A Study on Deep Convolutional Neural Network Based Approaches for Person Re-identification

Harendra Chahar(✉) and Neeta Nain

Malaviya National Institute of Technology, Jawahar Lal Nehru Margh, Jaipur, India
hchahar616@gmail.com, nnain.cse@mnit.ac.in

Abstract. Person re-identification is a process to identify the same person again viewed by disjoint field of view of cameras. It is a challenging problem due to visual ambiguity in a person's appearance across different camera views. These difficulties are often compounded by low resolution surveillance images, occlusion, background clutter and varying lighting conditions. In recent years, person re-identification community obtained large size of annotated datasets and deep learning architecture based approaches have obtained significant improvement in the accuracy over the years as compared to hand-crafted approaches. In this survey paper, we have classified deep learning based approaches into two categories, i.e., image-based and video-based person re-identification. We have also presented the currently ongoing under developing works, issues and future directions for person re-identification.

Keywords: Person re-identification · Convolutional neural network

1 Introduction

In automated multi-camera video-surveillance, person re-identification is defined as whether the same person has been already observed at another place by different camera field of view. It is used for behaviour recognition, person tracking, image retrieval and safety purpose at public place. For humans to manually monitor video-surveillance systems to identify a probe accurately and efficiently is a difficult task. It is vary challenging problem due to variation in a person's appearance across different cameras. Therefore, person observed at multi-camera views have small inter-class variations and large ambiguities in intra-class variations.

For person re-identification, few surveys have been already exist [1–4]. In recent years, the availability of large size annotated person re-identification datasets and great success of deep learning in computer vision for image classification and object recognition also have made great influence in person re-identification. In this survey paper, we have presented the deep learning based approaches for person re-identification on both image and video datasets.

Section 2 present various deep learning approaches for person re-identification on image datasets. Section 3 describes different types of deep learning approaches for person re-identification on video datasets and various currently ongoing issues and future works. In Sect. 4, we have drawn conclusion.

B.U. Shankar et al. (Eds.): PReMI 2017, LNCS 10597, pp. 543–548, 2017.
https://doi.org/10.1007/978-3-319-69900-4_69

2 Deep Learning Based Person Re-identification Approaches on Image Datasets

In year 2012, convolutional neural network based deep learning model has been presented by Krizhevsky et al. [7] in ILSVRC'12 competition. They won this competition with a large margin in accuracy. Since then convolutional neural network based deep learning models have been becomes more popular in computer vision comunity. Yi et al. [5] have been proposed a deep metric learning approach for person re-identification using a siamese convolutional neural network with a symmetry structure comprising two sub-networks connected by a cosine layer. A pair of images is used as a input, extracts features from each image separately and then uses their cosine distance for similarity matching. In [6] authors have been proposed a siamese architecture wherein a patch-matching layer is used which multiplies convolutional feature responses from the two inputs at a variety of horizontal stripes and uses product to compute patch similarity in similar latitude. Varior et al. [8] have been presented a method by inserting a gating function after each convolutional layer into the network to find effective subtle patterns in testing of paired images. In [9], a soft attention based model has been integrated with a siamese neural network to adaptively focus on the important local parts of paired input images. Cheng et al. [10] have been presented a triplet loss function, wherein a triplet of three images as input has been created. Each image is partitioned into four overlapping body parts after the first convolutional layer and fusion of all as a final one has been done in the fully-connected layer. In [12] authors have proposed a pipeline for learning generic feature representations from multiple domains. They combine all the datasets together and train a designed convolutional neural network from scratch on combined dataset and a softmax loss is used in the classification. In [13] authors has presented an approach wherein they construct a single fisher vector [14] for each image by using SIFT and color histograms aggregation. They have used fisher vectors as a input and build a fully connected network and linear discriminative analysis is used as an objective function. In [22] authors have proposed a deep

Table 1. Statistics of benchmark image datasets for person re-identification

Dataset	Time	#ID	#Image	#Camera	Label
VIPeR [18]	2007	632	1264	2	Hand
iLIDS [17]	2009	119	476	2	Hand
GRID [19]	2009	250	1275	8	Hand
CUHK01 [20]	2012	971	3884	2	Hand
CUHK02 [21]	2013	1816	7264	10	Hand
CUHK03 [6]	2014	1467	13164	2	Hand/DPM
PRID 450S [34]	2014	450	900	2	Hand
Market-1501 [32]	2015	1501	32668	6	Hand/DPM

Table 2. Rank-1 accuracy of different deep learning approaches for person re-identification on various image datasets, i.e., (VIPeR, CUHK-01, CUHK-03, PRID, iLIDS and Market-1501)

Authors/Year	Evolution	VIPeR	CUHK-01	CUHK-03	PRID	iLIDS	Market-1501
Yi [5] (2014)	CMC	28.23%	–	–	–	–	–
Li [6] (2014)	CMC	–	27.87%	20.65%	–	–	–
Wu [15] (2016)	CMC/mAP	–	71.14%	64.80%	–	–	37.21%
Xiao [12] (2016)	CMC	38.6%	66.6%	75.33%	64.0%	64.6%	–
Chi-Su [11] (2016)	CMC/mAP	43.5%	–	–	22.6%	–	39.4%
Liu [9] (2016)	CMC/mAP	–	81.04%	65.65%	–	–	48.24%
Varior [8] (2016)	CMC/mAP	37.8%	–	68.1%	–	–	65.88%
Wang [16] (2016)	CMC	35.76%	71.80%	52.17%	–	–	–
Geng [22] (2016)	CMC/mAP	**56.3**%	–	**85.4**%	–	–	**83.7**%

transfer learning approach wherein one stepped fine-tuning for large person re-identification datasets (Imagenet → Market-1501) and two stepped fine-tuning for small datasets (Imagenet → Market-1501 → VIPeR) have been used. We have taken all the result from existing approaches and observed overwhelming advantage of deep learning [22] in rank-1 accuracy on largest datasets CUHK03 and Market-1501 so far (Tables 1 and 2).

3 Deep Learning Based Person Re-identification Approaches on Video Dataset

The deep learning approaches for person re-identification on video datasets are [23, 25, 31] wherein appearance features have been used as the starting point into RNN to obtain the time flow information between frames. McLaughlin et al. [31] have been presented a framework wherein convolutional neural network is used to extract features from consecutive video frames and fedded through a recurrent final layer. In [23] authors have proposed the gated recurrent unit and an identification loss based recurrent neural network. Yan et al. [25] and Zheng et al. [33] have proposed a model in which each input video sequence is classifies into their respective subject by using the identification model. Color and local binary pattern features are fedded into LSTM cells. Wu et al. [24] has proposed a model to build a hybrid network by fusing color and LBP features to extract both spatial-temporal and appearance features from a video sequence. In [30] authors have presented a method to extract a compact and discriminative appearance features representation from selected frames based on flow energy profile instead of the whole sequence (Tables 3 and 4).

Computer vision community is always looking for annotated large size datasets for supervised learning. This is a challenging problem in person re-identification. Assigning an id to a pedestrian is not trivial. Open-world person re-identification can be viewed as a person verification task. Zheng et al. [35] has

Table 3. Statistics of benchmark video datasets for person re-identification

Dataset	Time	#ID	#Track	#Bbox	#Camera	Label
ETHZ [26]	2007	148	148	8580	1	Hand
3DPES [27]	2011	200	1000	200 k	8	Hand
PRID-2011 [28]	2011	200	400	40 k	2	Hand
iLIDS-VID [29]	2014	300	600	44 k	2	Hand
MARS [33]	2016	1261	20715	1 M	6	DPM&GMMCP

Table 4. Rank-1 accuracy of deep learning based approaches for person re-identification on different datasets, i.e.,(iLIDS-VID and PRIQ-2011)

Authors/Year	Evaluation	iLIDS-VID	PRIQ-2011
Wu [23] (2016)	CMC	46.1%	69.0%
Yan [25] (2016)	CMC	49.3%	58.2%
McLaughlin [31] (2016)	CMC	58%	70%
Zhang [30] (2017)	CMC	**60.2**%	**83.3**%

been presented a method to achieve low false and high true target recognition. Liao et al. [36] has proposed a method having two stages, in the first stage, it finds whether a query subject is present in the gallery or not. In second stage, assigns an id to the accepted query subject. Open-world person re-identification is still challenging task as evidenced by the low recognition rate under low false accept rate as shown in [35,36]. Therefore, there is need to design an efficient methods to improve both accuracy and efficiency of the person re-id systems.

4 Conclusion

Increasing the demand of saftey at public places gain more interest for person re-identification. In this survey paper, we have presented deep learning approaches in both image and video datasets. Solving the data volume issue, re-identification re-ranking methods, and open world re-identification systems are some important open issues that may attract further attention from the community.

References

1. D'Orazio, T., Grazia, C.: People re-identification and tracking from multiple cameras: a review. In 19th IEEE International Conference on Image Processing (ICIP), pp. 1601–1604 (2012)
2. Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. Image Vis. Comput. **32**(4), 270–286 (2014)
3. Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.): Person Re-Identification. ACVPR, vol. 1. Springer, London (2014). doi:10.1007/978-1-4471-6296-4

4. Satta, R.: Appearance descriptors for person re-identification: a comprehensive review. arXiv preprint arXiv1307.5748 (2013)
5. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Deep metric learning for person re-identification. In: Proceedings of International Conference on Pattern Recognition, pp. 2666–2672 (2014)
6. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 152–159 (2014)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
8. Varior, R.R., Haloi, M., Wang, G.: Gated Siamese convolutional neural network architecture for human re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 791–808. Springer, Cham (2016). doi:10.1007/978-3-319-46484-8_48
9. Liu, H., Feng, J., Qi, M., Jiang, J., Yan, S.: End-to-end comparative attention networks for person re-identification, arXiv preprint arXiv:1606.04404 (2016)
10. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1335–1344 (2016)
11. Su, C., Zhang, S., Xing, J., Gao, W., Tian, Q.: Deep attributes driven multi-camera Person re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 475–491. Springer, Cham (2016). doi:10.1007/978-3-319-46475-6_30
12. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1249–1258 (2016)
13. Wu, L., Shen, C., van den Hengel, A.: Deep linear discriminant analysis on fisher networks: a hybrid architecture for person re-identification. Pattern Recognit. (2016)
14. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 143–156. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15561-1_11
15. Wu, L., Shen, C., Hengel, A.V.D.: Personnet: person re-identification with deep convolutional neural networks. arXiv preprint arXiv:1601.07255 (2016)
16. Wang, F., Zuo, W., Lin, L., Zhang, D., Zhang, L.: Joint learning of single-image and cross-image representations for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1288–1296 (2016)
17. Wei-Shi, Z., Shaogang, G., Tao, X.: Associating groups of people. In: Proceedings of the British Machine Vision Conference, pp. 23.1–23.11 (2009)
18. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008). doi:10.1007/978-3-540-88682-2_21
19. Loy, C.C., Xiang, T., Gong, S.: Multi-camera activity correlation analysis. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1988–1995. IEEE (2009)

20. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7724, pp. 31–44. Springer, Heidelberg (2013). doi:10.1007/978-3-642-37331-2_3
21. Li, W., Wang, X.: Locally aligned feature transforms across views. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3594–3601 (2013)
22. Geng, M., Wang, Y., Xiang, T., Tian, Y.: Deep transfer learning for person re-identification. arXiv preprint arXiv:1611.05244 (2016)
23. Wu, L., Shen, C., Hengel, A.V.D.: Deep recurrent convolutional networks for video-based person re-identification: an end-to-end approach. arXiv preprint arXiv:1606.01609 (2016)
24. Wu, Z., Wang, X., Jiang, Y.G., Ye, H., Xue, X.: Modeling spatial-temporal clues in a hybrid deep learning framework for video classification. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 461–470 (2015)
25. Yan, Y., Ni, B., Song, Z., Ma, C., Yan, Y., Yang, X.: Person re-identification via recurrent feature aggregation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 701–716. Springer, Cham (2016). doi:10.1007/978-3-319-46466-4_42
26. Ess, A., Leibe, B., Van Gool, L.: Depth and appearance for mobile scene analysis. In: IEEE 11th International Conference on Computer Vision, pp. 1–8 (2007)
27. Baltieri, D., Vezzani, R., Cucchiara, R.: 3DPeS: 3D people dataset for surveillance and forensics. In: Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, pp. 59–64 (2011)
28. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 91–102. Springer, Heidelberg (2011). doi:10.1007/978-3-642-21227-7_9
29. Wang, T., Gong, S., Zhu, X., Wang, S.: Person re-identification by video ranking. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 688–703. Springer, Cham (2014). doi:10.1007/978-3-319-10593-2_45
30. Zhang, W., Hu, S., Liu, K.: Learning compact appearance representation for video-based person re-identification. arXiv preprint arXiv:1702.06294 (2017)
31. McLaughlin, N., Martinez del Rincon, J., Miller, P.: Recurrent convolutional network for video-based person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1325–1334 (2016)
32. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)
33. Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., Tian, Q.: MARS: a video benchmark for large-scale person re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 868–884. Springer, Cham (2016). doi:10.1007/978-3-319-46466-4_52
34. Roth, P.M., Hirzer, M., Köstinger, M., Beleznai, C., Bischof, H.: Mahalanobis distance learning for person re-identification. In: Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.) Person Re-Identification. ACVPR, pp. 247–267. Springer, London (2014). doi:10.1007/978-1-4471-6296-4_12
35. Zheng, W.S., Gong, S., Xiang, T.: Towards open-world person re-identification by one-shot group-based verification. IEEE Trans. Pattern Anal. Mach. Intell. **38**(3), 591–606 (2016)
36. Liao, S., Mo, Z., Zhu, J., Hu, Y., Li, S.Z.: Open-set person re-identification. arXiv preprint arXiv:1408.0872 (2014)