For example, if most pedestrians in the real data are wearing dark clothing (refer to Fig. 1), the generator has a great chance to generate many dark clothing pedestrians (refer to Fig. 1) based on this dark clothing distribution. Due to a large amount of dark color, the MpRL may assign large weights to pre-defined classes with dark color clothes. In this situation, bright color information may be ignored during the training process, which decreases the generalization ability of training model. In this work, we make the following contributions:

1) We propose the Sparse Regularized Multi-Pseudo Label (SRMpL) for generated data. Unlike MpRL, our proposed SRMpL can avoid over-fitting to certain characteristic attributes during the training process. That is, the proposed SRMpL can encourage the model to learn more generalized information.

2) Qualitative analyses are given to SRMpL. Also, comprehensive quantitative evaluations are carried out to verify the performance of the proposed SRMpL method on two large person re-ID datasets by adapting two kinds of classical person re-ID models.

II. The proposed Sparse Regularized Multi-Pseudo Label

In this section, we first review MpRL [18] for person re-ID. Then, we first describe our proposed SRMpL method and explain why SRMpL works better.

A. MpRL for Person Re-ID Revisit

Given an unlabeled generated image, MpRL proposes that the contribution of different pre-defined training classes in the original training set are different. To reflect such different contributions, MpRL records the weights through a dictionary α. Due to the LSRO may lead to ambiguous predictions in training [17], the MpRL proposes that each generated data needs to be dynamically assigned different virtual labels. The dictionary α in MpRL can show the likelihood of the affiliation of each generated data to all the pre-defined training classes. Given a generated image I_i, its label distribution is defined as follows:

$$q_{MpRL}(k) = \frac{\alpha_k}{K}, \tag{1}$$

where α_k represents the weights of k-th pre-defined training class in the dictionary α. MpRL first formulates the set of predicted probabilities $p(X)$ of a generated image over K pre-defined training classes, then all elements in $p(X)$ are sorted and saved to $p_s(X)$. Finally, a_k is obtained by taking the corresponding index of $p(X_k)$ in the set of $p_s(X)$. MpRL defined the cross-entropy loss as follow:

$$l_{MpRL} = -(1-y)\log\left(p\left(X_c\right)\right) - y\cdot\lambda\cdot\sigma\sum_{k=1}^{K}\frac{\alpha_k}{K}\cdot\log\left(p\left(X_k\right)\right), \tag{2}$$

where c represents the ground-truth label of a real image, λ is the parameter for the trade-off between losses of generated and real data, σ is a normalization factor.

B. The proposed Sparse Regularized Multi-Pseudo Label

As can be seen from Fig. 1, When real image samples are dominated by some characteristics (e.g., dark color clothes), a certain number of generated images may obey the distribution of such characteristics. The MpRL may encourage the model to learn feature information of such characteristics constantly. Consequently, the model may not learn other information existing in the training data generally (e.g., some bright clothes). In this situation, when two different persons wear similar bright clothes, the model may not learn such information to identify different persons.

We propose the SRMpL method that utilizes data sparsity to solve the above-mentioned problem. For the generated data, we randomly zero out the likelihood of existing pre-defined training classes instead of using all weights from pre-defined classes. As shown in Fig. 2, we introduce the sparsity and randomly zero out partial weights to avoid the issue of over-fitting in the model training. In this way, the re-ID model trained with our SRMpL can learn more generalized ID features to help identify different persons. The label distribution of our SRMpL can be defined as follows:

$$q_{SRMpL}(k) = \begin{cases} \frac{\alpha_k}{K} & k \in [1, K], k \notin S \\ 0 & k \in S \end{cases}, \tag{3}$$

where K is the number of classes, S represents the index set of weights that are randomly set to 0. We denote the

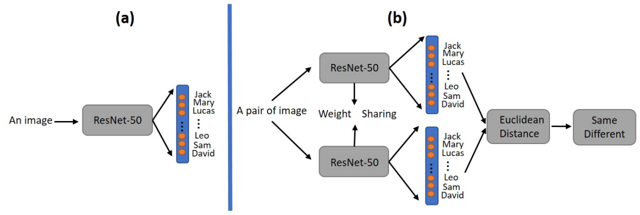

Fig. 3: The architectures of Identif Network (a) and Two-Stream Network (b).

sparse ratio as β in SRMpL, $\mathbb{N}(S)$ represents the number of set S. Thus, $\mathbb{N}(S) = \beta K$.

The loss function of using our SRMpL can be defined as below:

$$l_{SRMpL} = -(1-y)\log\left(p\left(X_c\right)\right) \\ -y\cdot\lambda\cdot\sigma\sum_{k\in[1,K],k\notin S}^{K}\frac{\alpha_k}{K}\cdot\log\left(p\left(X_k\right)\right). \tag{4}$$

where $p\left(X_c\right)\in(0,1)$ is the softmax predicted probability of X_c belonging to the pre-defined training class c.

The sparse ratio that we choose has the following advantages. First, it can reset some weights of the generated data to zero with a certain probability, so as to solve the over-fitting issue caused by MpRL. Our SRMpL uses sparsity to set some weights to zero randomly, which is similar to the dropout strategy of neural networks. Second, since the generated data are normally based on the most significant characteristic of real data. If the corresponding weights are relatively large, the model may focus on learning this characteristic in the generated data, but is hard to explore more generalized information. Our SRMpL method can reduce the some weights on generated data. Particularly, when the proposed SRMpL resets some larger weights to 0, it can encourage the network to learn more generalized information rather than some significant information. Through this to improve the generalisability of re-ID CNN models.

III. EXPERIMENTAL RESULTS

A. Datasets

Market-1501 [21] contains 32,668 images of 1,501 identities collected by 6 cameras. The training set contains 12,936 images of 751 identities, and the testing set contains 19,732 images of 750 identities. We follow the standard testing setting in [21].

DukeMTMC-reID [22] contains 36,411 images of 1,812 identities collected by 8 cameras. 16,522 images of 702 identities are utilized for training. The testing set contains 702 identities, 2,228 query images and 17,661 gallery images.

For a fair comparison, our training strategy and parameter setting is the same as MpRL [18]. After 20 epochs when the model becomes relatively stable, we assign SRMpL to generated data. That is, in Eq.4 y = 0, and until after 20 epochs, it is set to 1. Also, the loss is set to 0.1 and 1 for the generated and real data respectively. Therefore, under this training strategy, λ is set to 0.1 and σ is set to $\frac{2}{1+K}$ in Eq.4,. In addition, as shown in the Fig. 3, we use the same network architectures (i.e., Identif Network [23] and Two-Stream Network [24]) as MpRL [18].

We analyze the influence of different sparse ratio on dMpRL-II of MpRL with two person re-ID datasets. The dMpRL has been evaluated using the Identif Network [23] and the two-stream Network [24]. Therefore, we also evaluate our SRMpL using the Identif Network and the two-stream Network. We use the DCGAN model [30] to follow the same setting in [18] for a fair comparison. All the generated data are resized to 256×256 and are used to train CNN models with the proposed SRMpL. We choose 24,000 generated data in experiment which follow the best performance setting of MpRL in [18].

Our proposed SRMpL is evaluated on Market-1501 [21] and DukeMTMC-reID [22]. The cumulative matching characteristics (CMC) and mean average precision (mAP) are used as the performance evaluation protocol. Compared with the reported results in [18], we can see that our proposed SRMpL can further improve the performance of the two person re-ID models by using unlabeled generated data.

TABLE I: The performance of using different sparse ratio β on Market-1501

Identif			Two-stream		
β	Rank-1	mAP	β	Rank-1	mAP
0.1	80.65	59.78	0.1	85.86	67.52
0.2	80.82	59.96	0.2	85.98	67.72
0.3	81.20	59.25	0.3	86.50	68.21
0.4	81.52	59.56	0.4	86.67	68.34
0.5	81.22	59.30	0.5	86.57	68.22
0.6	80.89	59.05	0.6	86.31	67.95
0.7	80.58	58.74	0.7	86.01	67.72
0.8	79.55	57.94	0.8	84.83	66.71
0.9	79.11	57.56	0.9	84.41	66.34

TABLE II: The performance of using different sparse ratio β on DukeMTMC-reID

Identif			Two-stream		
β	Rank-1	mAP	β	Rank-1	mAP
0.1	68.58	48.87	0.1	77.07	58.70
0.2	68.86	49.07	0.2	77.39	58.97
0.3	69.16	49.43	0.3	77.52	59.15
0.4	69.50	49.74	0.4	77.70	59.34
0.5	69.12	49.30	0.5	77.62	59.18
0.6	68.88	49.03	0.6	77.41	59.07
0.7	68.49	48.76	0.7	77.02	58.74
0.8	67.46	47.91	0.8	75.09	57.78
0.9	66.92	47.34	0.9	75.52	57.35

B. Comparison with different sparse ratio

In the experiment on the two networks, we select different sparse ratio to verify the impact on person re-ID performance. It can be seen in Table I and II that

TABLE III: Two performance by using different sparse strategies on Market-1501 and DukeMTMC-reID

	Market-1501				DukeMTMC-reID			
	Identif		Two-Stream		Identif		Two-Stream	
Mode	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
max	81.85	59.78	87.03	68.95	69.81	49.83	78.01	58.59
random	81.52	59.56	86.67	68.34	69.50	49.74	77.70	59.34
min	79.59	58.01	84.81	66.68	67.32	48.95	75.88	57.74

when the sparse ratio is around 0.4 SRMpL can achieve the best performance. Through this experiment, it can be observed that when a appropriate sparse ratio is set, the over-fitting issue can be reduced, so that the model can obtain better performance. When the sparse ratio is large, the relationship between the generated data and the real data cannot be fully considered. In this case, the generated images may become to noise data for model training. In this situation, the model may not use the generated data properly, which has a negative impact on the re-ID performance.

C. Comparison with different sparse strategies

In Table III, we compare the effectiveness of different sparse strategies, and the sparse ratios are all set as 0.4. We use three different sparse strategies: maximum (max), random and minimum (min). We first sort the weights of pre-defined classes from large to small. The maximum mode is to zero out the top 40% weights, the minimum mode is zero out the smallest 40% weights. We found that when we use the maximum sparse strategy, the performance improvement is limited. It can be observed that the model is over-fitting on the pre-defined classes with a large weight. When the weights are set to 0, the model could learn more information and improve its robustness.

D. Comparison with MpRL

Due to the MpRL may cause the model to be easily over-fitting on the generated data. Also, the MpRL may cause the model to be unable to explore more fine-grained features of the real data. Our SRMpL can deal with these problems better than MpRL by using the sparsity. By setting the weights of some pre-defined classes to zero with a certain probability, such a practice can be regarded as equivalent to adding noise to these weights, which can reduce the problem of over-fitting to some extent. When the sparse ratio is used to set partial weights to zero, the model can obtain more fine-grained features from the real data to update the parameters of a re-ID network .

In Table IV and V, we compare performance with MpRL on Market-1501 [21] and DukeMTMC-reID [22]. For the Identif Network, our SRMpL increases Rank-1 accuracy from 80.37% (68.24%) of dMpRL-II to 81.52% (69.50%) on Market-1501 (DukeMTMC-reID). For the two-stream network, our SRMpL increases the performance of model from 85.75% (76.81%) to 86.67% (77.70%) on Market-1501 (DukeMTMC-reID).

TABLE IV: Comparison with MpRL on Market-1501

	Identif		Two-stream	
Method	Rank-1	mAP	Rank-1	mAP
dMpRL-II	80.37	58.59	85.75	67.53
SRMpL	81.52	59.56	86.67	68.34
Improvement	+1.15	+0.97	+0.92	+0.81

TABLE V: Comparison with MpRL on DukeMTMC-reID

	Identif		Two-stream	
Method	Rank-1	mAP	Rank-1	mAP
dMpRL-II	68.24	48.58	76.81	58.56
SRMpL	69.50	49.74	77.70	59.34
Improvement	+1.26	+1.16	+0.89	+0.78

E. Comparison with some state-of-the-art methods

We also compare our SRMpL method with some other state-of-the-art methods. Table VI shows that the performance of the two-stream network with SRMpL can defeat the performance of MSCAN [25], ResNet+OIM [26], SSM[29], Triplet Loss [27], JLML[28], and SVDNet[11].

TABLE VI: Comparison with state-of-the-art methods

Method	Market-1501		DukeMTMC-re-ID	
	Rank-1	mAP	Rank-1	mAP
MSCAN [25]	80.31	57.53	-	-
ResNet+OIM [26]	82.10	-	68.10	-
SSM [29]	82.21	68.80	-	-
SVDNet [11]	82.30	62.10	76.70	56.80
Triplet Loss [27]	84.90	69.10	-	-
JLML [28]	85.10	65.50	-	-
Identif [23]	74.08	52.68	61.94	42.20
Identif+MpRL [18]	80.37	58.59	68.24	48.58
Identif+SRMpL	81.52	59.56	69.50	49.74
Two-stream [24]	81.83	64.09	72.62	51.40
Two-stream+MpRL [18]	85.75	67.53	76.81	58.56
Two-stream+SRMpL	86.67	68.34	77.70	59.34

IV. Conclusion

In this paper, we propose a new virtual label SRMpL for the generated data by unconditional GAN. Due to the MpRL is prone to over-fitting, it may decrease the ability to learn information from real data when some kinds of generated data are assigned relatively large weights. We deal with this problem by introducing the sparsity in SRMpL. We combine real data and generated data into model training to improve the robustness of re-ID models. We use the Identif network and the Two-stream network to verify the effectiveness of our proposed SRMpL. Our experiments show that the generated data can effectively improve the performance of the two CNNs re-ID models with the proposed SRMpL. In addition, compared with the previous state-of-the-art method MpRL [18], SRMpL can achieve better performance in a fair comparison.

References

[1] J. Lu, and Y. he, and T. Liu, and X. chen, “Centralized and Clustered Features for Person Re-Identification,” in IEEE Signal Process. Lett., vol. 26, no. 6, pp. 933-937, Apr. 2019.

[2] M. Dikmen, E. Akbas, T. Huang, and N. Ahuja, “Pedestrian recognition with a learned metric,” in ACCV, NZ, 2010. pp. 501-512.

[3] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. W Bang, “Pedestrian recognition with a learned metric,” in ECCV, GER, 2018. pp. 480-496.

[4] 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 CVPR, US, 2017. pp. 1077-1085.

[5] C. Luo, Y. Chen, N. Wang, and Z. Zhang, “Spectral feature transformation for person re-identification,” in ICCV, KR, 2019. pp. 4976–4985.

[6] Y. Huang, Q. Wu, J. XU, and Y. Zhong, “SBSGAN: Suppression of Inter-Domain Background Shift for Person Re-Identification,” in ICCV, KR, 2019. pp. 9527-9536.

[7] Y. Huang, H. Sheng, Y. Zheng, and Z. Xiong, ”DeepDiff: Learning deep difference features on human body parts for person re-identification,” in Neurocomputing, 2017, 241, 191-203.

[8] Z. Zheng, X. Yang, Z. Yu, L. Zheng, Y. Yang, and J. Kautz, ”Joint discriminative and generative learning for person re-identification,” in CVPR, 2019.

[9] Y. Huang, J. Xu, Q. Wu, Y. Zhong, P. Zhang, and Z. Zhang, ”Beyond Scalar Neuron: Adopting Vector-Neuron Capsules for Long-Term Person Re-Identification,” in IEEE TCSVT, 2019.

[10] Y. Huang, Q. Wu, J. Xu, and Y. Zhong, ”Celebrities-ReID: A Benchmark for Clothes Variation in Long-Term Person Re-Identification,” in IJCNN, 2019.

[11] Y. Sun, L. Zheng, W. Deng, and S. Wang, “Svdnet for pedestrian retrieval,” in ICCV, UK, 2017. pp. 3820-3828.

[12] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, “Beyond Part Models Person Retrieval with Refined Part Pooling,” in ECCV, GER, 2018. pp. 480-496.

[13] J. Chen, Z. Zhang, and Y. Wang, “Relevance metric learning for person re-identification by exploiting global similarities,” in ICPR, SE, 2014. pp. 1657–1662.

[14] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu,D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in NeurIPS, CA, 2014. pp. 2672-2680.

[15] X. Qian, Y. Fu, T. Xiang, W. Wang, J. Qiu, Y. Wu, Y. Jiang, and X. Xue, “Pose-Normalized Image Generation for Person Re-identification,” in ECCV, GER, 2018. pp. 650-667.

[16] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L, Van Gool, “Pose guided person image generation,” in NeurIPS, USA, 2017. pp. 406-416.

[17] Zheng, Zhedong and Zheng, Liang and Yang, Yi, “Unlabeled samples generated by gan improve the person re-identification baseline in vitro,” in ICCV, UK, 2017. pp. 3754-3762.

[18] Y. Huang, J. Xu, Q. Wu, Z. Zheng, Z. Zhang, and J. Zhang, “Multi-pseudo Regularized Label for Generated Samples in Person Re-Identification,” in IEEE Trans. Image Process., US, 2019. pp. 1391-1403.

[19] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” in NeurIPS, ES, 2016. pp. 2234-2242.

[20] D.-H. Lee, “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks,” in ICML, US, 2013.

[21] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in ICCV, UK, 2015. pp. 1116-1124.

[22] E. Ristani, F. Solera, R. Zou, R. Cucchiara, C. Tomasi, “Performance measures and a data set for multi-target, multi-camera tracking,” in ECCV, NL, 2016. pp. 17-35.

[23] L. Zheng, Y. Yang, and H. Alexander G, “Person re-identification: Past, present and future,”, in arXiv, 2016, arXiv: 1610.02984.

[24] Z. Zheng, L. Zheng, and Y. Yang, “A discriminatively learned cnn embedding for person reidentification,” in ACM, US, 2018.

[25] D. Li, X. Chen, Z. Zhang, and K. Huang, “Learning deep context-aware features over body and latent parts for person re-identification,” in CVPR, US, 2017. pp. 384-393.

[26] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, “Joint detection and identification feature learning for person search,” in CVPR, US, 2017. pp. 3415-3424.

[27] Hermans, Alexander Beyer, Lucas and Leibe, Bastian, “In defense of the triplet loss for person re-identification,”, in arXiv, 2017, arXiv: 1703.07737.

[28] W. Li, X. Zhu, and S. Gong, “Person re-identification by deep joint learning of multi-loss classification,” in arXiv, 2017, arXiv: 1705.04724.

[29] S. Bai, X. Bai, and Q, Tian, “Scalable person re-identification on supervised smoothed manifold,” in CVPR, US, 2017. pp. 2530-2539.

[30] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,”, in arXiv, 2015, arXiv: 1511.06434.