Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı, S. 291-296, Kasım 2020

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue, pp. 291-296, November 2020

Research Article

Person Re-Identification in Surveillance Videos using Deep Learning based Body Part Partition and Gaussian Filtering

Fatih Aksu[1], Cem Direkoğlu[2*]

[1] Middle East Technical University Northern Cyprus Campus, Centre for Sustainability, Department of Electrical and Electronics Engineering, 99738 Kalkanlı, Güzelyurt, Northern Cyprus, Mersin 10, Turkey (ORCID: 0000-0001-5677-2024)
[2] Middle East Technical University Northern Cyprus Campus, Centre for Sustainability, Department of Electrical and Electronics Engineering, 99738 Kalkanlı, Güzelyurt, Northern Cyprus, Mersin 10, Turkey (ORCID: 0000-0001-7709-4082)

(International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) 2020 – 22-24 October 2020)

(**DOI:** 10.31590/ejosat.823257)

ATIF/REFERENCE: Aksu, F. & Direkoğlu, C. (2020). Person Re-Identification in Surveillance Videos using Deep Learning based Body Part Partition and Gaussian Filtering. *European Journal of Science and Technology*, (Special Issue), 291-296.

Abstract

In this paper, we concentrate on Person Re-Identification (Re-ID) that consists of searching for a person who has been previously observed over a camera network. Person Re-ID is important for searching suspicious or missing persons if we have sample images of the person of interest. Despite the fact that there are many researches on vision-based Person re-identification, it still remains a challenging problem. We propose a person re-identification system using a deep learning based human body part segmentation, and Gaussian filtering based smooth mask generation. A semantic partition technique is used to segment human body parts and generate local binary masks. These masks are deterministic binary images. These binary masks have strict boundaries, and we lose some features with these deterministic masks. Therefore, we apply Gaussian filter for smoothing masks so that features near the boundaries are also taken into account slightly. These smooth masks are applied to the final feature maps generated at the end of network on contrary to other methods which apply mask at the beginning or in the middle of the deep learning network. Therefore, our work is new and different from other works because of using semantic partition and masking at the end of network, as well as our mask are smoothed with Gaussian filter to handle errors during the partitioning stage. We use a well-known pre-trained network, namely ResNet-50, to extract global features, and a method called Cross-Domain Complementary Learning for human body partitioning. Applying Gaussian filtered smooth local masks to the global features, which are extracted at the end of Resnet-50 network, increases the performance of Person Re-Identification system. Evaluation is conducted on a commonly accepted Market-1501 dataset, and results are promising.

Keywords: Person Re-Identification, feature extraction, classification, Deep Learning, Convolutional Neural Networks.

Derin Öğrenme Tabanlı Vücut Bölme ve Gaussian Filtreleme Kullanarak Gözetim Videolarında Kişiyi Yeniden Tanıma

Öz

Bu makalede, daha önce bir kamera ağı üzerinden gözlemlenen bir kişiyi aramak için Kişiyi Yeniden Tanıma (Person Re-Identification) sistemi üzerine yoğunlaşıyoruz. Kişiyi Yeniden Tanıma önemli bir iştir. Örnek olarak kayıp veya şüpheli bir kişinin görüntüleri bulunuyorsa, Kişiyi Yeniden Tanıma sistemi, kişiyi video görüntülerinden bulunmasını sağlayabilir. Bu alanda kişinin yeniden tanımlanmasına ilişkin birçok araştırma olmasına rağmen, bu hala zor bir problem olmaya devam etmektedir ve birçok araştırma bu alanda devam etmektedir. Bu sorunu çözmek amacıyla, makalemizde derin öğrenme tabanlı insan vücudu bölümü bölümleme ve Gaussian filtreleme tabanlı pürüzsüz maske üretimi kullanarak Kişiyi Yeniden Tanıma Sistemi sunuyoruz. İnsan vücudu parçalarını bölümlere ayırmak ve yerel ikili maskeler oluşturmak için anlamsal bir bölme tekniği kullanıyoruz. Bu maskeler deterministik ikili görüntülerdir. Bu ikili maskelerin katı sınırları vardır ve bu deterministik maskelerle bazı özelliklerin kaybedilmesine sebep verip, performansı düşürmektedir. Bu nedenle, maskeleri pürüzsüz hale getirebilmek için Gaussian filtresi

[*] Corresponding Author: Middle East Technical University Northern Cyprus Campus, Centre for Sustainability, Department of Electrical and Electronics Engineering, Kalkanlı, Güzelyurt, Northern Cyprus, Mersin 10, Turkey. ORCID: 0000-0001-7709-4082, cemdir@metu.edu.tr

uyguluyoruz, böylece sınırlara yakın özellikler de performansa biraz katkı sağlıyor. Bizim geliştirdiğimiz metodumuzda, bu pürüzsüz maskeler, derin öğrenme ağının başında veya ortasında maske uygulayan diğer yöntemlerin aksine, ağın sonunda oluşturulan son özellik haritalarına uygulanmaktadır. Bu nedenle, işimiz yeni ve diğer çalışmalardan farklıdır çünkü ağın sonunda anlamsal bölümleme ve maskeleme kullanmanın yanı sıra bölümleme aşamasında hataları gidermek için maskemiz Gaussian filtresi ile pürüzsüz hale getirilmiştir. Global özellikleri çıkarmak için iyi bilinen önceden eğitilmiş bir ağ olan ResNet-50'yi ve insan vücudunun bölümlenmesi için Alanlar Arası Tamamlayıcı Öğrenme adlı bir yöntemi kullanıyoruz. Resnet-50 ağının sonunda çıkarılan global özelliklere Gaussian filtreli pürüzsüz yerel maskelerin uygulanması, kişiyi yeniden tanıma sisteminin performansını artıryor. Değerlendirme, yaygın olarak kabul edilen Market-1501 veri kümesi üzerinde gerçekleştirilmiştir ve sonuçlar umut vericidir.

Anahtar Kelimeler: Kişiyi Yeniden Tanıma, özellik çıkarma, sınıflandırma, Derin Öğrenme, Evrişimli Sinir Ağları.

1. Introduction

Surveillance video analysis is required for safety and security of human beings. There are surveillance control rooms where the videos captured by surveillance cameras are monitored by human operators. However, monitoring many recorded videos continuously is a difficult process, and almost impossible for a person to follow many videos continuously. Computer vision technology is required to automatically detect and recognize objects, and analyze videos for action recognition and behaviour analysis. In our work, we developed a Person Re-Identification (Re-ID) algorithm that is about searching for a person who has been previously captured by a camera network. In particular, Person Re-ID is important for searching suspicious or missing persons if we have sample images of the person of interest. A typical application scenario of person Re-ID can be considered as follows: We have a network of video surveillance cameras monitoring a large public space. A little girl wearing green dress was lost in a shopping mall, or on a street. The girl is seen at first by a camera located at some point. Then the images of the little girl are input to the system for person re-identification in a camera network. A person Re-ID system should associate these images to the same identity and retrieve the same person and the most similar individuals as a result of Re-ID. Despite the fact that there are some researches on vision-based Person re-identification (Leng, et al., 2020; Zheng et al., 2016; Lavi et al., 2018), it still remains a challenging problem and practically not applicable. To achieve Person Re-ID, two important problems should be solved: Person detection and recognition. If an image of a person is given as a query image, the system retrieves the same and similar persons, together with location and time information in the surveillance network. In our research, we concantrate on recognition part that consist of feature extraction and similarity calculation between the queried person and the persons detected in the video frames.

Person Re-ID techniques mainly divided into two approaches: Deep Learning systems (Zheng et al., 2016; Lavi et al., 2018) and Hand-Crafted Feature systems. In Hand-Crafted Feature systems (Hamdoun et al., 2008; Farenzena et al., 2010; Cong et al., 2009; Karaman and Bagdanov, 2012), features and information are extracted from the image itself. For example, some simle Hand-Crafted features which can be extracted from images are edges, corners and color. In (Farenzena et al., 2010), the SURF features have been used as local features to detect and represent points of interest (corners) inside of short video segments. In order to optimize the matching performance and speed up the matching process, these SURF features are indexed in the KD-tree. (Cong et al., 2009) used the manifold geometric structures within video segments to generate more compact spatial descriptors with color-based features. (Karaman and Bagdanov, 2012) designed a conditional random field (CRF) model to incorporate constraints in the spatial and temporal domains for Person Re-ID. The Hand-Crafted Feature systems are considered to be older works, they are less accurate than Deep Learning methods since the features are extracted with human decision. On the other side, Deep Learning based systems automatically learns the features from the example images of objects. As a result, Deep Learning-based systems are more robust than Hand-Crafted systems. Additionally, Hand-Crafted systems may be computationally more complex than Deep Learning methods. Hand-Crafted systems may also be more problematic on large datasets in comparison to Deep Learning systems. Since in Hand-Crafted systems, it is difficult to generalize the features using human decision. Conversely, Deep Learning systems are designed to work well with large datasets since as the number of image samples increases, they can better learn and discriminate features.

Recently, convolutional neural networks (CNNs) (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014) proved to be very succesful in different computer vision problems. CNN is also known as Deep Learning (DL) system in research community. CNNs have been used for Person Re-ID application as well. A number of DL models have been employed to improve the accuracy of Person Re-ID either by changing the existing DL architectures or designing a new DL architecture. Two main types of DL techniques have been used in Person Re-ID (Lavi et al., 2018): (i) a classification technique for Person Re-ID problem, and (ii) a Siamese model based on either pair or triplet comparisons. In a classification model for Person Re-ID, the network takes the image of an individual (i.e. in a bounding box), and calculated class probability of the individual (Wu et al., 2016; Xiao et al., 2016; Su et al., 2017; Li et al., 2017). (Wu et al., 2016) introduce a DL architecture in order to combine CNN features with handcrafted features. (Xiao et al., 2016) presents a work that uses CNN for automatically learning effective neurons in every training dataset by utilizing deep feature representations. Another challenge of Person Re-ID methods is pose variations and misalignments of persons. (Su et al., 2017) aims to tackle this problem of person pose variations and misalignments by proposing a deep convolutional model. Another interesting work that is presented by (Li et al., 2017) which is based on extracting features from the parts of the body and the whole body. In particular, they introduced a context aware deep network consisting of multiple scales to acquire powerful features for Person Re-ID. Conversely, in person re-identification, Siamese network models have been widely used because of small number of training samples. Siamese neural network is a neural network architecture that consist of two or more identical sub-networks. These sub-networks have identical network architecture and identical parameters and weights with each other. A Siamese network can be generally designed as pairwise (Zhang et al., 2014; Wang et al., 2017; Tao et al., 2018; Varior et al., 2016): when there are two sub-networks, or triplet (Ding et al., 2015; Liu et al., 2016; Bai et al., 2017): when there are three sub-networks. A Siamese model basically outputs a similarity score of the network. In triplet Siamese model, an objective function is utilized in order to train the network models. These models generates a margin between distance metric of positive pair and distance metric of negative pair. A softmax layer is used at the end of the network for both distance outputs. The network models are trained with the triplet loss function that makes the similarity distance between the matched pairs less than the mismatched pairs in the learning feature space.

In this paper, we introduce a person re-identification system based on human body part segmentation. This segmentation is achieved with a semantic partitioning method called Cross-Domain Complementary Learning (CDCL) (Lin et al., 2020). After segmentation, we create binary masks for human body part regions. These regions are smoothed with Gaussian fileter to overcome the segmentation error. These smooth masks are appied to the feature maps obtined at the end on feature maps produced at the end of ResNet-50 network (He et al., 2016). The masked feature maps are then processed and classified for Person Re-ID. Results shows that applying Gaussian filtered masks at the end of Resnet-50 network increase the discremination power, and therefore the accuracy of Person Re-ID. Experiments are performed on a publicly available popular Market-1501 dataset (Zheng et al., 2016B). In particular, we compare our method with two baseline methods. The first baseline method is ResNet-50 network. Person Re-ID is achieved directly with ResNet-50 without any body part segmentation and Gaussian filtering. Global features of full body are directly used for recognition in this baseline method. In the second baseline method, horizontal partitioning is used with individual classifier for person re-ID. In this baseline method, the image is divided into equal horizontal stripes which does not contain semantic information about body parts. Then, we compare the results of the baseline methods with the results of; (1) the semantic partition with individual classifier (no Gaussian filtering), (2) the semantic partition with Gaussian filtering and individual classifier, and (3) the semantic partition with Gaussian filtering and shared classifier. In this way, gains of the proposed methods against the baselines can be assessed. Results are very promising.

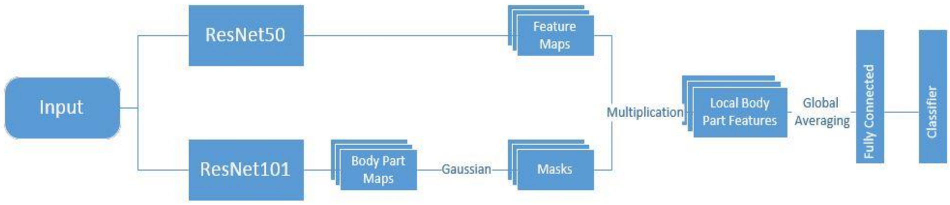


Figure 1. Proposed Framework for Person Re-ID

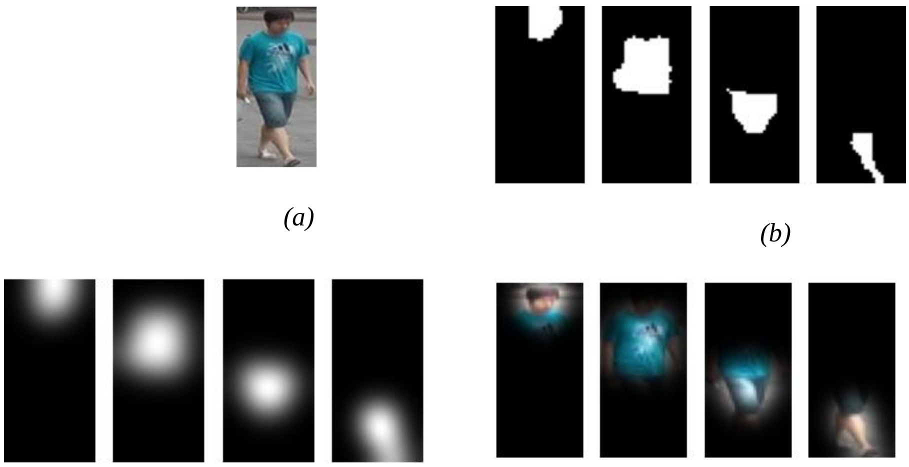
2. Proposed Method

2.1. System Architecture

To develop a person re-identification system, we use CNN based human body part segmentation and feature extraction algorithms. We extract local part features of individuals and then measure distances between features for identification. Our proposed algorithm is illustrated in Figure 1. The upper part of the system shows the main backbone network to generate global feature maps. We use a well-known pre-trained network, ResNet-50 (He et al., 2016) for global feature map generation. ResNet-50 is a commonly used, 50 layer, pretrained convolutional neural network. It is trained on the well-known ImageNet (Deng et al., 2009) dataset. The bottom part shows the human body part segmentation algorithm to partition the human body into parts. We use the state-of-art algorithm for human body partitioning which is called CDCL (Lin et al., 2020). CDCL employes ResNet-101 for body parts segmentation. These part regions are represented with binary images. Then, we apply Gaussian filter to smooth these binary images (masks) features since binary segmentations (part labels) are not 100% accurate. Applying a Gaussian filter optimizes the body part mask distributions, and make them more reliable. After extracting global feature maps as shown at the upper part of Figure 1, and creating smooth local masks at the bottom part, we mask these feature maps to separate local features. We have 7 different feature maps for masking. Finally, global average pooling is utilized to take the average of all parts separately and concatenate the final values into a single fully connected layer. The right part of the Figure 1 shows the averaging and classification blocks.

2.2. Local Body Part Feature Extraction and Our Contributions

As described above, we use ResNet-50 (He et al., 2016) to extract global features of the images. ResNet-50 is a commonly used, 50 layer, pretrained convolutional neural network. It is trained on the well-known ImageNet (Deng et al., 2009) dataset. We first use ResNet-50 with raw images to extract global features of the person. A sample raw image is shown in Figure 2(a). In literature, there are two common approaches for feature generation. Some works (Bai et al., 2017B; Wang et al., 2018) use global features to calculate the similarity and others (Cheng et al., 2016; Sun et al., 2018) use local features. We choose to use local features because global features can be affected more by the distortions, occlusions, background etc. We take into account every body part individually. Since the color, texture and other features of the parts are similar, the features that are extracted are more consistent than the global features. These part-based methods are also divided into a few categories. Some works (Cheng et al., 2016; Sun et al., 2018) partition the image into equal horizontal stripes. However, this kind of partition causes some problems because of the pose variations and improper bounding boxes. For instance, the first part of one image can contain a head but the other can contain only background or middle parts can be both contain body and legs. Therefore, we use a semantic partition technique. Our work partitions the final feature map according to body parts on the contrary to horizontal partition as in other works. In literature, partition at the end of the network is done as horizontally in general. The semantic partition takes places mostly at the beginning or in the middle of the network. Therefore, we use a new approach which is using semantic partition at the end. Moreover, we use a different algorithm to extract body part maps and use them as masks in our partition method. To do this, first, we use a state-of-art algorithm (Lin et al., 2020) to label the body parts accordingly like head, body, arms, legs etc. After that, we use these labels to mask the global features that are the outputs of the ResNet-50 in order to separate the features according to each body part. We convert all the labels to binary map. A sample binary mask is shown in Figure 2 (b). However, these binary maps have strict boundaries. Since the body parts are not labeled perfectly, we lose some features with these deterministic masks. Therefore, we apply Gaussian filter to smooth these masks to take the features near the labels into account. Figure 2 (c) shows the resulting masks after Gaussian filtering. This is a new approach as other works generally does not use part labelling or when it is used, they use different attention-based methods. After this process, we multiply these masks with the global image feature map one-by-one to get the features of each part separately. Figure 2 (d) shows the masked images if they were applied to input image directly. However, it is important to mention that those smooth masks are applied to feature maps at the output of Resnet-50 with multiplication as illustrated in Figure 2. As a last step, we perform adaptive average pooling to take the means of each part in order to reduce dimensions. Then these mean values, final features, are concatenated and fed into the classifier. Some of the researches (Li et al., 2017B; Gray and Brennan, 2007) uses multi-loss method which means that every part is fed into a separate classifier and has separate loss. However, we observed that combining feature and using one classifier increases accuracy considerably. Results are presented in experiments section.



(a) *(b)*

(c) *(d)*

Figure 2. (a) Raw image (b) binary body part masks (c) masks after applying Gaussian filter (d) masking operation is illustrated on a raw image.

3. Evaluation

We conduct experiments on a commonly used dataset, namely Market-1501 dataset (Zheng et al., 2016B). Market-1501 dataset has 32,668 images of 1,501 persons. Each individual is captured by at least two different cameras from the front of a supermarket, and there are 6 different camera views in total. Some of the images can be seen in Figure 3. We implement the algorithm that is explained in Section 2 with a well-known deep learning library, namely Pytorch. We choose batch size 32, epoch number 30 and learning rate 0.1 with a decay by a factor of 0.1 in every 10 epochs. The parameters of the Gaussian filter is as follows; kernel size is 41x41 and standard deviation is 5. Filter is applied on spatial domain. We did experiments with horizontal partitioning, body part partitioning, and different classifiers. Most of the benchmarks uses CMC evaluation which stands for cumulative matching characteristics. CMC represents the probability of a person appears in different-sized candidate lists. Generally, rank-1, rank-5 and rank-10 lists are used which represent first, first five, and first ten guesses. Particularly, achieving good results in rank-1 is important for retrieving tasks. Market-1501 also uses another evaluation metric that is called mean average precision (mAP). To calculate mAP, first, it needed to be calculated the area under the Precision-Recall curve for each query. This area is the average precision (AP). Then, the mean of average precision of all queries is calculated.

Results are given in Table 1. First, we present the results of two baseline methods. The first baseline method is defined to be ResNet-50 network. Person Re-ID is achieved directly with ResNet-50 without any body part segmentation and Gaussian filtering. Global features of full body are directly used for recognition in this baseline method. In the second baseline method, horizontal partitioning is used with individual classifier for person re-ID. In this baseline method, the image is divided into equal horizontal stripes which does not contain semantic information about body parts. Then, we compare the results of the baseline methods with the results of; (1) the semantic partition with individual classifier (no Gaussian filtering), (2) the semantic partition with Gaussian filtering and individual classifier, and (3) the semantic partition with Gaussian filtering and shared classifier. In this way, gains of the proposed methods against the baselines can be assessed. Results show that the baseline method #1 (ResNet-50) achieves 0.858, 0.949 and 0.969 rank-1, rank-5 and rank-10 respectively. The baseline #1 also achieves a mAP of 0.67. The baseline method #2 (horizontal partition) achieves worse than the baseline method #1. In particular, baseline #2 achieves 0.796, 0.912, 0.941 and 0.59 for rank-1, rank-5, rank-

10 and mAP respectively. The semantic partition with individual classifier (without the Gaussian filtering) performs the worse among the all compared methods (except rank-1) with 0.817, 0.912, 0.94 and 0.468 for rank-1, rank-5, rank-10 and mAP respectively. The semantic partition with Gaussian filtering and individual classifier performs better than the semantic partition without the Gaussing filtering but worse than the baseline method #1 with 0.845, 0.94, 0.961 and 0.634 for rank-1, rank-5, rank-10 and mAP respectively. Finally, the semantic partition with Gaussian filtering and shared classifier improves the performance at rank-1, rank-5 and rank-10 compared to the baseline method #1 with 0.87, 0.952, 0.971 for rank-1, rank-5, rank-10. In particular, at rank-1, it improves the performance comparing to the baseline methods significantly. In addition, the semantic partition with Gaussian filtering and shared classifier achives similar results as the baseline #1 with a mAP of 0.688. These results show that integration of semantic partition with Gaussian filtering and shared classifier with an existing pre-trained network can improve person re-identification performance. In future, we plan to integrate our semantic partitioning based body parts segmentation with Gaussing filtering approach to other networks that is particularly designed for person re-identification tasks rather than a general network used for image classification (Resnet-50). In particular, Harmonious attention network (Li et al., 2018), Auto-ReID (Quan et al., 2019) and PLR-OSNet (Xie et al., 2020) can be tested. In addition, semantic approaches (Sah and Direkoglu, 2017) can be integrated for person re-ID that allows intelligent search interfaces (Hatirnaz et al., 2020).

Figure 3. Sample images from Market-1501 dataset [23].

Table 1. Experimental results

Model	Rank-1	Rank-5	Rank-10	mAP
Baseline Method #1 (ResNet-50) [22]	0.858	0.949	0.969	**0.670**
Baseline Method #2 (Horizontal partition with individual classifiers) [28]	0.796	0.912	0.941	0.590
Semantic partition with individual classifiers (no Gaussian filtering) (Proposed Method)	0.817	0.912	0.940	0.468
Semantic partition with Gaussian filtering and individual classifiers (Proposed Method)	0.845	0.940	0.961	0.634
Semantic partition with Gaussian filtering and shared classifier (Proposed Method)	**0.870**	**0.952**	**0.971**	0.668

4. Conclusions and Future Work

Person Re-Identification (Person Re-ID) is an important task for searching suspicious or missing persons in surveillance domain. To address this problem, in this paper, we developed a person re-identification algorithm. Our algorithm involves body part partitioning and Gaussian filtering based mask generation for Person Re-ID. First, a pre-trained network (Resnet-50) is employed to extract global features of the person. A semantic partition technique is also used to segment human body parts, and create local binary masks where the masks are deterministic binary images. Mask smoothing is performed with a Gaussian filter, and then we perform masking operation to the global feature maps produced at the end of Resnet-50. Our masks are smoothed binary images. Our approach is experimented on Market-1501 dataset that is a publicly available popular dataset. Results indicates that the our method is very promising. Using Gaussian filter with semantic partioning in the algorithm increases the accuracy of baseline algorithms. In our future work, we can integrate our semantic partitioning based body parts segmentation with Gaussing filtering approach to other networks that is particularly designed for person re-identification tasks. We will also experiment our method on different datasets and conduct time evaluations.

5. Acknowledge

We would like to thank Middle East Technical University – Northern Cyprus Campus for supporting us with Scientific Research Project (SRP) Fund (Grant no: FEN-19-D-3).

References

Bai, X., & Yang, M., & Huang, T., & Dou, Z., & Yu, R., & Xu, Y. (2017). Deep-Person: Learning Discriminative Deep Features for Person Re-Identification. arXiv preprint, arXiv:1711.10658.

Bai, X. et al. (2017B). Deep-Person: Learning Discriminative Deep Features for Person Re-Identification, arXiv:1711.10658.

Cheng, D., & Gong, Y., & Zhou, S., & Wang, J., & Zheng, N. (2016). Person Re-identification by Multi-Channel Parts-Based CNN with Improved Triplet Loss Function, IEEE Conference on Computer Vision and Pattern Recognition, 1335-1344.

Cong, D. N. T., & Achard, C., & Khoudour, L. & Douadi, L. (2009). Video Sequences Association for People Re-Identification Across Multiple Non-Overlapping Cameras. International Conference on Image Analysis and Processing, 179–189.

Deng, J., & Dong, W., & Socher, R., & Li, L., & Kai, L., Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database, IEEE Conference on Computer Vision and Pattern Recognition, 248-255.

Ding, S., & Lin, L., & Wang, G., & Chao, H. (2015). Deep Feature Learning with Relative Distance Comparison for Person Re-Identification. Pattern Recognition, 48(10), 2993–3003.

Farenzena, M. & Bazzani, L., & Perina, A., & Murino, V., & Cristani, M. (2010). Person Re-Identification by Symmetry-Driven Accumulation of Local Features. IEEE Computer Vision and Pattern Recognition (CVPR), 2360–2367.

Gray, D., & Brennan, S., & Tao, H. (2007). Evaluating appearance models for recognition, reacquisition, and tracking, IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), 1–7.

Hamdoun, O., & Moutarde, F., & Stanciulescu, B., & Steux, B. (2008). Person Re-Identification in Multi-Camera System by Signature based on Interest Point Descriptors Collected on Short Video Sequences. ACM/IEEE International Conference on Distributed Smart Cameras, 1–6.

Hatirnaz, E., & Sah, M., & Direkoglu, C. (2020). A novel framework and concept-based semantic search Interface for abnormal crowd behaviour analysis in surveillance videos. Multimedia Tools and Applications, 79, 17579–17617.

He, K., & Zhang, X., & Ren, S., & Sun, J. (2016). Deep residual learning for image recognition, IEEE Conference on Computer Vision and Pattern Recognition, 770-778.

Karaman, S. & Bagdanov, A. D. (2012). Identity Inference: Generalizing Person Re-Identification Scenarios. European Conference on Computer Vision, 443–452.

Krizhevsky, A. & Sutskever, I., & Hinton, G. E. (2012). Imagenet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 1097–1105.

Lavi, B., & Serj, M. F., & Ullah, I. (2018). Survey on Deep Learning Techniques for Person Re-Identification Task. arXiv preprint arXiv:1807.05284.

Leng, Q., & Ye, M., & Tian, Q. (2020). A Survey of Open-World Person Re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 30(4), 1092-1108.

Li, D. & Chen, X., & Zhang, Z., & Huang, K. (2017). Learning Deep Context-Aware Features over Body and Latent Parts for Person Re-Identification. IEEE Conference on Computer Vision and Pattern Recognition, 384–393.

Li, W. et al. (2017B). Person Re-Identification by Deep Joint Learning of Multi-Loss Classification, International Joint Conference on Artificial Intelligence, 2194–2200.

Li, W., & Zhu, X., & Gong, S. (2018). Harmonious attention network for person re-identification, arXiv preprint, arXiv:1802.08122.

Lin, K., & Wang, L., & Luo, K., & Chen, Y., & Liu, Z., & Sun, M. (2020). Cross-Domain Complementary Learning Using Pose for Multi-Person Part Segmentation, IEEE Transactions on Circuits and Systems for Video Technology.

Liu, J., & Zha, Z.-J., & Tian, Q., & Liu, D., & Yao, T., & Ling, Q., & Mei, T. (2016). Multi-Scale Triplet CNN for Person Reidentification. ACM Conference on Multimedia, 192–196.

Quan, R., & Dong, X., & Wu, Y., & Zhu, L., & Yang, Y. (2019). Auto-reid: Searching for a part-aware convnet for person re-identification, in Proc. ICCV.

Sah, M., & Direkoglu, C. (2017). Semantic annotation of surveillance videos for abnormal crowd behaviour search and analysis, IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-6.

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-scale Image Recognition. arXiv preprint arXiv:1409.1556.

Su, C., & Li, J., & Zhang, S., & Xing, J., & Gao, W., & Tian, Q. (2017). Pose-driven Deep Convolutional Model for Person Re-Identification. IEEE International Conference on Computer Vision (ICCV), 3980–3989.

Sun, Y., & Zheng, L., & Yang, Y., & Tian, Q., & Wang, S. (2018). Beyond Part Models: Person Retrieval with Refined Part Pooling (And A Strong Convolutional Baseline), European Conference on Computer Vision, 480-496.

Tao, D., & Guo, Y., & Yu, B., & Pang, J. & Yu, Z. (2018). Deep Multi-View Feature Learning for Person Re-Identification, IEEE Transactions on Circuits and Systems for Video Technology, 28(10), 2657 – 2666.

Varior, R. R., & Shuai, B., & Lu, J., & Xu, D., & Wang, G. (2016). A Siamese Long Short-Term Memory Architecture for Human Re-Identification. European Conference on Computer Vision, 135–153.

Wang, J., & Wang, Z., & Gao, C., & Sang, N., & Huang, R. (2017). Deeplist: Learning Deep Features with Adaptive Listwise Constraint for Person Re-Identification. IEEE Transactions on Circuits and Systems for Video Technology. 27(3), 513 – 524.

Wang, G. et al. (2018). Learning Discriminative Features with Multiple Granularities for Person Re-Identification. ACM international conference on Multimedia, 274–282.

Wu, S., & Chen, Y-C, & Li, X., & Wu, A. C., & You, J. J., & Zheng, W. S. (2016). An Enhanced Deep Feature Representation for Person Re-Identification. IEEE Winter Conference Applications of Computer Vision (WACV), 1–8.

Xiao, T., & Li, H., & Ouyang, W., & Wang, X. (2016). Learning Deep Feature Representations with Domain Guided Dropout for Person Re-Identification. IEEE Conference on Computer Vision and Pattern Recognition, 1249–1258.

Xie, B., & Wu, X., & Zhang, S., & Zhao, S., & Li, M. (2020). Learning Diverse Features with Part-Level Resolution for Person ReIdentification, arXiv preprint, arXiv:2001.07442.

Zhang, G., & Kato, J., & Wang, Y., & Mase, K. (2014). People Re-Identification using Deep Convolutional Neural Network. IEEE International Conference on Computer Vision Theory and Applications (VISAPP), 216–223.

Zheng, L., & Yang, Y., & Hauptmann, A. G. (2016). Person Re-identification: Past, Present and Future. arXiv preprint arXiv:1610.02984.

Zheng, L., & Shen, L., & Tian, L., & Wang, S., & Wang, J., & Tian, Q. (2016B). Scalable person re-identification: a benchmark, IEEE International Conference on Computer Vision, 1116–1124.