Deep Transfer Learning for Person Re-identification

Haoran Chen[1], Yemin Shi[2], Ke Yan[2], Yaowei Wang[1], Tao Xiang[3], Mengyue Geng[2], Yonghong Tian[2]
[1] School of Information and Electronics, Beijing Institute of Technology, Beijing, China,
[2] National Engineering Laboratory for Video Technology, School of EE&CS,
Peking University, Beijing, China
[3] School of EECS, Queen Mary University of London

***Abstract*—Person re-identification (Re-ID) poses an inevitable challenge to deep learning: how to learn a robust deep model with millions of parameters on a small training set of few or no labels. In this paper, two deep transfer learning methods are proposed to address the training data sparsity problem, respectively from the supervised and unsupervised settings. First, a two-stepped fine-tuning strategy with proxy classifier learning is developed to transfer knowledge from auxiliary datasets. Second, given an unlabelled Re-ID dataset, an unsupervised deep transfer learning model is proposed based on a co-training strategy. Extensive experiments show that the proposed models achieve a good performance of deep Re-ID models.**

***Index Terms*—Person Re-ID, Deep Transfer Learning, Unsupervised Learning**

I. Introduction

Person re-identification (Re-ID) is the problem of matching people across non-overlapping camera views, which typically arises in a surveillance application. Despite the best efforts from the computer vision researchers, it remains an unsolved problem [1]. Earlier works focus on either designing view-insensitive feature representations [2], [3], or learning an effective distance metric [4], [5], or both [6]. Recently, deep Re-ID models started to attract attention [7], [8].

Given insufficient training samples, transferring feature representations learned from a larger auxiliary dataset becomes critical. Indeed, transfer learning has been considered in most existing deep Re-ID works. In particular, given a small Re-ID dataset with only a few hundreds of labelled identities, existing models are typically pretrained on larger auxiliary Re-ID datasets followed by fine-tuning on the target set. However, the domain gaps between different Re-ID datasets are typically large due to the often drastically different camera viewing conditions. As a result, models that adopt this one-stepped fine-tuning strategy would often gain only limited performance improvements with a possibility of even inducing negative transfer [1].

In this work, we aim to address the problem of lacking labelled training data in Re-ID by proposing two deep transfer learning methods. The first method is a two-stepped fine-tuning strategy based on proxy classifier learning, which is designed for situations where a small number of labelled training data are available. Specifically, we formulate a deep Re-ID network based on GoogLeNet [9] with an identity classification loss and a verification loss to learn a discriminative feature representation.

Corresponding author: Yaowei Wang (yaoweiwang@bit.edu.cn).

The second proposed deep transfer learning method is designed for unsupervised deep Re-ID. Transfer learning from labelled source data to unlabelled target data is an unsupervised (by 'unsupervised', we mean target-unsupervised domain adaptation, a definition adopted by [10]–[12]) domain adaptation problem which is still an open problem for deep models and has not been attempted by existing deep Re-ID works. In this paper, a novel co-training based unsupervised transfer learning model is proposed. We show that such a deep/non-deep hybrid co-training framework can effectively prevent model drift and yield Re-ID performance that is better than most existing supervised deep Re-ID models.

Finally, our models achieve a good performance of deep Re-ID models: Rank-1 accuracy of 85.4%, 83.7% and 56.3% on CUHK03, Market1501, and VIPeR respectively; and our unsupervised model rank-1 accuracy 45.1% on VIPeR.

II. Related Work

Deep Re-ID models Existing deep Re-ID models [7], [8], [13]–[15] differ significantly in their network architectures, which are largely determined by the training objectives/losses. Specifically, most existing works cast the Re-ID problem as a deep metric learning problem and employ pairwise verification loss [13], [15] or triplet ranking loss [7], [14], or both [16]. Correspondingly the overall network architecture is a Siamese CNN network with either two or three branches for the pairwise or triplet loss respectively.

Deep transfer learning Transfer learning or domain adaptation is an extensively studied topic [17]. Transfer learning is widely used for deep learning when a target task is short of labelled data. The most common deep transfer learning strategy is fine-tuning [18]. A systematic study is presented in [18] which examines how transferable the feature outputs of different layers are between the source and target domains. It concludes that the generalisation ability diminishes when the discrepancy between the source and target domains increases.

Deep unsupervised domain adaptation In theory, any unsupervised deep learning methods can be applied for unsupervised domain adaptation. Recently a number of deep unsupervised transfer learning models are proposed [11], [12] which aim to align the data distributions of different domains. Nevertheless, the domain gap between different Re-ID datasets is significant and cannot be overcome by just aligning the data distributions, making them less effective than the proposed model, as demonstrated in our experiments (see Sec. V-C).

978-1-5386-5321-0/18/$31.00 ©2018 IEEE

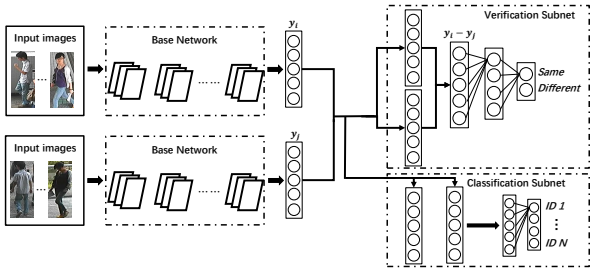

Fig. 1. The proposed deep Re-ID network architecture.

Our contributions are as follows: (1) A two-stepped fine-tuning with proxy classifier learning is developed for supervised deep transfer learning. (2) A co-training based unsupervised domain adaptation method is proposed for unsupervised deep Re-ID. (3) Comprehensive evaluations are presented to provide insights regarding how to design the optimal network architecture and learning objectives to facilitate deep transfer learning.

III. Deep Re-ID Model

A. Network Architecture

Overview The overall network architecture of the proposed deep Re-ID model is illustrated in Fig. 1. The model has two training objectives/losses: an ID classification loss and a pairwise verification loss. As a result, the network contains three parts (see Fig. 1): a base network shared by the two branches, an ID classification subnet, and a pairwise verification subnet. The two main branches of the network have the same base network architecture and share their parameters, hence the name Siamese. After feature vectors are computed for the input images using the base network, the pairwise verification subnet takes a pair of features and learn to distinguish whether they come from the same person or not. In the meantime, the person ID classification subnet learns to classify each feature output of the base network into a class corresponding to the input image person ID.

Base network The classification subnet learns a softmax person ID (SID) classifier with a cross-entropy loss that distinguishes different people from each other. The pairwise verification (PV) subnet first takes two feature vectors $\mathbf{y}_i$ and $\mathbf{y}_j$ as input. They are first fused with element-wise subtraction. Subsequently, the difference vector is multiplied by a random dropout mask $\mathbf{r}_v$ and then passed to a rectified linear unit (ReLU). After a fully connected (FC) layer, the last layer of the verification network is a softmax layer with two output nodes, corresponding to whether or not the input image pair contains the same person. Let $\Psi_0(\mathbf{x}, \mathbf{x}')$ and $\Psi_1(\mathbf{x}, \mathbf{x}')$ be the functions learned by the network in the two nodes, where $\mathbf{x}$ and $\mathbf{x}'$ are the input images. In the training stage, the negative log-likelihood is used as a cost function J.

Once trained, we pre-compute the output vector of the base network $\phi(\mathbf{x}_g)$ for each gallery image $\mathbf{x}_g$; and when any probe $\mathbf{x}_p$ comes in, we compute its feature output $\phi(\mathbf{x}_p)$ and calculate its distance to $\mathbf{x}_g$ by:

$$Dist(\mathbf{x}_p, \mathbf{x}_g) = ||\phi(\mathbf{x}_p) - \phi(\mathbf{x}_g)||_2 \quad (1)$$

which is about three magnitudes faster in our model than entering the verification subnet and computing the softmax score as the distance.

IV. Deep Transfer Learning for Re-ID

A. Supervised Transfer Learning

Coping with large domain discrepancy Two scenarios are typically considered in Re-ID under the supervised setting: (1) The target Re-ID dataset is 'large', i.e. having more than 1,000 identities, for instance CUHK03 [19] and Market1501 [20], and (2) it is 'small' with less than 1,000, e.g. VIPeR [21]. Specifically, for large Re-ID datasets, our model is pretrained on ImageNet and then fine-tuned, whilst for smaller datasets, one more stage of fine-tuning from large to small Re-ID datasets is performed. In both scenarios, the classification subnet is problematic because the source classification task and the target one have completely different class labels (object categories vs. person IDs, or different sets of person IDs). This means that the classification subnet has to be initialised randomly. This would inevitably generate a great deal of noisy gradients, which after being back-propagated to the base network will generate 'garbage gradients' that could derail the model adaptation.

Two-stepped fine-tuning with proxy classifier learning Suppose we have a large source Re-ID dataset S and a small target dataset T with N_s and N_t unique person identities respectively. Given an initial model trained using S, our goal is to transfer the learned feature representation from S to T. Formally, we need to learn the base network's mapping function ϕ_t and an identity classifier f_t such that:

$$< \phi_t, f_t >= \arg\max_{\phi, f} \mathbb{E}_{\mathbf{X}, l\sim p_t} p(\mathbf{L}|\mathbf{X}, \phi, f) \quad (2)$$

where p_t is the target data distribution and $p(\mathbf{L}|\mathbf{X}, \theta, f)$ denotes the probability distribution learned by our model. Using the architecture described in Sec. III-A we first learn a base network ϕ_s and its corresponding ID classifier f_s on S. During domain adaptation, f_s cannot be re-used because the N_s source and N_t target identities have no overlapping. The original N_s-nodes classifier layer thus has to be replaced with a randomly initialised one with N_t nodes, which we denote as f_r. From f_r, we first learn a proxy classifier f_{proxy} from T but with the mapping function ϕ_s:

$$f_{proxy} = \arg\max_{f} \mathbb{E}_{\mathbf{X}, \mathbf{L}\sim p_t} p(\mathbf{L}|\mathbf{X}, \phi_s, f). \quad (3)$$

f_{proxy} is thus designed as a bridge (proxy) between the source and target domain. After f_{proxy} has been learned, in the second fine-tuning step, the whole network is updated (i.e. both ϕ_t and f_t), but we limit the parameter updating of f to small steps so that the proxy classifier f_{proxy} acts as a regulariser to estimate the final ϕ_t and f_t.

Algorithm 1 Pseudo Label Generation

Input: Training image set $\mathbf{X} = \{\mathbf{x}_1^a, ..., \mathbf{x}_{M_a}^a, \mathbf{x}_1^b, ..., \mathbf{x}_{M_b}^b\}$, base network ϕ_s, hyper parameter K
Output: Pseudo label set $\mathbf{L} = \{l_1^a, ..., l_{M_a}^a, l_1^b, ..., l_{M_b}^b\}$
1: **For each** $l_i \in \mathbf{L}$ **do**:
2: $l_i = \varnothing$;
3: Compute feature representation $R = \{\phi_s(\mathbf{x}_1^b), ...\phi_s(\mathbf{x}_{M_b}^b)\}$;
4: **For each** $\mathbf{x}_i^a \in \mathbf{X}$ **do**:
5: Compute $\phi_s(\mathbf{x}_i^a)$;
6: $l_i^a = \{i\}$;
7: Calculate the KNN set N_i of $\phi_s(\mathbf{x}_i^a)$ in R;
8: **For each** $\phi_s(\mathbf{x}_i^b) \in N_i$ **do**:
9: $l_i^b = l_i^b \cup \{i\}$;

B. Unsupervised Transfer Learning

Now the M_t target training images of an unknown number of identities are unlabelled. For simplicity of symbols, we assume that they are collected from two camera views denoted as A and B respectively. Let's denote the training set as $\mathbf{X} = \{\mathbf{X}^a, \mathbf{X}^b\}$, where $\mathbf{X}^a = \{\mathbf{x}_1^a, ..., \mathbf{x}_{M_a}^a\}$ contains M_a images in view A, while $\mathbf{X}^b = \{\mathbf{x}_1^b, ..., \mathbf{x}_{M_b}^b\}$ for the M_b images in view B; we thus have $M_t = M_a + M_b$. For each image $\mathbf{x}$, an D-dimensional feature vector $\mathbf{y} = \phi_s(\mathbf{x})$ is computed by the base network learned using the source dataset S. We wish to learn a better network using T with M_t unlabelled images yielding an updated mapping function ϕ_t.

Semantic bootstrapping with pseudo labels A simple solution to the unsupervised transfer learning problem is to use the supervised two-stepped fine-tuning method described in Sec. IV-A but with pseudo labels. The generation process of pseudo labels is detailed in Algorithm 1. Specifically, for each of the M_a images from camera A $\mathbf{x}_i^a \in \mathbf{X}^a$, we assign it with a unique identity label. After that, each of the M_b images from camera B is assigned with the same label as its K-nearest neighbour (KNN) from A based on $||\phi_s(\mathbf{x}_i^a) - \phi_s(\mathbf{x}_j^b)||_2$. With these pseudo labels, a semantic bootstrapping strategy [22] is used to train the deep model, in which the base network will produce an updated mapping function $\widetilde{\phi}$ which will again be used to generate another set of pseudo labels for retraining using Algorithm 1 with ϕ_s replaced by $\widetilde{\phi}$.

Solving the model drift problem by co-training In particular, pseudo labels that generated by Algorithm 1 clearly do not correspond to the real identity labels: For a start, there could be multiple images per person in each camera, so there are less than M_a identities; second, the KNN can only give a visually similar person which by no means is always the same person. These pseudo labels are thus highly noisy. Model drift is thus a big problem: The errors in the soft labels will be propagated with the iterations and quickly magnified. To address the issue of model drift, co-training [23], [24] is used in our model. It was first designed for using the same model with two sufficient and yet conditionally independent views (feature representations) as inputs to label some unlabelled instances for each other [23]. Since in most problem settings such views do not exist, in practice one often has a co-training style algorithm whereby two different models with the same features or even same model with same feature but different parameter settings are used [24]. The key is that both models need to be effective and importantly complementary to each other.

In our case, we have already got the semantic bootstrapping deep CNN as one of the two unsupervised models. The other model needs to be both effective on its own and complementary. To this end, we choose a graph regularised subspace learning model [25], [26]. Such a model aims to learn a discriminative subspace where the data distribution is smooth with regard to a KNN graph constructed in the input feature space. In such a learned subspace, data clusters can be formed to provide the pseudo labels for the semantic bootstrapping deep model. In the meantime, it uses the deep model learned feature vector $\mathbf{y} = \phi(\mathbf{x})$ as model input as well as to construct the graph for regularisation.

Formally, given our pretrained deep Re-ID model, we obtain a feature matrix from the base network output $\mathbf{Y} = [\mathbf{Y}^a, \mathbf{Y}^b] \in \mathbb{R}^{D \times M_t}$, where $\mathbf{Y}^a = [\mathbf{y}_1^a, \ ... \ , \mathbf{y}_{M_a}^a] \in \mathbb{R}^{D \times M_a}$ and $\mathbf{Y}^b = [\mathbf{y}_1^b, \ ... \ , \mathbf{y}_{M_b}^b] \in \mathbb{R}^{D \times M_b}$. We aim to learn a subspace defined by a dictionary $\mathbf{D}$ and a new representation $\mathbf{Z}$ in the subspace. $\mathbf{D}$ and $\mathbf{Z}$ can be estimated jointly by solving the following optimisation problem:

$$(\mathbf{D}^*, \mathbf{Z}^*) = \min_{\mathbf{D},\mathbf{Z}} \|\mathbf{Y} - \mathbf{D}\mathbf{Z}\|_F^2 + \lambda\Omega(\mathbf{Z}) \ \ s.t. \ \ \|\mathbf{d}_i\|_2^2 \leq 1, \ (4)$$

where the first term is the reconstruction error evaluating how well a linear combination of the learned atoms can approximate the input data, and $||.||_F$ denotes the matrix Frobenious norm. $\Omega(\mathbf{Y})$ is the graph regularisation term that is weighted by λ:

$$\Omega(\mathbf{Z}) = \sum_{ij} W_{ij} \|\mathbf{z}_i - \mathbf{z}_j\|_2^2. \quad (5)$$

where the graph is encoded by an affinity matrix $\mathbf{W} \in \mathbb{R}^{M_t \times M_t}$ for M_t data points where $W_{i,j} \neq 0$ only when $\mathbf{y}_i$ and $\mathbf{y}_j$ are from two different camera views and are nearest neighbours. With the learned new representation $\mathbf{Z}$, we can generate pseudo labels for the unlabelled target data, that is, the cross-view nearest neighbours are obtained by $||\mathbf{z}_i^a - \mathbf{z}_j^b||_2$ instead of $||\phi(\mathbf{x}_i^a) - \phi(\mathbf{x}_j^b)||_2$. With these pseudo labels, another round of semantic bootstrapping of the deep model is carried out and the updated base network then produces input vectors and a new graph for the subspace learning model. This iterative process normally converges after 2-3 iterations in our experiments.

V. Experiments

A. Datasets

Five widely used datasets are used including two large datasets (CUHK03 [19] and Market1501 [20]) and three small ones (VIPeR [21], PRID [27] and CUHK01 [28]).

B. Supervised Transfer Learning

Results on large datasets On the two large Re-ID datasets, namely CUHK03 and Market1501, transfer learning using our model takes place between ImageNet (ILSVRC 2012)

	Manual	Detected
DNS [5]	62.5	54.7
LSSCDL [29]	57.0	51.2
Siamese LSTM [30]	-	57.3
EDM [15]	61.3	52.0
Joint Learning [16]	-	52.1
CAN [14]	65.7	63.1
Ours	**85.4**	**84.1**

TABLE I
SUPERVISED RESULTS (RANK 1 MATCHING ACCURACY IN %) ON THE CUHK03 DATASET. '-' MEANS NO REPORTED RESULT IS AVAILABLE.

	Single query		Multi-query	
	R1	mAP	R1	mAP
SCSP [4]	51.9	26.3	-	-
DNS [5]	61.0	35.6	71.5	46.0
Siamese LSTM [30]	-	-	61.6	35.3
Gated S-CNN [13]	65.8	39.5	76.0	48.4
CAN [14]	48.2	24.4	-	-
Ours	**83.7**	**65.5**	**89.6**	**73.8**

TABLE II
SUPERVISED RESULTS ON MARKET-1501

and the target Re-ID dataset. The results of our model are compared with the state-of-the-art deep and non-deep Re-ID models in Table I and Table II respectively (they are grouped together in the tables). Due to space limit, only the most competitive ones since 2015 are chosen. We can make the following observations: (1) Our model significantly achieves good performance: on CUHK03, the gap is 10.1% using the manually cropped images and 16.0% using the detected ones. The gap is even bigger for Market, particular on the mAP metric: 26.0% over Gated S-CNN [13] under the single query setting. (2) The best competitors on these two large datasets are all deep learning based. However, their advantages over the hand-crafted feature based models are modest (especially on Market) and far less pronounced than what is widely observed in other visual recognition tasks. This is because the large Re-ID datasets are still relatively small to release the full potential of a deep model. However, with our model, the gap is clear now.

Results on small datasets On the three smaller datasets, the ImageNet-pretrained base model is first trained using

	VIPeR	PRID	CUHK01 (N_t=871/485)
TMA [31]	43.8	-	-
ℓ1 GL [32]	41.5	30.1	-/50.1
Siamese LSTM [30]	42.4	-	-
DNS [5]	51.1	40.9	-/69.0
MCP-CNN [7]	47.8	22.0	-/53.7
Gated S-CNN [13]	37.8	-	-
EDM [15]	40.9	-	86.6/-
Joint Learning [16]	35.8	-	72.5/-
CAN [14]	-	-	81.0/-
Ours	**56.3**	**43.6**	**93.2 / 77.0**

TABLE III
SUPERVISED RESULTS ON VIPER, PRID AND CUHK01. *THE DGD RESULTS ON PRID WERE OBTAINED BY USING 10 TIMES MORE TRAINING IMAGES FROM THE ORIGINAL PRID VIDEO DATASET, GIVING IT A HUGE UNFAIR ADVANTAGE.

	R1	R5	R10	R20
One-stepped	47.6	77.2	86.8	93.1
Two-stepped	**56.3**	**83.3**	**90.5**	**96.0**

TABLE IV
TWO-STEPPED VS. ONE-STEPPED FINE-TUNING ON VIPER

	VIPeR	PRID	CUHK01
CDTL [33]	31.5	24.2	27.1
ℓ1 GL [32]	33.5	25.0	41.0
Ours	**45.1**	**36.2**	**68.8**

TABLE V
UNSUPERVISED TRANSFER LEARNING RESULTS

CUHK03+Market. We then apply the two-stepped fine-tuning strategy on the target VIPeR/PRID/CUHK01 dataset. The comparative results are presented in Table III. Note that the compared hand-crafted feature based models have two sub-groups: those with one type of feature and those using multiple features based on model fusion/ensemble. In addition, most compared deep models use transfer learning, but with the standard one-stepped fine-tuning (typically from CUHK03+Market to the target dataset). It can be seen that our deep Re-ID model achieves the best results on all three datasets.

Ablation study In this experiment, we evaluate the contribution of the proposed two-stepped fine-tuning with proxy classifier learning strategy, in comparison with the standard ones-stepped fine-tuning strategy used by most previous deep Re-ID models. Table IV shows that the two-stepped fine-tuning strategy brings about 8.7% at Rank 1 on VIPeR. This suggests that the two-stepped fine-tuning strategy is much more effective for knowledge transfer in deep Re-ID.

C. Unsupervised Transfer Learning

Comparative results Our co-training based unsupervised transfer learning model is compared against the best reported results on the three small datasets in Table V. Note that to the best of our knowledge, no published deep Re-ID model has attempted this challenging setting. The results clearly show that we can beat the existing hand-crafted features based models by big margins. Compared with the supervised learning results in Table III, our unsupervised model is very competitive, beating most of them, particularly the deep learning based ones.

Ablation study In this experiment, the effectiveness of the co-training strategy is evaluated. Our unsupervised model alternates between a pseudo-label semantic bootstrapping deep model and a graph-regularised subspace learning model. Table VI shows that both models are effective on their own and when combined in our co-training framework, boost the performance by 2-3%.

	R1	R5	R10	R20
Semantic Bootstrapping	42.8	66.9	77.3	85.9
Subspace	42.3	71.5	79.8	87.5
AE	36.4	62.3	74.0	81.9
Adversarial [10]	22.8	38.6	50.3	63.9
Ours	**45.1**	**73.1**	**81.7**	**89.4**

TABLE VI
EVALUATIONS ON ALTERNATIVE UNSUPERVISED MODEL ON VIPER

Alternative Unsupervised Transfer Learning Models We also compare our model with two alternative unsupervised transfer learning methods. The first one combines a CNN with an autoencoder. In our CNN+autoencoder model, the input layer of the autoencoder is the feature output of the base network; the middle layer dimension is set to 512 and the output layer has the same dimension as the input layer (1,024). Note that since the size of target dataset is too small to train the AE from scratch, we initialize the parameters of the AE layers by first pretraining them using images in the source dataset. The second model compared is the deep unsupervised domain alignment model using gradient reversal and adversarial learning [10]. Specifically, we add a domain classifier connected to the feature extractor (i.e. our base network) via a gradient reversal layer that multiplies the gradient by a certain negative constant during the back-propagation based training. The results in Table VI show that both the compared models yield much weaker performance than the proposed co-training based model. The autoencoder model is weaker because it is not discriminative.

VI. Conclusion

We have proposed a couple of deep transfer learning strategies to tackle the challenging person Re-ID problem with small datasets. Our experiments validated the claim that a two-stepped fine-tuning method with proxy classifier learning is effective for supervised transfer learning with our deep Re-ID model. More importantly, we show a co-training based deep unsupervised transfer learning model can achieve good Re-ID performance without any labelled data.

Acknowledgement. This work is partially supported by grants from the National Key R&D Program of China under grant 2017YFB1002401, the National Natural Science Foundation of China under contract No. U1611461, No. 61471042, No. 61390515, No. 61425025, and No. 61650202, also supported by grants from NVIDIA and the NVIDIA DGX-1 AI Supercomputer.

References

[1] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," *arXiv preprint*, 2016.
[2] S. Liao and S. Z. Li, "Efficient psd constrained asymmetric metric learning for person re-identification," in *ICCV*, 2015.
[3] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato, "Hierarchical gaussian descriptor for person re-identification," in *CVPR*, 2016.
[4] D. Chen, Z. Yuan, B. Chen, and N. Zheng, "Similarity learning with spatial constraints for person re-identification," in *CVPR*, 2016.
[5] L. Zhang, T. Xiang, and S. Gong, "Learning a discriminative null space for person re-identification," in *CVPR*, 2016.
[6] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *CVPR*, 2015.
[7] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng, "Person re-identification by multi-channel parts-based cnn with improved triplet loss function," in *CVPR*, 2016.
[8] T. Xiao, H. Li, W. Ouyang, and X. Wang, "Learning deep feature representations with domain guided dropout for person re-identification," in *CVPR*, 2016.
[9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *CVPR*, 2015.
[10] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *ICML*, 2015.
[11] X. Zhang, F. X. Yu, S. Chang, and S. Wang, "Deep transfer network: Unsupervised domain adaptation," *CoRR*, 2015.
[12] M. Long, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," *CoRR*, 2016.
[13] R. R. Varior, M. Haloi, and G. Wang, "Gated siamese convolutional neural network architecture for human re-identification," in *ECCV*, 2016.
[14] H. Liu, J. Feng, M. Qi, J. Jiang, and S. Yan, "End-to-end comparative attention networks for person re-identification," *arXiv preprint*, 2016.
[15] H. Shi, Y. Yang, X. Zhu, S. Liao, Z. Lei, W. Zheng, and S. Z. Li, "Embedding deep metric for person re-identification: A study against large variations," in *ECCV*, 2016.
[16] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zhang, "Joint learning of single-image and cross-image representations for person re-identification," in *CVPR*, 2016.
[17] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE TKDE*, 2010.
[18] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *NIPS*, 2014.
[19] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in *CVPR*, 2014.
[20] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *ICCV*, 2015.
[21] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE International Workshop on PETS*, 2007.
[22] X. Wu, R. He, and Z. Sun, "A light CNN for deep face representation with noisy labels," *CoRR*, 2015.
[23] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *COLT*, 1998.
[24] W. Wang and Z. Zhou, "Analyzing co-training style algorithms," in *ECML*, 2007.
[25] H. Hu, Z. Lin, J. Feng, and J. Zhou, "Smooth representation clustering," in *CVPR*, 2014.
[26] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE TPAMI*, 2016.
[27] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Scandinavian conference on Image analysis*, 2011.
[28] W. Li and X. Wang, "Locally aligned feature transforms across views," in *CVPR*, 2013.
[29] Y. Zhang, B. Li, H. Lu, A. Irie, and X. Ruan, "Sample-specific svm learning for person re-identification," in *CVPR*, 2016.
[30] R. R. Varior, B. Shuai, J. Lu, D. Xu, and G. Wang, "A siamese long short-term memory architecture for human re-identification," in *ECCV*, 2016.
[31] N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Temporal model adaptation for person re-identification," in *ECCV*, 2016.
[32] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Person re-identification by unsupervised $\ell 1$ graph learning," in *ECCV*, 2016.
[33] P. Peng, T. Xiang, Y. Wang, M. Pontil, S. Gong, T. Huang, and Y. Tian, "Unsupervised cross-dataset transfer learning for person re-identification," in *CVPR*, 2016.