



# Two-stream person re-identification with multi-task deep neural networks

Liang Hu<sup>1</sup> · Chaoqun Hong<sup>1</sup> · Zhiqiang Zeng<sup>1</sup> · Xiaodong Wang<sup>1</sup>

Received: 30 October 2017 / Revised: 29 December 2017 / Accepted: 18 January 2018 / Published online: 1 March 2018  
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

## Abstract

Person re-identification (re-id) with images is very useful in video surveillance to find specific targets. However, it is challenging due to the complex variations of human poses, camera viewpoints, lighting, occlusion, resolution, background clutter and so on. The key to tackle this problem is how to represent the body and match these representations among frames. Current methods usually use the features of the whole bodies, and the performance may be reduced because of part invisibility. To solve this problem, we propose a two-stream strategy to use parts and bodies simultaneously. It utilizes a multi-task learning framework with deep neural networks (DNNs). Part detection and body recognition are performed as two tasks, and the features are extracted by two DNNs. The features are connected to multi-task learning to compute the mapping model from features to identifications. With this model, re-id can be achieved. Experimental results on a challenging task show the effectiveness of the proposed method.

**Keywords** Person re-identification · Multi-task learning · Deep learning

## 1 Introduction

Vision-based person re-identification (re-id) aims at identifying a target person with a gallery of pedestrian images. It is important in many video surveillance applications, such as finding criminals, cross-camera person tracking, and person activity analysis. This problem is challenging because of complex variations of human poses, camera viewpoints, lighting, occlusion, resolution, background clutter, etc., and thus draws plenty of research attention in recent years [28,29,51].

Traditional routine of the person re-id methods consists of two stages: detection and recognition.

1. Person detection tries to find people in images. Histograms of oriented gradients (HOG) proposed by Dalal et al. have been proven success in the past few years [7]. Recently, researchers focus on complex situations such as

occlusion [20], crowded scenarios [15], mobile cameras [21]. In addition, people also try to tackle this problem using multiple features [19,35,38–41].

2. Person recognition tries to match the detected person with a specific target in the dataset. In these two stages, person recognition is more difficult. Even with perfect detecting results, recognition is still challenging. Therefore, researchers devote themselves into different aspects of this problem, such as features [27,37], metrics [25], and matching [18].

Similar to many other applications of computer vision, the key to re-id is finding descriptive features. To achieve it, deep learning has been used. Yi et al. [36] and Li et al. [16] both employ a siamese neural network [3] to determine whether a pair of input images belong to the same ID. The reason for choosing the siamese model is probably that the number of training samples for each identity is limited (usually two).

Existing person re-identification benchmarks and methods mainly focus on matching cropped pedestrian images between queries and candidates. However, it is different from real-world scenarios where the annotations of pedestrian bounding boxes are unavailable and the target person needs to be searched from a gallery of whole scene images. Recently, some researchers propose end-to-end methods to tackle the

✉ Chaoqun Hong  
cqhong@xmut.edu.cn

Liang Hu  
lhu@xmut.edu.cn

<sup>1</sup> Xiamen University of Technology, Ligong Road #600, Houxi, Jimei, Xiamen, Fujian, China

re-id problem. Xiao et al. use fasterRCNN framework [31]. Zheng et al. discuss the relationship between detection and recognition [45–47]. In this way, they try to find the best combination of detectors and recognizers.

Although numerous person re-id datasets and methods have been proposed, there is still a big gap between the problem setting itself and real-world applications. In most benchmarks, the gallery only contains manually cropped pedestrian images, while in real applications, the goal is to find a target person in a gallery of whole scene images. Following the protocols of these benchmarks, most of the existing person re-id methods assume perfect pedestrian detections. However, these manually cropped bounding boxes are unavailable in practical applications. Off-the-shelf pedestrian detectors would inevitably produce false alarms, misdetections, and misalignments, which could harm the final searching performance significantly.

To tackle the above problem, we propose a method to combine body part detection and person recognition. Specifically, we design a novel deep architecture named multi-task deep neural networks (MDNN) for person re-id. Different from existing methods, the proposed method defines part detection and person recognition as two tasks. In this way, it applies deep neural network (DNN)-based feature extraction to represent pedestrian images and multi-task learning based modal to construct the mapping relationship from images to identification. The contributions of this paper are summarized below:

1. First, we propose a new multi-task learning framework based on Deep Neural Network (DCNN). In this framework, DCNN-based feature mapping and multi-task learning are connected to obtain a DNN-based regression for re-id, which unified multimodal problem in a single model.
2. Second, in the proposed framework, re-id consists of two tasks. They are part detection and person recognition. In this way, a two-stream network is applied and the outputs of these two tasks are unified to compute the results.
3. Finally, the proposed framework is naturally multimodal. We conduct comprehensive experiments to on a challenging dataset [31]. The experimental results validate the effectiveness of our method.

The remainder of this paper is organized as follows. Related works on multi-task learning and deep learning are reviewed in Sect. 2. Then, the proposed MDNN-based re-id method is presented in Sect. 3. After that, we demonstrate the effectiveness of MDNN by experimental comparisons with other state-of-the-art methods in Sect. 4. We conclude in Sect. 5.

## 2 Related works

### 2.1 Multi-task learning

Multi-task learning has recently been employed in image classification [43], visual tracking [44], multi-view action recognition [34], egocentric daily activity recognition [33], and image privacy protection [42]. Given a set of related tasks, MTL [4] seeks to simultaneously learn a set of task-specific classification or regression models. The intuition behind MTL is that a joint learning procedure accounting for task relationships is more efficient than learning each task separately. Traditional MTL methods [1,8] assume that all the tasks are related and their dependencies can be modeled by a set of latent variables. However, in many real-world applications, not all tasks are related, and enforcing erroneous (or nonexistent) dependencies may lead to negative knowledge transfer.

Recently, sophisticated methods have been introduced to counter this problem. These methods assume a-priori knowledge (e.g., a graph) defining task dependencies [6], or learn task relationships in conjunction with task-specific parameters [9,11,14,48,49]. Among these, our work is most similar to [6].

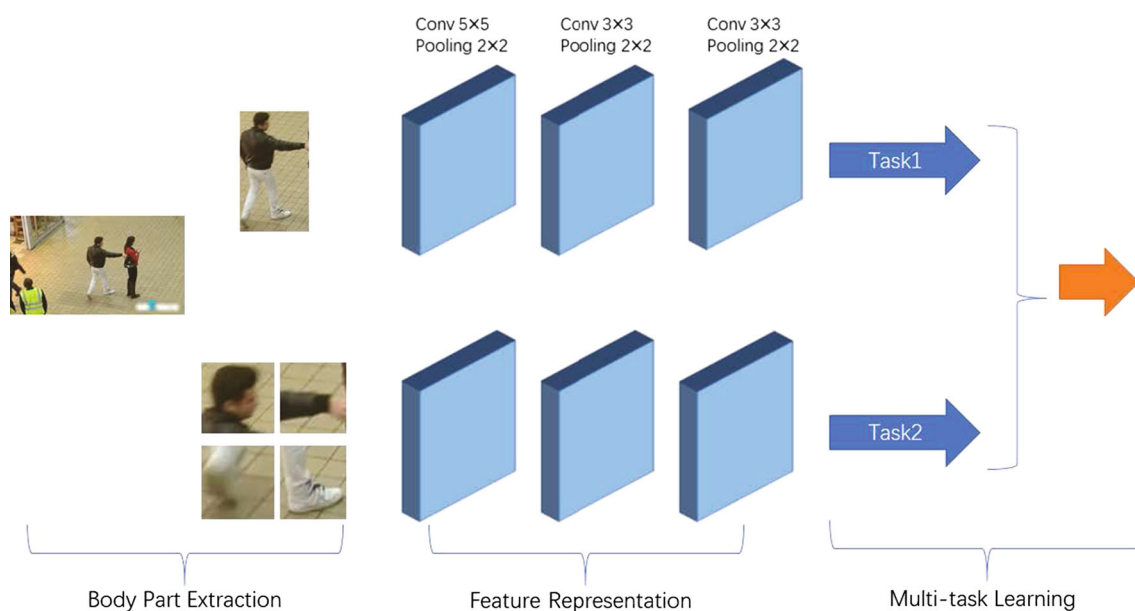
### 2.2 Deep learning

Feature description of images is critical to image-based analysis [22,32]. To obtain descriptive representation, deep learning architectures [23,24] have been efficient in exploring hidden representations in natural images and have achieved proven success in a variety of vision applications. For example, an autoencoder [2] is an efficient unsupervised feature learning method in which the internal layer acts as a generic extractor of inner image representations. A double-layer structure, which efficiently maps the input data onto appropriate outputs, is obtained by using a multilayer perceptron. In addition, deep learning can exploit parallel GPU computation and deliver high speeds in the forward pass. These advantages make deep models an attractive option for handling the re-id problem.

## 3 Multi-task deep neural networks

### 3.1 Overview of the proposed method

The flowchart of the proposed method is shown in Fig. 1. The training process consists of two tasks. They are part detection and person recognition. They are multimodal and trained in two deep neural networks. Then, the output of DNNs is connected to a multi-task regression learning. With the trained model, re-id can be achieved.



**Fig. 1** Flowchart of the proposed method is shown. Generally speaking, it consists of three stages. First, we extract parts from the images. Second, we use deep neural networks to compute features. Finally, we

connect multi-task learning with neural networks to compute the regression model for re-id. With this model, we can map image features to the person identities

### 3.2 The framework of multi-task learning

As mentioned before, the traditional routine to achieve re-id is mapping images to people with pre-computed regression models, specifically the key to define a reasonable loss function between the estimation and the groundtruth while training. Therefore, we aim at computing a well-defined regression model. In data mining and machine learning, a common paradigm for classification and regression is to minimize the penalized empirical loss:

$$\arg \min_W \ell(W) + \Phi(W), \tag{1}$$

where  $W$  is the parameter to be estimated from the training samples,  $\ell(W)$  is the loss function, and  $\Phi(W)$  is the regularization term that encodes task relatedness.

In our application, two-stream re-id with  $V$  modals can be considered as a multi-task process with  $V$  tasks while  $V = 2$ . The training data for  $v$ -th task can be denoted by  $x_i^v, y_i^v$ , where  $v = 1, 2, i = 1, \dots, N$  and  $N$  is the number of samples.  $X = x_i^v R^{d1}$  are image features, where  $d1$  is the dimension. The goal of multi-task learning can be defined as:

$$\arg \min \sum_{v=1}^V \sum_{i=1}^N \ell(y_i^v, f(x_i^v; w^v)) + \Phi(w^v), \tag{2}$$

where  $f(x_i^v; w^v)$  is a function of  $x_i^v$  and parameterized by a weight vector  $w^v$ . There are several existing choices of  $\ell(\cdot)$ .

- Trace norm regularized learning with least squares loss (LeastTrace) [12]: the loss function is defined as:

$$\arg \min_W \sum_{i=1}^t (0.5 * \text{norm}(Y_i - X_i' * W(:, i))^2) + \rho_1 \| W \|_*, \tag{3}$$

where  $\| W \|_* = \sum(\text{SVD}(W, 0))$  is the trace norm.

- L21 joint feature learning with least squares loss (LeastL21) [1]: the loss function is defined as:

$$\arg \min_W \sum_{i=1}^t (0.5 * \text{norm}(Y_i - X_i' * W(:, i))^2) + \text{opts}.\rho_{L2} * \| W \|_2^2 + \rho_1 \| W \|_{2,1}. \tag{4}$$

- Sparse structure-regularized learning with least squares loss (LeastLasso) [26]: the loss function is defined as:

$$\arg \min_W \sum_{i=1}^t (0.5 * \text{norm}(Y_i - X_i' * W(:, i))^2) + \text{opts}.\rho_{L2} * \| W \|_F^2 + \rho_1 \| W \|_1. \tag{5}$$

- Incoherent sparse and low-rank learning with least squares loss (Least-SparseTrace) [5]: the loss function is defined as:

$$\begin{aligned} \arg \min_W \sum_{i=1}^t (0.5 * \text{norm}(Y_i - X_i' * W(:, i))^2) \\ + \gamma * \| P \|_1 \text{.subject to: } W = P + Q, \| Q \|_* \leq \tau \end{aligned} \quad (6)$$

where  $\| Q \|_* = \sum(\text{SVD}(Q, 0))$  is the trace norm.

According to existing reports and our experiments, Least-SparseTrace outperforms the other loss functions. Due to the sparse constraints, Least-SparseTrace can improve the descriptive ability with features from different tasks.  $\Phi(w^v)$  is the regularization term that penalizes the complexity of weights. In this way, the objective function can be rewritten as:

$$\arg \min_W \frac{1}{2} \sum_{v=1}^V (0.5 * \text{norm}(Y - \mathcal{F}(X^v; W^v))^2) + \sum_{v=1}^V \| W^v \|^2, \quad (7)$$

where  $W = w^v$  is the weighted matrix with the same meaning as Eq. (1). To solve the above function, the key is how to define an optimized regression function  $\mathcal{F}(X^v; W^v)$ .

### 3.3 Deep neural network based regression

Deep neural networks (DNNs) has been proven success in image description, especially with multi-task learning [17]. In our method, we solve  $f(\cdot)$  by using DNNs. In DNNs, this function is called the activation function. In computational networks, the activation function of a node defines the output of that node given an input or set of inputs. In the scenario of the deep neural network, activation functions project  $x_i^v$  to higher level representation gradually by learning a sequence of nonlinear mappings, which can be defined as:

$$(x_i^v)^0 \xrightarrow{\mathcal{R}_w} (x_i^v)^1 \xrightarrow{\mathcal{R}_w} \dots \xrightarrow{\mathcal{R}_w} (x_i^v)^l, \quad (8)$$

where  $l$  is the number of layers and  $\mathcal{R}$  is the mapping function from input to estimated output.

To optimize the weighted matrix  $W$  which contains the mapping parameters, we use a back-propagation strategy. For each echo of this process, the weighted matrix is updated by  $\Delta W$ , which is defined by:

$$\Delta W = -\eta \frac{\partial E}{\partial W} \quad (9)$$

$\eta$  is the learning rate and:

$$\frac{\partial E}{\partial W} = (y_i^v - \mathcal{R}(x_i^v))(x_i^v)^T, \quad (10)$$

In this way, we try to minimize the differences between the groundtruth  $y_i^v$  and the estimated output  $\mathcal{R}(x_i^v)$ . The back-propagation strategy can be modeled by:

$$(x_i^v)^0 \xleftarrow{\frac{\mathcal{R}}{W}} (x_i^v)^1 \xleftarrow{\frac{\mathcal{R}}{W}} \dots \xleftarrow{\frac{\mathcal{R}}{W}} (x_i^v)^l. \quad (11)$$

### 3.4 Implementation details

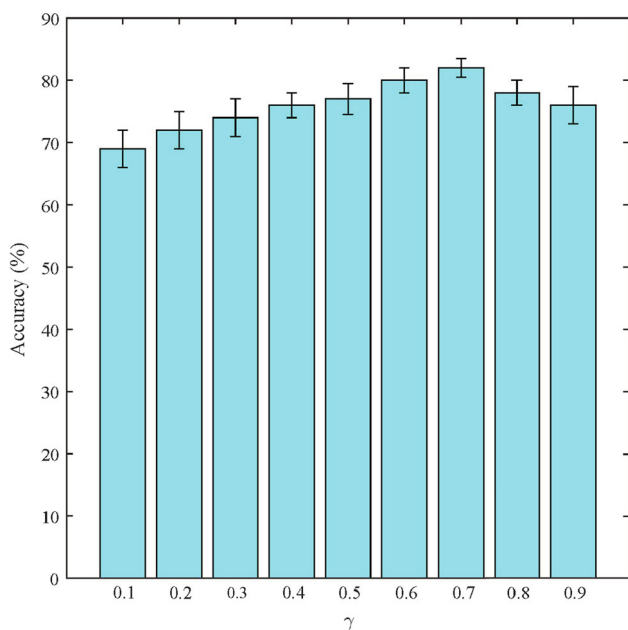
In most of the re-id datasets, the body parts are not labeled. To get them, we use EdgeBox to automatically extract body parts [52]. If the detected parts are located inside the labeled bounding boxes of people, we add them in the body part stream. To get a uniform representation, all the body parts are cropped and normalized to  $16 \times 16$ . On the other hand, the people within the bounding boxes are used in the whole person stream. Then, the features of body parts and whole bodies are extracted in two networks. These two DNNs contain three hidden layers, and the parameters are shown in Fig. 1. DNNs are implemented based on Caffe [13]. Finally, the outputs of DNNs are used as features of multi-task learning and the final re-id decision is made. Multi-task learning is implemented based on MALSAR [50]. Besides, the experiments are conducted on a workstation equipped with 4 Titan X (Pascal).

## 4 Experimental results and discussion

### 4.1 Dataset and settings

In our paper, we use the dataset proposed by Person Search [31]. In this dataset, a large-scale person search image set is collected and annotated. Two data sources are exploited to diversify the scenes. On one hand, hand-held cameras are used to shoot street snaps around an urban city. On the other hand, images are collected from movie snapshots that contain pedestrians, as they could enrich the variations of viewpoints, lighting, and background conditions. In the 18184 images of this dataset, 96143 bounding boxes of pedestrians are annotated. Then, the same person that appears across different images is associated, resulting in 8432 identities.

In the experiment, the dataset is split into a training subset and a test subset, ensuring no overlapped images or labeled identities between them. The test identity instances are further divided into queries and galleries. For each of the 2900 test identities, one of his/her instances are randomly chosen as the query, while the corresponding gallery set consists of two parts all the images containing the other instances and some randomly sampled images not containing this person. Different queries have different galleries, and jointly they cover all the 6978 test images. This process is repeated 20 times, and the average results are recorded.



**Fig. 2** Performance with different values of  $\gamma$

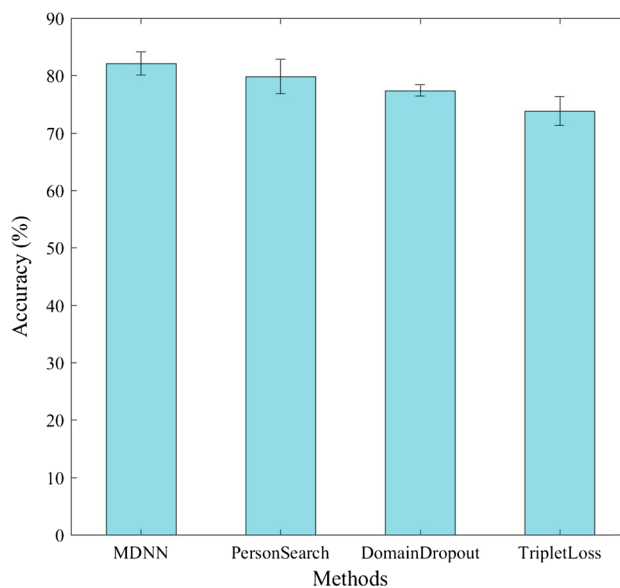
To evaluate the performance, we employ cumulative matching characteristics (CMC top-K). The first one is inherited from the person re-id problem, where a matching is counted if there is at least one of the top-K predicted bounding boxes overlaps with the ground truths with intersection-over-union (IoU) greater or equal to 0.5. To simplify the demonstration, we use top-one as the evaluation metric.

## 4.2 Experimental results

In the proposed multi-task learning framework,  $\gamma$  in Eq. (6) may influence the performance. Therefore, we try different settings of  $\gamma$  to optimize the performance, which is shown in Fig. 2. We can figure out that the best performance is achieved when  $\gamma = 0.7$  and this setting is used in the following experiments.

In our experiments, we refer to several state of the arts and the performance is compared with them. They are PersonSearch, DomainDropout, and TripletLoss. They are all based on deep learning and achieve the best performance on the dataset we use. Besides, the theoretical contributions of them and the proposed method are similar.

1. PersonSearch [31]: instead of breaking it down into two separate tasks pedestrian detection and person re-identification, we jointly handle both aspects in a single convolutional neural network. An online instance matching (OIM) loss function is proposed to train the network effectively, which is scalable to datasets with numerous identities.



**Fig. 3** Comparison of re-id accuracy

2. DomainDropout [30]: different from the standard Dropout, which treats all the neurons equally, our method assigns each neuron a specific dropout rate for each domain according to its effectiveness on that domain.
3. TripletLoss [10]: firstly we introduce variants of the classic triplet loss which render mining of hard triplets unnecessary and we systematically evaluate these variants. And secondly, we show how, contrary to the prevailing opinion, using a triplet loss and no special layers, we achieve state-of-the-art results both with a pre-trained CNN and with a model trained from scratch.

The performance of different methods is shown in Fig. 3. We can clearly figure out the following conclusion:

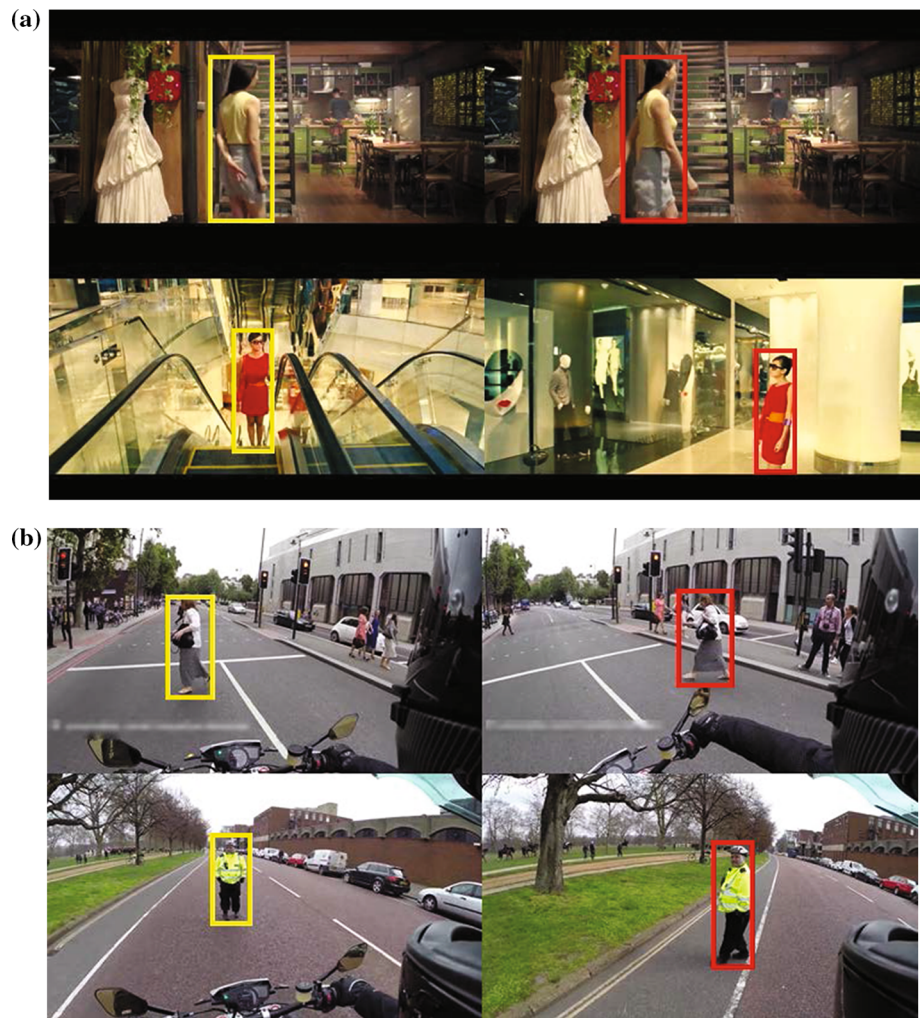
1. The proposed MDNN achieves the best performance.
2. DomainDropout can produce stable achievement.
3. PersonSearch is able to give a good performance, but the stabilities of PersonSearch and TripletLoss are lower than the other methods.

Some of the typical re-id results are shown in Fig. 4. The proposed method is effective for re-id.

## 5 Conclusion

In this paper, we propose a novel person re-identification method. It improves previous methods by employing deep learning and multi-task learning. First, we define the task of re-id as the combination of two tasks. They are part detection and person recognition. In this way, a two-stream strategy is

**Fig. 4** Some sample results of re-id are shown



designed. Second, to train the mapping model from images to identifications, we propose a multi-task learning framework based on deep neural network (DCNN). In this framework, DCNN-based feature mapping and multi-task learning are connected to obtain a DNN-based regression for re-id. Second, experimental results show that the proposed method outperforms state-of-the-art methods of re-id.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China (61622205, 61472110), the Fujian Provincial Natural Science Foundation of China (2016J01327, 2016J01324), the Fujian Provincial High School Natural Science Foundation of China (JZ160472), and Foundation of Fujian Educational Committee (JAT160357, JAT160358).

## References

- Argyriou, A., Evgeniou, T., Pontil, M.: Multi-task feature learning. In: Conference on Advances in Neural Information Processing Systems, pp. 41–48 (2007)
- Bengio, Y., et al.: Learning deep architectures for ai. *Foundations and trends* ©. *Mach. Learn.* **2**(1), 1–127 (2009)
- Bromley, J., Guyon, I., Lecun, Y., Sackinger, E., Shah, R.: Signature verification using a "siamese" time delay neural network. In: International Conference on Neural Information Processing Systems, pp. 737–744 (1993)
- Caruana, R.: Multitask learning. *Mach. Learn.* **28**(1), 41–75 (1997)
- Chen, J., Liu, J., Ye, J.: Learning incoherent sparse and low-rank patterns from multiple tasks. *ACM Trans. Knowl. Discov. Data* **5**(4), 22 (2012)
- Chen, X., Lin, Q., Kim, S., Carbonell, J.G., Xing, E.P.: Smoothing proximal gradient method for general structured sparse learning. In: Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, pp. 105–114 (2011)
- Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886–893. IEEE (2005)
- Evgeniou, T., Pontil, M.: Regularized multi-task learning. In: International Conference on Knowledge Discovery and Data Mining, pp. 109–117 (2004)
- Gong, P., Ye, J., Zhang, C.: Robust multi-task feature learning. In: International Conference on Knowledge Discovery & Data Mining, pp. 895–903 (2012)
- Hermans, A., Beyler, L., Leibe, B.: In defense of the triplet loss for person re-identification. arXiv preprint [arXiv:1703.07737](https://arxiv.org/abs/1703.07737) (2017)
- Jalali, A., Sanghavi, S., Ruan, C., et al.: A dirty model for multi-task learning. In: Lafferty, J.D., Williams, C.K.I., Shawe-Taylor,

- J., Zemel, R.S., Culotta, A. (eds.) *Advances in Neural Information Processing Systems*, pp. 964–972. Curran Associates, Inc. (2010)
12. Ji, S., Ye, J.: An accelerated gradient method for trace norm minimization. In: *International Conference on Machine Learning*, pp. 457–464 (2009)
  13. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093* (2014)
  14. Kang, Z., Grauman, K., Sha, F.: Learning with whom to share in multi-task feature learning. In: *International Conference on Machine Learning*, pp. 521–528 (2011)
  15. Karpagavalli, P., Ramprasad, A.V.: An adaptive hybrid gmm for multiple human detection in crowd scenario. *Multimed. Tools Appl.* **76**, 1–21 (2016)
  16. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: deep filter pairing neural network for person re-identification. In: *Computer Vision and Pattern Recognition*, pp. 152–159 (2014)
  17. Li, X., Zhao, L., Wei, L., Yang, M.H., Wu, F., Zhuang, Y., Ling, H., Wang, J.: Deepsaliency: multi-task deep neural network model for salient object detection. *IEEE Trans Image Process* **25**(8), 3919 (2016)
  18. Lin, W., Shen, Y., Yan, J., Xu, M., Wu, J., Wang, J., Lu, K.: Learning correspondence structures for person re-identification. *IEEE Trans Image Process* **26**(5), 2438–2453 (2017)
  19. Liu, W., Yang, X., Tao, D., Cheng, J., Tang, Y.: Multiview dimension reduction via hessian multiset canonical correlations. *Inf Fusion* **41**, 119–128 (2017)
  20. Mar, N.J., Vazquez, D., Lopez, A.M., Amores, J., Kuncheva, L.I.: Occlusion handling via random subspace classifiers for human detection. *IEEE Trans. Cybern.* **44**(3), 342–354 (2017)
  21. Miseikis, J., Borges, P.V.K.: Joint human detection from static and mobile cameras. *IEEE Trans. Intell. Transp. Syst.* **16**(2), 1018–1029 (2015)
  22. Sang, J., Xu, C., Liu, J.: User-aware image tag refinement via ternary semantic analysis. *IEEE Trans. Multimed.* **14**(3), 883–895 (2012)
  23. Shao, L., Wu, D., Li, X.: Learning deep and wide: a spectral method for learning deep networks. *IEEE Trans. Neural Netw. Learn. Syst.* **25**(12), 2303–2308 (2014)
  24. Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
  25. Sun, C., Wang, D., Lu, H.: Person re-identification via distance metric learning with latent variables. *IEEE Trans. Image Process.* **26**(1), 23–34 (2016)
  26. Tibshirani, R.: Regression shrinkage and selection via the lasso: a retrospective. *J. R. Stat. Soc.* **73**(3), 267–288 (2011)
  27. Varior, R.R., Wang, G., Lu, J., Liu, T.: Learning invariant color features for person reidentification. *IEEE Trans. Image Process.* **25**(7), 3395–3410 (2016)
  28. Xiao, F., Liu, W., Li, Z., Chen, L., Wang, R.: Noise-tolerant wireless sensor networks localization via multi-norms regularized matrix completion. *IEEE Trans. Veh. Technol.* **PP**(99), 1–1 (2017)
  29. Xiao, F., Wang, Z., Ye, N., Wang, R., Li, X.Y.: One more tag enables fine-grained rfid localization and tracking. *IEEE ACM Trans. Netw.* **PP**(99), 1–14 (2017)
  30. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification. In: *CVPR* (2016)
  31. Xiao, T., Li, S., Wang, B., Lin, L., Wang, X.: Joint detection and identification feature learning for person search. In: *CVPR* (2017)
  32. Xu, C.: Exploiting social-mobile information for location visualization. *ACM* **8**, 39 (2017)
  33. Yan, Y., Ricci, E., Liu, G., Sebe, N.: Egocentric daily activity recognition via multitask clustering. *IEEE Trans. Image Process.* **24**(10), 2984–2995 (2015)
  34. Yan, Y., Ricci, E., Subramanian, R., Liu, G., Sebe, N.: Multitask linear discriminant analysis for view invariant action recognition. *IEEE Trans. Image Process.* **23**(12), 5599–5611 (2014)
  35. Yang, X., Liu, W., Tao, D., Cheng, J.: Canonical correlation analysis networks for two-view image recognition. *Inf. Sci.* **385**(C), 338–352 (2017)
  36. Yi, D., Lei, Z., Liao, S., et al.: Deep metric learning for person re-identification. In: *ICPR '14 Proceedings of the 2014 22nd International Conference on Pattern Recognition*, pp. 34–39. IEEE Computer Society, Washington, DC, USA (2014)
  37. Yogarajah, P., Chaurasia, P., Condell, J., Prasad, G.: Enhancing gait based person identification using joint sparsity model and -norm minimization. *Inf. Sci.* **308**, 3–22 (2015)
  38. Yu, J., Rui, Y., Chen, B.: Exploiting click constraints and multi-view features for image re-ranking. *IEEE Trans. Multimed.* **16**(1), 159–168 (2013)
  39. Yu, J., Rui, Y., Tao, D.: Click prediction for web image reranking using multimodal sparse coding. *IEEE Trans. Image Process.* **23**(5), 2019–32 (2014)
  40. Yu, J., Tao, D., Wang, M., Rui, Y.: Learning to rank using user clicks and visual features for image retrieval. *IEEE Trans. Cybern.* **45**(4), 767–779 (2015)
  41. Yu, J., Yang, X., Fei, G., Tao, D.: Deep multimodal distance metric learning using click constraints for image ranking. *IEEE Trans. Cybern.* **PP**(99), 1–11 (2016)
  42. Yu, J., Zhang, B., Kuang, Z., Lin, D., Fan, J.: Iprivacy: image privacy protection by identifying sensitive objects via deep multi-task learning. *IEEE Trans. Inf. Forensics Secur.* **12**(5), 1005–1016 (2017)
  43. Yuan, X.T., Liu, X., Yan, S.: Visual classification with multitask joint sparse representation. *IEEE Trans. Image Process.* **21**(10), 4349–4360 (2012)
  44. Zhang, T., Ghanem, B., Liu, S., Ahuja, N.: Robust visual tracking via structured multi-task sparse learning. *Int. J. Comput. Vis.* **101**(2), 367–383 (2013)
  45. Zheng, L., Huang, Y., Lu, H., Yang, Y.: Pose invariant embedding for deep person re-identification. *CoRR abs/1701.07732* (2017). <http://arxiv.org/abs/1701.07732>
  46. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: past, present and future (2016)
  47. Zheng, L., Zhang, H., Sun, S., Chandraker, M., Yang, Y., Tian, Q.: Person re-identification in the wild (2016)
  48. Zhong, W., Kwok, J.: Convex multitask learning with flexible task clusters. In: *International Conference on Machine Learning*, pp. 49–56 (2012)
  49. Zhou, J., Chen, J., Ye, J.: Clustered multi-task learning via alternating structure optimization. In: *Advances in Neural Information Processing Systems*, p. 702 (2011)
  50. Zhou, J., Chen, J., Ye, J.: Malsar: multi-task learning via structural regularization, vol. 21. Arizona State University, Tempe (2011)
  51. Zhu, H., Xiao, F., Sun, L., Wang, R., Yang, P.: R-ttwd: robust device-free through-the-wall detection of moving human with wifi. *IEEE J. Sel Areas Commun.* **PP**(99), 1–1 (2017)
  52. Zitnick, C.L., Dollr, P.: Edge boxes: locating object proposals from edges. In: *European Conference on Computer Vision*, pp. 391–405 (2014)

**Liang Hu** received the M.S. degree from Changchun University of Technology, Changchun, China, in 2011. She is currently a Lecturer in the Department of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China. Her research interests include high-performance computing, pattern recognition, image processing and data mining.

**Chaoqun Hong** received the Ph.D. degree from Zhejiang University, Zhejiang, China, in 2011. He is currently an Associate Professor in the Department of Computer and Information Engineering, Xiamen University of Technology, Xiamen, China. His research interests include high-performance computing, image processing, computer vision, and pattern recognition. He has authored or coauthored more than 30 scientific articles.

**Zhiqiang Zeng** received M.Sc. and Ph.D. degrees in Computer Science from Xi'an Jiaotong University and Zhejiang University, China, in 2004 and 2007, respectively. He holds a Bachelor degree in Automation from Sichuan University, Chian, in 1994. In 2008, he joined the Computer Science Department of the Xiamen University of Technology as a Research Associate. His interests include pattern recognition and machine learning, in particular, support vector machines and general kernel methods.

**Xiaodong Wang** received his B.E and M.E. degree in College of Computer Science and Electronic Engineering from Hunan University, China, in 2007 and 2010. He is currently a lecturer in the college of Computer and Information Engineering, Xiamen University of Technology. His research interests cover pattern recognition, image processing and embedded system structure.