Center Based Pseudo-Labeling for Semi-supervised Person Re-Identification

Conference Paper · May 2018

Guodong Ding
Nanjing University of Science and Technology

CENTER BASED PSEUDO-LABELING FOR SEMI-SUPERVISED PERSON RE-IDENTIFICATION

Guodong Ding, Shanshan Zhang*, Salman Khan† and Zhenmin Tang**

*Nanjing University of Science and Technology, †Australian National University

ABSTRACT

Generative Adversarial Networks (GAN) have shown promising results on data modeling and can generate high quality synthetic samples from the data distribution. However, how to effectively use the generated data for improved feature learning still remains an open question. This work proposes a Center based Pseudo-Labeling (CPL) method dedicated to this purpose. The network is trained with both labeled real data and unlabeled synthetic data, under a joint supervision of cross-entropy loss together with a center regularization term, which simultaneously predicts pseudo-labels for unlabeled synthetic data. Experimental results on two standard benchmarks show our approach achieves superior performance over closely related competitors and comparable results with state-of-the-art methods.

***Index Terms*—** pseudo-labels, semi-supervised learning, person re-identification, convolutional networks.

1. INTRODUCTION

Person re-identification aims at matching pedestrians across images captured from multiple non-overlap cameras. Thus, labeling for this task involves manually associating images from different cameras which is rather demanding. The emergence of Generative Adversarial Network (GAN) [1] in 2014 provides an option to tackle this problem by generating images with perceptual quality. However, it is still an open question how to properly adopt them for training. Previous works towards this purpose [2, 3] either assign one single label for all generated samples or use a fixed label for each sample during the training procedure. In this paper we aim to design a dynamic labeling strategy, which progresses through the whole training process. On the other hand, the labeling should take into account the distances between the generated sample and all other labeled samples.

The main contributions of this work are summarized as follows:

- We propose a multi-task loss formulation for the person re-identification task that jointly considers conventional cross-entropy loss for supervised learning and a center loss term for unsupervised clustering.

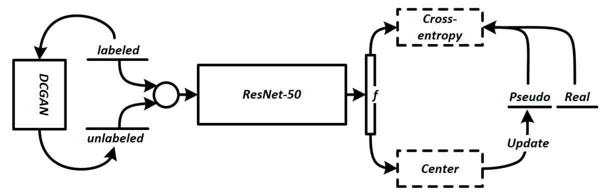

Fig. 1. Overall architecture with our proposed CPL for person re-identification task. DCGAN is trained to generate synthetic data, followed by semi-supervised learning with our proposed multi-task loss (Denoted by dashed boxes).

- We propose a clustering based pseudo-labeling approach for GAN generated samples which helps regularize our proposed model for a better performance over closely related approaches and is on par with state-of-art methods.

The remainder of this paper is organized as follows. Section 2 reviews related work and Section 3 provides details of our proposed semi-supervised learning. Section 4 reports experimental results, and Section 5 concludes this paper.

2. RELATED WORK

We review related work in the fields of person re-identification and pseudo-labeling, respectively.

Person Re-identification. Majority efforts in this area can be grouped into two categories: (a) metric learning and (b) feature extraction based approaches. Metric learning usually takes input in the form of image pairs or triplets and learn a similarity metric using pairwise or triplet loss [4, 5, 6]. However, this stream of work suffers from huge data expansion when constituting image pairs and triplets. The other type of works focus on feature learning, addressing this task in the form of classification. Common practices include first training a pedestrian identity predicting model and then extract last fully connected layer as pedestrian descriptor for retrieval during testing [7, 8].

Pseudo-labeling. Pseudo-labeling is a technique to produce approximate labels for unlabeled data on the basis of labeled data instead of manually labeling them. Existing pseudo-labeling approaches can be categorized as follows:

978-1-5386-1737-3/18/$31.00 ©2018 IEEE

- **All-in-one** [9, 10] simply introduces an extra new and fixed class label for all unlabeled data without considering any relationships between labeled and unlabeled data.
- **One-hot pseudo** [11] improves labeling strategy on the basis of all-in-one and proposes to dynamically assign labels for unlabeled data every forward process according to its maximum class prediction probability.
- **LSRO** [2], **MpRL** [3] uses distributed (multiple) pseudo class labels for a single unlabeled data. LSRO proposed a uniformly distributed pseudo-label while MpRL considers different class contributions. Labels are also updated every iteration.

None of the above methods consider the relationships between labeled and unlabeled data samples to improve the feature representations during semi-supervised learning. A major drawback of [2] approach is the underlying assumption that the synthetic data does not belong to any class, therefore considering a uniform distribution for all unlabeled samples. Our work aims to address this limitation and propose a novel loss function that automatically discovers patterns in the unlabeled data.

3. PROPOSED APPROACH

In this section, we describe our proposed semi-supervised learning approach in person re-identification and some discussions.

3.1. Overview

We propose a semi-supervised learning approach for person re-identification that does not require any extra data besides the training dataset. The whole work-flow is displayed in Fig. 1. The proposed semi-supervised learning approach can be separated in two modules. The first module on the left side is the data generation module, where a generative model (DCGAN) is learned using adversarial training to estimate the data distribution based on existing real samples(e.g., labeled images from training set). Then, the trained generator can be used to obtain large amounts of synthetic image samples lying on the approximated data manifold. The second module takes as input these unlabeled generated data samples alongside the labeled ones to learn feature representations with the joint supervision of a center regularization term which simultaneously predict pseudo labels for generated samples. During testing, the output representations from the convolutional network are used as pedestrian descriptors for a Euclidean distance based retrieval task.

To be specific, we introduce a new learning objective for the semi-supervised training. We propose a new method of pseudo-labeling for unlabeled data samples by exploring the underlying data patterns. Our proposed approach is based on the proposition that a good pseudo-label should reflect the data sample's similarity to real ground truth data. Furthermore, the similarity needs to be more broadly defined and measured with a defined distance metric rather than to a single closest data point.

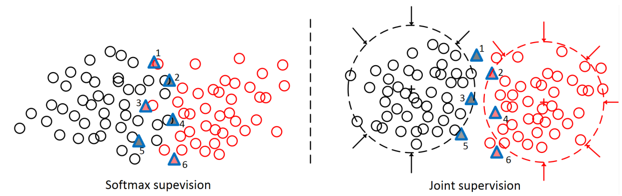

Fig. 2. A toy example illustrating representation distributions in feature space before and after the center regularization term is imposed. Circles with different colors (e.g. red and black) stand for feature representations of *real (labeled)* samples form different categories. Blue triangles represent GAN generated *fake (unlabeled)* data, which are denoted 1-6 from top to bottom. The filled color of each triangle denotes pseudo-label predicted in each case. Cross in each dashed circle denotes the center of that class. (Best viewed in color.)

3.2. Proposed Semi-supervised Learning

We now introduce our proposed center based pseudo-labeling approach. It takes into account the patterns in the labeled data and leverages those to infer pseudo-labels for unlabeled data. Further, it jointly learns more discriminative feature representations for both labeled real data and unlabeled synthetic data.

Loss Function. Common CNN models for classification adopt classical cross-entropy loss during training whose output deeply learned features often have large intra-class variations. Contrastive and triplet losses have been proposed to enhance the discriminative power of these features. However, both these losses suffer from drastic data expansion when constituting the sample pairs or sample triplets from training set. Here, we use the center loss [12] to provide joint supervision with conventional cross-entropy loss to minimize intra-class variations while keeping it simple to train. The final loss for the proposed model is a combination of cross-entropy loss L_S and center regularization term L_C defined as follows:

$$\begin{aligned} L &= L_S + \lambda L_C \\ &= -\sum_{i=1}^{m} \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{k=1}^{K} e^{W_{y_k}^T x_i + b_{y_k}}} + \frac{\lambda}{2} \sum_{i=1}^{m} \|x_i - c_{y_i}\|_2^2 \end{aligned} \quad (1)$$

where x_i is the ith feature representation, belonging to the class y_i, c_{y_i} denotes the y_ith class center of the features. W is the layer parameter of cross-entropy loss. λ is a trade-off parameter to balance two loss functions. K is the total number of classes.

The L_C term seeks to pull closer representations from same class and with the joint supervision, network is therefore encouraged to learn discriminative deep features while assuring low intra-class variations.

Pseudo-label generation. Pseudo-label conceptually defines which class the unlabeled sample belongs to based on some measurement. Distinguished from approaches with the choice of unlabeled sample's maximum class prediction probability, we consider similarity between its feature representation and class centers to decide which class it belongs to, which itself can be regarded as a clustering process. We formulate our similarity definition as cosine similarity given by:

$$sim(u_i, c_k) = \frac{u_i \cdot c_k}{\|u_i\| \|c_k\|} \tag{2}$$

where, u_i is the feature vector for an unlabeled data sample i and c_k is the cluster center. Label $Y(u_i)$ is defined according to its clustering results as:

$$Y(u_i) = \arg\max_k sim(u_i, c_k), \quad s.t., k \in [1, K]. \tag{3}$$

We assign one-hot label, same as one-hot pseudo, for unlabeled data which is in format consistent with ground truth labels for easier implementation and training.

Gradients. It is clear that our loss Eq. (1) is differentiable and thus it can be optimized by Stochastic Gradient Descent (SGD). The gradient of loss L with respect to x_i and the center c_{y_i} update equation are written as:

$$\begin{aligned} \frac{\partial L}{\partial x_i} &= \frac{\partial L_S}{\partial x_i} + \frac{\lambda}{2}\frac{\partial L_C}{\partial x_i} \\ &= \frac{W_{y_i} \cdot e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{k=1}^{K} e^{W_{y_k}^T x_i + b_{y_k}}} - W_{y_i} + \lambda(x_i - c_{y_i}) \end{aligned} \tag{4}$$

$$\Delta c_j = \frac{\sum_{i=1}^{m} \delta(x_i \in L) \cdot \delta(y_i = k) \cdot (c_k - x_i)}{1 + \sum_{i=1}^{m} \delta(x_i \in L) \cdot \delta(y_i = k)} \tag{5}$$

where δ denotes delta function i.e., $\delta(condition) = 1$ if $condition$ is satisfied, and otherwise 0.

Once pseudo-labels are obtained, we train our network following the same procedure as [12] with two notable modifications: (a) Instead of taking the entire training set into account, centers are updated based on mini-batches. (b) Only labeled samples in mini-batches are allowed to update class centers, enabling a more stable training procedure.

We present our proposed semi-supervised feature learning procedure in Algorithm 1.

3.3. Discussion

Why center better labels data? As mentioned before, GAN generator is trained to generate samples following the original data distribution to fool the discriminator. Once GAN is well trained, one can assume its generated samples are close to original ones. However, it will be inappropriate to use the class which has the maximum predicted probability to label data as proposed in [11]. A toy example is shown in Fig. 2 for illustration. Left part in Fig. 2 shows large intra-class representation variations trained only with cross-entropy loss and adopts one-hot pseudo-label proposed in [11]. As the closest real samples to unlabeled data samples 1 and 3 are red, high chances are that 1 and 3 will be classified as red (due to their maximum class prediction is more likely red) despite the fact they are more inclined towards black if seen from a global view. Similar argument stands for data samples 2 and 4. Feature distribution after center regularization term imposed is displayed on the right part with much lower intra-class variations and the unlabeled samples are plausibly and correctly classified according to the distance to each center.

Algorithm 1: The semi-supervised feature learning approach

Input: Labeled data set L, Unlabeled data set U, Maximum iteration T, Parameters θ, Center update rate α
Output: Updated parameters θ
Initialization: Training set $X = L \cup U$, Initialize θ, Class centers $\{c_k = \mathbf{0} | k = 1, 2, ..., K\}$

```
for t = 1 : T do
    Shuffle X and sample mini batch x^t;
    Feed forward x^t through M and calculate
      sim(x_u^t, c_k) using Eq. (2);
    Update pseudo-label y_u^t for x_u^t using Eq. (3);
    Compute the joint loss L^t using Eq. (1);
    Update class center c_j^t = c_j^{t-1} + α Δc_j^t with Eq (5);
    Update the parameter set θ^t;
```

Another worth mentioning fact is that this line of re-id researches use an identity predicting network in training, but extract last fully connected layer activations as the final descriptor to perform similarity calculation when testing. That is to say, final retrieval is based on feature similarity rather than class predictions. Thus, pseudo-labels derived from predictions in varying degrees introduce extra errors. In contrast, center based labeling directly considers feature similarity.

Comparison with related methods. Unlike all-in-one approach with an extra class for unlabeled data, ours takes into consideration the distribution property of GAN generated images and propose to classify them into existing classes. LSRO assigns one single label for all samples while ignoring the variance within generated samples. CPL is superior to one-hot pseudo and MpRL in the sense that it leverages the relationships between labeled and unlabeled data instead of purely based on single class predictions. In summary, our proposed center based pseudo-labeling is better designed than other pseudo-labeling methods.

Table 1. Rank-1 accuracy (%) and mAP (%) on the Market-1501 dataset with varying numbers of unlabeled training data. Best results amongst approaches are in bold whilst best results for different unlabeled data incorporated is underlined.

#GAN images	All-in-one [9, 10]		One-hot Pseudo [11]		LSRO [2]		dMpRL-I [3]		dMpRL-II [3]		CPL(Ours)	
	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP	rank-1	mAP
0(baseline)	72.74	50.99	72.74	50.99	72.74	50.99	72.74	50.99	72.74	50.99	72.74	50.99
12000	76.96	55.68	76.52	55.69	77.17	55.22	77.88	55.84	79.22	58.14	**81.38**	**60.31**
18000	77.40	55.59	77.95	55.04	76.96	55.28	78.36	56.21	79.81	58.31	**82.10**	**62.31**
24000	77.21	56.07	77.62	56.90	78.21	56.33	77.79	56.10	80.37	58.59	**82.04**	**61.26**
30000	77.17	56.19	77.95	56.54	77.46	55.40	78.65	57.15	79.16	57.69	**82.10**	**61.42**
36000	75.92	55.24	77.42	56.38	77.91	55.82	78.95	57.42	79.90	57.61	**82.12**	**60.70**
improvement	4.66	5.20	5.21	5.91	5.47	5.34	6.21	6.43	7.63	7.60	**9.38**	**11.32**

4. EXPERIMENTS

In this section, we evaluate our proposed method on two standard person re-identification benchmarks: Market-1501 and DukeMTMC-reID. We will start with the implementation details, followed by the evaluations and an ablation study.

4.1. Implementation Details

GAN network. We follow [2] and train a DCGAN to generate synthetic unlabeled pedestrian images to enlarge the training set. The generator first applies a linear function on a 100-dim random noise to form a $4\times4\times16$ tensor, then further adds 6 deconvolution layers with a kernel size of 5×5 to generate the final $128\times128\times3$ image. The discriminator uses 5 convolutional layers with 5×5 kernels to separate real and synthetic samples from its input. After training, we use the generator to produce up to 36,000 synthetic images for following semi-supervised learning. Some samples of real and generated synthetic data are displayed in Fig. 3. Although these generated images can be easily recognized as fake by human eyes, they can still help regularize the model and result in better matching accuracy.

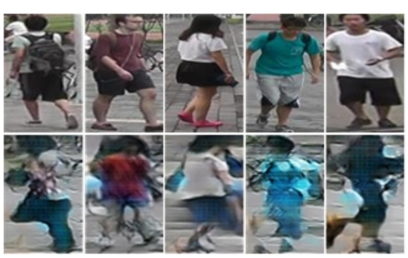

Fig. 3. Above are real samples from Market-1501 dataset and below are synthetic images generated by a trained DCGAN.

Re-id Baseline. For fair comparison, we followed [2] and adopted ResNet-50 which is commonly used as our backbone network. In our experiments, no changes were made to the architecture expect for substituting the last 1000 class activation neurons to the number of identities in each dataset.

4.2. Evaluations on Standard Benchmarks

Comparison with existing pseudo-label methods. To evaluate CPL, we compare it with four competitive pseudo-labeling methods: all-in-one, one-hot pseudo, LSRO and MpRL. MpRL is the current state-of-art approach tackling semi-supervised learning problem on person re-identification with GAN generated images. Detailed results on Market-1501 are reported in Table 1.

Compared with existing virtual label methods, ours outperforms the top competitor dMpRL-II by a margin of 1.75% and 3.72% in rank-1 accuracy and mAP respectively on the Market-1501 dataset. dMpRL-II and dMpRL-I differs in the starting point of dynamical label process. dMpRL-II slightly performs better (around 1-2%) because labels are assigned after certain number of epochs when the network is relatively stabilized while dMpRL-I generates labels in the very beginning. This is intuitive since a stable network can make stable predictions for label ranking in their case. Our proposed CPL follows the same setting as dMpRL-I and produces pseudo-labels from scratch. This is reasonable, because our approach allows automatic adaptation for the pseudo-labels.

We also test our approach on the DukeMTMC-reID dataset to show the generalization ability, results are shown in Table 2. Whole training procedure is identical to that on Market-1501. Our method achieves the better results than other works on pseudo-labeling and outperforms the most recent competitor [3] by a margin of 2.68% in rank-1 accuracy and 3.41% in mAP.

Table 2. Comparison of related approaches for pseudo-labeling on the DukeMTMC-reID dataset. Rank-1 accuracy (%) and mAP (%) are reported.

Method	rank-1	mAP
baseline	65.22	44.99
LSRO [2]	67.68	47.13
dMpRL [3]	68.24	48.58
CPL (Ours)	**70.92**	**51.99**

Amount of unlabeled data. In our experiments, different numbers of unlabeled data samples were used during training

Table 3. Comparison with state-of-the-art methods on the Market-1501 dataset. Best and second best results are denoted in bold and underlined, respectively.

Method	Market 1501	
	rank-1	mAP
Gate-reID (ECCV'16) [14]	65.88	39.55
SCSP (CVPR'16) [15]	51.90	26.35
DNS (CVPR'16) [16]	61.02	35.68
ResNet+OIM (CVPR'17) [17]	82.10	-
Latent Parts (CVPR'17) [18]	80.31	57.53
P2S (CVPR'17) [19]	70.72	44.27
Consistent-Aware (CVPR'17) [20]	80.90	55.60
Spindle (CVPR'17) [21]	76.90	-
SSM (CVPR'17) [22]	82.21	<u>68.80</u>
JLML (IJCAI'17) [23]	**85.10**	65.50
SVDNet (ICCV'17) [24]	82.30	62.10
Part Aligned (ICCV'17) [25]	81.00	63.40
PDC (ICCV'17) [26]	84.14	63.41
LSRO (ICCV'17) [2]	78.06	56.23
dMpRL-II (Arxiv'18) [3]	80.37	58.59
Baseline	72.74	50.99
Ours	82.10	62.31
Ours+re-rank	<u>84.47</u>	**75.90**

to show its influence. A considerable performance increase in both rank-1 accuracy (9.38%) and mAP (11.32%) compared to baseline is observed when our proposed approach is applied. However, an increase in the amount of unlabeled images (from 12000 to 36000) failed to lead to further boost and the final result fluctuates around 82%. Same trend is observed amongst all other pseudo-labeling methods. For this phenomenon, we speculate it is due to the inherent representation ability of generated images. Those images are sampled from a specific distribution (manifold of real data), so simply increasing samples can not produce any new information that benefits the retrieval task eventually.

Comparison with state-of-art methods. We compare our proposed approach with other state-of-art works on Market-1501 to show its competence. Our method (rank-1:82.10%, mAP: 62.31%) shows to be very competitive with many state-of-art methods except for JLML (rank-1:85.1%, 65.50%), PDC and SVDNet. The main reason why JLML outperforms by a relatively large margin is because JLML adopted a much stronger baseline (around 3% higher than ours) and incorporates three extra networks focusing on different local areas compared to our single branch architecture. After applying a re-ranking technique from [13] applied, a further boost is observed showing reciprocal relationships are encoded in learned identity representations.

4.3. Ablation Study

We provide ablative experiment results on Market-1501 to evaluate each component in our proposed approach. The network is under full supervision of labeled data for baseline and center loss but turns into a semi-supervision case when unlabeled data is provided.

Table 4. Ablative experiments about effectiveness of center regularization term and pseudo-labeling on Market-1501.

methods	rank-1	mAP	supervision
baseline	72.74	50.99	full
center	79.45	57.25	full
center + pseudo	82.10	62.31	semi

Regularization term. In order to study the effect of center regularization term on final results, we discard all generated data and train only on labeled data with the regularization term imposed. Note in this case, it is a fully supervised mode. As show in Table. 4, imposing the regularization term leads to a 6.71% rank-1 increase to get 79.45% and 6.26% gain in mAP to achieve 57.25%. This improvement is intuitive since the regularization term helps to reduce the intra-class variations and thus leads to stronger feature representations.

Pseudo-label. Adopting generated images from DCGAN with pseudo-labels predicted by CPL turns previous full-supervised model into a semi-supervised one. With synthetic data labeled by our proposed CPL, the network gains a further performance boost on both metrics to 82.1% rank-1 accuracy and 62.31% mAP, which justify the effectiveness of our labeling solution. In this experiment, the number of unlabeled images is set to 18,000.

5. CONCLUSION

In this paper, we proposed a center based pseudo-labeling (CPL) approach for GAN generated synthetic images and evaluated its effectiveness on the person re-identification task. Specifically, unlabeled each synthetic sample is assigned a one-hot class label according to its closest class center. Training with a combination of labeled real and pseudo-labeled synthetic samples better address the re-id problem in a semi-supervised manner. Experimental results show that our proposed approach outperforms other pseudo-labeling methods on the person re-identification task and achieves competitive accuracy compared to state-of-the-art methods.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 61473154 and 61702262), Jiangsu Key Laboratory of Image and Video Understanding for Social Safety (Grant No. 30920140122007) and Fundamental Research Funds for the Central Universities (Grant No.30918011322).

6. REFERENCES

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *NIPS*, 2014.

[2] Zhedong Zheng, Liang Zheng, and Yi Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," *ICCV*, 2017.

[3] Yan Huang, Jinsong Xu, Qiang Wu, Zhedong Zheng, Zhaoxiang Zhang, and Jian Zhang, "Multi-pseudo regularized label for generated samples in person re-identification," *arXiv preprint arXiv:1801.06742*, 2018.

[4] De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng, "Person re-identification by multi-channel parts-based cnn with improved triplet loss function," in *CVPR*, 2016.

[5] Rahul Rama Varior, Bing Shuai, Jiwen Lu, Dong Xu, and Gang Wang, "A siamese long short-term memory architecture for human re-identification," in *ECCV*, 2016.

[6] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li, "Deep metric learning for person re-identification," in *ICPR*, 2014.

[7] Liang Zheng, Yi Yang, and Alexander G Hauptmann, "Person re-identification: Past, present and future," *arXiv preprint arXiv:1610.02984*, 2016.

[8] Yanbei Chen, Xiatian Zhu, and Shaogang Gong, "Person re-identification by deep learning multi-scale representations," in *CVPR*, 2017.

[9] Augustus Odena, "Semi-supervised learning with generative adversarial networks," *arXiv preprint arXiv:1606.01583*, 2016.

[10] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen, "Improved techniques for training gans," in *NIPS*, 2016.

[11] Dong-Hyun Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on Challenges in Representation Learning, ICML*, 2013.

[12] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao, "A discriminative feature learning approach for deep face recognition," in *ECCV*, 2016.

[13] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li, "Re-ranking person re-identification with k-reciprocal encoding," in *CVPR*, 2017.

[14] Rahul Rama Varior, Mrinal Haloi, and Gang Wang, "Gated siamese convolutional neural network architecture for human re-identification," in *ECCV*, 2016.

[15] Dapeng Chen, Zejian Yuan, Badong Chen, and Nanning Zheng, "Similarity learning with spatial constraints for person re-identification," in *CVPR*, 2016.

[16] Li Zhang, Tao Xiang, and Shaogang Gong, "Learning a discriminative null space for person re-identification," in *CVPR*, 2016.

[17] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang, "Joint detection and identification feature learning for person search," in *CVPR*, 2017.

[18] Dangwei Li, Xiaotang Chen, Zhang Zhang, and Kaiqi Huang, "Learning deep context-aware features over body and latent parts for person re-identification," in *CVPR*, 2017.

[19] Sanping Zhou, Jinjun Wang, Jiayun Wang, Yihong Gong, and Nanning Zheng, "Point to set similarity based deep feature learning for person re-identification," in *CVPR*, 2017.

[20] Ji Lin, Liangliang Ren, Jiwen Lu, Jianjiang Feng, and Jie Zhou, "Consistent-aware deep learning for person re-identification in a camera network," in *CVPR*, 2017.

[21] Haiyu Zhao, Maoqing Tian, Shuyang Sun, Jing Shao, Junjie Yan, Shuai Yi, Xiaogang Wang, and Xiaoou Tang, "Spindle net: Person re-identification with human body region guided feature decomposition and fusion," in *CVPR*, 2017.

[22] Song Bai, Xiang Bai, and Qi Tian, "Scalable person re-identification on supervised smoothed manifold," in *CVPR*, 2017.

[23] Wei Li, Xiatian Zhu, and Shaogang Gong, "Person re-identification by deep joint learning of multi-loss classification," *arXiv preprint arXiv:1705.04724*, 2017.

[24] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang, "Svdnet for pedestrian retrieval," *arXiv preprint*, 2017.

[25] Liming Zhao, Xi Li, Jingdong Wang, and Yueting Zhuang, "Deeply-learned part-aligned representations for person re-identification," *arXiv preprint arXiv:1707.07256*, 2017.

[26] Chi Su, Jianing Li, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian, "Pose-driven deep convolutional model for person re-identification," in *ICCV*, 2017.