16 papers and related papers

**left column gives a brief overview of the algorithm used of the cited papers if it used NN

1. **Person Transfer GAN to Bridge Domain Gap for Person Re-Identification**

“Thanks to the development of deep learning and the availability of many datasets, person ReID performance has been significantly boosted. For example, the Rank-1 accuracy of single query on *Market1501* [38] has been im- proved from 43.8% [21] to 89.9% [30].“ [30] L. Wei, S. Zhang, H. Yao, W. Gao, and Q. Tian. Glad: Global-local-alignment descriptor for pedestrian retrieval. In *ACM MM*, 2017. 1, 2, 6	[30] Global-local alignment descriptor first detects several body parts, then learn descriptors from both global and local regions. After detecting more subtle body parts, GLAD aims to be robust to misalignment and gain more discriminative power by explicitly embedding global and local cues.
“Deep learning based descriptors have shown substan- tial advantages over hand-crafted features on most of per- son ReID datasets. Some works [32, 40] learn deep de- scriptors from the whole images with classification models, where each person ID is treated as a category. “ [32] T.Xiao,H.Li,W.Ouyang,andX.Wang.Learningdeepfea- ture representations with domain guided dropout for person re-identification. In *CVPR*, 2016. 2, 5 “Some other works [39, 6] combine verification models with classifica- tion models to learn descriptors. “ [39] Z. Zheng, L. Zheng, and Y. Yang. A discriminatively learned cnn embedding for person re-identification. *arXiv preprint arXiv:1611.05666*, 2016. 2 Hermans *et al.* [9] show that triplet loss effectively improves the performance of per- son ReID. [9] A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017. 2 Similarly, Chen *et al.* [1] propose the quadruplet network to learn representations. [1] W. Chen, X. Chen, J. Zhang, and K. Huang.	[32] Learns deep feature representations. First, mix data and labels from all domains together and trains CNN on mix with single softmax loss. Next, for each domain, perform forward pass on all samples and compute for each neuron average impact on the objective function. Next, replace standard dropout layer, and continue to train CNN model for more epochs.

Beyond triplet loss: a deep quadruplet network for person re-identification. In *CVPR*, 2017. 2 Dif- ferently, extra constraints on person identity are applied to ensure the transferred images can be used for model train- ing. Zheng *et al.* [40] adopt GAN to generate new samples for data augmentation in person ReID.	

2. **Diversity Regularized Spatiotemporal Attention for Video-Based Person Re-Identification**

Existing video-based person re-identification methods represent each frame as a feature vector and then com- pute an aggregate representation across time using aver- age or maximum pooling [52, 28, 46]. [52] Z. Zhou, Y. Huang, W. Wang, L. Wang, and T. Tan. See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification. In *Computer Vision and Pattern Recognition*, 2017. 2, 3, 8 [28] H. Liu, Z. Jie, K. Jayashree, M. Qi, J. Jiang, S. Yan, and J. Feng. Video-based person re-identification with accumula- tive motion context. *arXiv preprint arXiv:1701.00193*, 2017. 2, 8 [46] J. You, A. Wu, X. Li, and W.-S. Zheng. Top-push video- based person re-identification. In *Computer Vision and Pat- tern Recognition*, pages 1345–1353, 2016. 2, 8	
Image-based person re-identification mainly focuses on two categories: extracting discriminative features [13, 9, 33, 19, 43] and learning robust metrics [37, 50, 18, 36, 2]. [13] D. Gray and H. Tao. Viewpoint invariant pedestrian recog- nition with an ensemble of localized features. In *European Conference on Computer Vision*, pages 262–275. Springer, 2008. 2 [37] B. J. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary. Person re-identification by support vector ranking. In *British Machine Vision Conference*, 2010. 2	[13] This paper uses Adaboost to learn an object class specific representation and a discriminative recognition model. [37] This paper reformulates the person re-id problem as a ranking problem. Uses Ensemble RankSVM instead of SVM-based ranking methods. [50]This paper reformulates the person re-id problem as a distance learning problem. It uses Probabilistic Relative Distance Comparison model (PRDC). It is different from existing distance learning methods in that, rather than minimising intra-class variation whilst maximising intra-class variation, it aims to maximise the probability of a

[50] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition*, pages 649–656. IEEE, 2011.	pair of true match having a smaller distance than that of a wrong pair.
In [46], You *et al*. present a top-push distance learning model accompanied by the minimization of intra-class variations to optimize the matching accuracy at the top rank for person re-identification. McLaughlin *et al*. [35] introduce an RNN model to encode temporal information. They utilize tem- poral pooling to select the maximum activation over each feature dimension and compute the feature similarity of two videos. Wang *et al*. [41] select reliable space-time features from noisy/incomplete image sequences while simultane- ously learning a video ranking function. Ma *et al*. [34] encode multiple granularities of spatiotemporal dynamics to generate latent representations for each person. A Time Shift Dynamic Time Warping model is derived to select and match data between inaccurate and incomplete sequences.	
Atten- tion models [45, 22, 21] have grown in popularity since [45]. [45] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudi- nov, R. Zemel, and Y. Bengio. Show, attend and tell: Neural image caption generation with visual attention. In *Interna- tional Conference on Machine Learning*, pages 2048–2057, 2015. 2 Zhou *et al*. [52] combine spatial and temporal infor- mation by building an end-to-end deep neural network. [52] Z. Zhou, Y. Huang, W. Wang, L. Wang, and T. Tan. See the forest for the trees: Joint spatial and temporal recurrent neural networks for video-based person re-identification. In *Computer Vision and Pattern Recognition*, 2017. 2, 3, 8 Liu *et al*. [30] proposed a multi-directional	**This portion of the paper talks about re-id using attention models [45] Uses attention based model that automatically describes the content of images (must be salient objects to be recognisable) [52]Focuses on video-based person re-id and uses end-to-end deep neural network architecture to jointly learn features and metrics. Claims to be able to pick out most discriminative frames in a given video by temporal attention model. [30] Uses attention-based deep nn called HydraPlus-Net and is capable of capturing multiple attentions from low-level to semantic-level.

attention module to exploit the global and local contents for image-based person re-identification. [30] X. Liu, H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, and X. Wang. Hydraplus-net: Attentive deep features for pedestrian analysis. *arXiv preprint arXiv:1709.09930*, 2017.	

3. **A Pose-Sensitive Embedding for Person Re-Identification With Expanded Cross Neighborhood Re-Ranking**

This correspondence can be established by explicitly using full body pose infor- mation for alignment [32] or locally through matching corresponding detected body parts [41, 42]. [32] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Pose- driven deep convolutional model for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision ICCV*, pages 3960–3969, 2017. [41] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang. Spindle net: Person re-identification with hu- man body region guided feature decomposition and fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1077–1085, 2017.	[32] Uses Pose-driven Deep Convolutional (PDC) model to learn improved feature extraction and matching models from end-to-end. [41] Uses Spindle Net, based on human body region guided multi-stage feature decomposition and tree-structured competitive feature fusion. First time human body structure information is considered in a CNN framework to facilitate feature learning.
A person's body pose is an important cue for successful re-identification. The popular SDALF fea- ture by Farenza *et al*. [8] uses two axes dependent on the body's pose to derive a feature description with pose invari- ance. Cho *et al*. [6] define four view angles (front, left, right, back) and learn corresponding matching weights to emphasize matching of same-view person images. A more fine-grained pose representation based on Pictorial Struc- tures was first used in [5] to focus on matching between individual body parts. More recently, the success of deep learning architectures in the context of re-id has lead to sev- eral works that include pose information into a CNN-based matching. In [43] Zheng *et al*. propose to use a CNN-based external pose estimator to normalize person images based on their pose. The original and normalized images are then used to train a single deep re-id embedding. A similar ap- proach is described by Su *et al*. in [32]. Here, a sub-network first estimates a pose map which is then used to crop the lo- calized body parts. A local and a global	[43] Introduces Pose Invariant Embedding as a descriptor to address alignment problem by using Posebox. [32] Uses Pose-driven Deep Convolutional (PDC) model to learn improved feature extraction and matching models from end-to-end.

person representa- tion are then learned and fused. Pose variation has also been addressed by explicitly detecting body parts through detec- tion frameworks [41], by relying on visual attention maps [25], or body part specific attention modeling [42]. [43] L. Zheng, Y. Huang, H. Lu, and Y. Yang. Pose invariant embedding for deep person re-identification. *arXiv preprint arXiv:1701.07732*, 2017. [32] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Pose- driven deep convolutional model for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision ICCV*, pages 3960–3969, 2017.	

4. **Image-Image Domain Adaptation With Preserved Self-Similarity and Domain-Dissimilarity for Person Re-Identification**

2) when models trained on one dataset are di- rectly used on another, the re-ID accuracy drops dramati-cally [6] due to *dataset bias* [41] [6] H. Fan, L. Zheng, and Y. Yang. Unsupervised person re- identification: Clustering and fine-tuning. *arXiv preprint arXiv:1705.10444*, 2017. 1, 3, 6, 8	[6] Deals with no or few labels. Uses progressive pretrained model to transfer deep representations to unseen domains by 1) pedestrain clustering and 2) fine tuning of CNN.
Some methods are based on saliency statistics [50, 44]. In [49], K-means clus- tering is used for learning an unsupervised asymmetric met- ric. Peng *et al*. [35] propose an asymmetric multi-task dic- tionary learning for cross-data transfer. [50] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013. 3 [49] H.Yu,A.Wu,andW.Zheng.Cross-viewasymmetricmetric learning for unsupervised person re-identification. In *ICCV*, 2017. 3, 8	

5. **Human Semantic Parsing for Person Re-Identification**

Also, there has been some attempts [33, 35] to im- prove person re-identification performance using person at- tributes. These attributes usually contain high-level seman- tic information that are supposedly invariant to pose, illu- mination and	[33] “We train an attribute classifier on separate data and include its responses into the training process of our person re-id model which is based on convolutional neural networks (CNNs). This allows us to learn a person representation which

camera point of view. [33] A. Schumann and R. Stiefelhagen. Person re-identification by deep learning attribute-complementary information. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pages 1435–1443. IEEE, 2017. 3, 8	contains information complementary to that contained within the attributes. Our approach is able to identify attributes which perform most reliably for re-id and focus on them accordingly. We demonstrate the performance improvement gained through use of the attribute information on multiple large-scale datasets and report insights into which attributes are most relevant for person re-id.”
The person re-identification back- bone in SPReID is exactly Inception-V3 [37] with a mi- nor modification of removing global average pooling layer. [37] C.Szegedy,V.Vanhoucke,S.Ioffe,J.Shlens,andZ.Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016. 2, 3, 4, 6	

6. **Video Person Re-Identification With Competitive Snippet-Similarity Aggregation and Co-Attentive Snippet Embedding**

Early methods on video-based person Re-ID focus on handcrafting video representations. Wang *et al.* [27] com- bined HOG3D features and optical flow energy profile to… [27] T. Wang, S. Gong, X. Zhu, and S. Wang. Person re- identification by video ranking. In *ECCV*, 2014. 1, 2, 6, 8 Liu *et al.* [17] further incorporated spatial pictorial structures for spa- tially aligning person videos with different poses and view- points. [17] K. Liu, B. Ma, W. Zhang, and R. Huang. A spatio-temporal appearance representation for video-based pedes- trian re-identification. In *ICCV*, 2015. 2, 3	

Metric learning algorithms were also developed for video-based Re-ID. Zhu *et al.* [40] and You *et al.* [31] imposed set-based constraints to better distinguish intra-person variations from the inter-person ones. [40] X. Zhu, X.-Y. Jing, F. Wu, and H. Feng. Video-based person re-identification by simultaneously learning intra-video and inter-video distance metrics. In *IJCAI*, 2016. 2, 6 [31] J. You, A. Wu, X. Li, and W.-S. Zheng. Top-push video- based person re-identification. In *CVPR*, 2016. 2, 6, 8	

7. **Mask-Guided Contrastive Attention Model for Person Re-Identification**

Re- cently, several deep learning based methods are proposed to learn identity features from the body parts which are generated by either part region detection [24], or pose and key- points estimation [32, 23, 49, 44]. [24] D. Li, X. Chen, Z. Zhang, and K. Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *CVPR*, 2017. 2, 3, 4, 6, 7, 8 [32] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Pose- driven deep convolutional model for person re-identification. In *ICCV*, 2017. 2, 3 [49] L. Zheng, Y. Huang, H. Lu, and Y. Yang. Pose invariant embedding for deep person re-identification. *arXiv preparXiv:1701.07732*, 2017. 2, 3 [44] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang. Spindle net: Person re-identification with hu- man body region guided feature decomposition and fusion. In *CVPR*, 2017. 2, 3, 7, 8	[24] Uses Multi-Scal Context Aware Network (MSCAN) to learn powerful features over full body and body parts. Uses Spatial Tranformer Networks (STN) with novel spatial constraints to learn body parts instead of predefining them. [32] Uses Pose-driven Deep Convolutional (PDC) model to learn improved feature extraction and matching models from end to end. [49] Fix pedestrian misalignment using pose invariant embedding (PIE)
Secondly, the mask contains body shape information which can be regarded as the important gait features. It has been proved that the body mask is robust to illumination, cloth colors, and thus is useful for identifying a person [35]. [35] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis- based gait recognition for	

human id	
These methods usually learn the ID-discriminative Embedding (IDE) fea- ture [48] via training a deep classification network. In addi- tion, some works try to introduce the pair-wise contrastive loss [14], triplet ranking loss [54] and quadruplet loss [8] to further enhance the IDE feature. To combine the clas- sification and pair-wise loss, Chen et al. attempt to ap- ply a multi-task model to simultaneously learn classification and ranking tasks [9]. There are also some works trying to implement the multi-scale context [24] or multi-resolution method [29] in person ReID. [48] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian. Mars: A video benchmark for large-scale person re-identification. In *ECCV*, 2016. 2, 5, 7, 8 **paper not found online	

8. **Person Re-Identification With Cascaded Pairwise Convolutions**

Like the Hard Negative Mining strategy [1, 35] which selects the negative samples with large loss on current model, the SRL strategy tries to increase the expected loss in each batch by adap- tively adjusting the ratio of positive samples and negative samples. [1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *CVPR*, [35] Y. Zhang, X. Li, L. Zhao, and Z. Zhang. Semantics-aware deep correspondence structure learning for robust person re-identification. In *IJCAI*, pages 3545–3551, 2016.	[1] DNN formulates the problem of person re-id as binary classification. Network archi has two layers of tied convolution with max pooling, cross-input neighborhood differences, patch summary features, across-patch features, higher-order relationships, softmax function to yield final estimate of whether the input images are of the same person or not.

9. **Exploring Disentangled Feature Representation Beyond Face Identification**

Facial Recognition Paper -> Irrelevant

10. **Multi-Level Factorisation Net for Person Re-Identification**

Most recent person Re-ID approaches [42, 38, 24, 32, 27, 25] employ deep neural networks (DNNs) to learn view- invariant discriminative features. For matching, the fea- tures are typically extracted	[42] Pipeline for learning deep feature representations from multiple domains with CNN. In addition, proposed domain guided dropout algorithm to improve learning procedure as it

from the very top feature layer of a trained model [42] T.Xiao,H.Li,W.Ouyang,andX.Wang.Learningdeepfea- ture representations with domain guided dropout for person re-identification. In *CVPR*, 2016. 1, 2, 5, 6 [38] Y. Sun, L. Zheng, W. Deng, and S. Wang. Svdnet for pedes- trian retrieval. *ICCV*, 2017. 1, 2, 5, 6	discards useless neurons for each domain.
A number of recent deep Re-ID models started to model discriminative factors of multiple levels. Some focused on learning *semantic visual features* with additional super- vision in the form of attributes [32, 35, 26, 15, 31, 41]. [32] N. McLaughlin, J. M. del Rincon, and P. C. Miller. Person reidentification using deep convnets with multitask learning. *IEEE Trans. CSVT*, 27(3):525–539, 2017. 1, 2, 3 [35] A. Schumann and R. Stiefelhagen. Person re-identification by deep learning attribute-complementary information. In *CVPR Workshops*, 2017. 1, 3, 6 [26] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, and Y. Yang. Improv- ing person re-identification by attribute and identity learning. *arXiv:1703.07220*, 2017. 1, 3, 6, 7	[32] “We propose to tackle this problem by training a deep convolutional network to repre- sent a person’s appearance as a low-dimensional feature vector that is invariant to common appearance variations encountered in the re-identification problem. Specifically, a Siamese-network architecture is used to train a feature extraction network using pairs of similar and dissimilar images. “ [26] “*this paper proposes a very sim- ple convolutional neural network (CNN) that learns a re-ID embedding and predicts the pedestrian attributes simulta- neously. This multi-task method integrates an ID classifica- tion loss and a number of attribute classification losses, and back-propagates the weighted sum of the individual losses. “*
The training procedure of MLFN for Person Re-ID follows the standard identity clas- sification paradigm [42, 53, 38] where each person’s iden- tity is treated as a distinct class for recognition. [42] T.Xiao,H.Li,W.Ouyang,andX.Wang.Learningdeepfea- ture representations with domain guided dropout for person re-identification. In *CVPR*, 2016. 1, 2, 5, 6 [53] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang. Spindle net: Person re-identification with hu- man body region guided feature decomposition and fusion. In *CVPR*, 2017. 2, 3, 5, 6	[53] “*In this study, we propose a novel Convolutional Neural Network (CNN), called Spindle Net, based on human body region guided multi-stage feature de- composition and tree-structured competitive feature fusion… The proposed Spindle Net brings unique advantages: 1) it separately captures semantic features from different body regions thus the macro- and micro-body features can be well aligned across images, 2) the learned region features from different semantic regions are merged with a competitive scheme and discriminative features can be well preserved. State of the art performance can be achieved on multiple datasets by large margins. “*

11. Attention-Aware Compositional Network for Person Re-Identification

There are two categories of methods addressing the problem of person re-identification, namely feature rep- resentation and distance metric learning. The first cat- egory mainly includes the traditional feature descriptors [53, 52, 51, 27, 7, 31, 35] and deep learning features [43, 41, 42, 12, 25, 40]. These approaches dedicate to de- sign view-invariant representations for person images. The second category [28, 12, 27, 45, 19, 9, 23, 14, 33] mainly targets on learning a robust distance metric to measure the similarity between images. [41] 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. [42] L. Wu, C. Shen, and A. v. d. Hengel. Personnet: Person re-identification with deep convolutional neural networks. *arXiv:1601.07255*, 2016.	[42] "The network takes a pair of raw RGB images as input, and outputs a similarity value indicating whether the two input images depict the same person. A layer of computing neighborhood range differences across two input images is employed to capture local relationship between patches [1]. This operation is to seek a robust feature from input images. By increasing the depth to 10 weight layers and using very small (3×3) convolution filters, our architecture achieves a remarkable improvement on the prior-art configurations. Meanwhile, an adaptive Root- Mean-Square (RMSProp) gradient decent algorithm is integrated into our architecture, which is beneficial to deep nets."
For example, Zhao *et al.* [49] proposed Spindle Net, that extracted and fused three level part features. Parts were extracted by PRN. Su *et al.* [37] proposed a Pose-driven Deep Convolutional model (PDC) that utilized Spatial Transformer Network (STN) to localize and crop body regions based on pre-defined centers. Zheng *et al.* [54] introduced to extract Pose Invariant Embedding (PIE) through aligning pedestrians to standard pose. [37] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Pose- driven deep convolutional model for person re-identification. In *ICCV*, 2017.	[37] Uses Pose-driven Deep Convolutional (PDC) model to learn improved feature extraction and matching models from end-to-end.
Inspired by the multi-stage CNN [6] for human pose es- timation, we utilize a two-stage network to learn part atten- tions. [6] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi- person 2d pose estimation using part	

affinity fields. In *CVPR*, 2016.	

12. -

Focus on person recognition

13. **Unifying Identification and Context Learning for Person Recognition**

With the emergence of deep learning techniques, state-of-the-art person re-identification methods adopted Convolutional Neural Networks (CNN) for learning person features. Li *et al*. [17] designed a filter pairing neural net- work to jointly handle misalignment and geometric trans- formations. Ahmed *et al*. [17] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *Pro- ceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 152–159, 2014. 2, 6, 7, 8 [1] proposed a Cross-Input Dif- ference CNN to capture local relationships between the two input images based on mid-level features from each input image. [1] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *Proceed- ings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3908–3916, 2015. 2 Ding *et al*. [8] exploited triplet samples for training CNNs to minimize the feature distance between positive samples and maximize the distance between negative sam- ples. [8] S. Ding, L. Lin, G. Wang, and H. Chao. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognition*, 48(10):2993–3003, 2015. 2 Xiao *et al*. [33] proposed a Domain Guided Dropout technique to mitigate the domain gaps between different person Re-ID datasets. [33] T. Xiao, H. Li, W. Ouyang, and X. Wang. Learning deep feature representations with domain guided dropout for per- son re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1249– 1258, 2016. 2 Chen *et al*. [7] proposed quadru- plet loss to train a deep network, which aims to learn fea- tures	[17]Uses novel filter pairing neural network (FPNN) to jointly handle misalignments, photometric and geometric transforms, occlusions and background clutter. [1]"*We present a deep convolutional architecture with layers specially designed to address the problem of re-identification. Given a pair of images as input, our network outputs a similarity value indicating whether the two input images depict the same person. Novel elements of our architecture include a layer that computes cross-input neighborhood differences, which capture local relationships between the two input images based on mid- level features from each input image. A high-level summary of the outputs of this layer is computed by a layer of patch summary features, which are then spatially integrated in subsequent layers.*" [7] "*a quadruplet deep network using a margin-based online hard negative mining is proposed based on the quadruplet loss for the person ReID.* *a quadruplet loss, which can lead to the model output with a larger inter-class variation and a smaller intra-class variation compared to the triplet loss.*"

with large inter-class variations and smalle intra-class variations. [7] W. Chen, X. Chen, J. Zhang, and K. Huang. Beyond triplet loss: A deep quadruplet network for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2, 8 Zhao *et al*. [40] and Su *et al*. [29] integrated human pose information for tackling the pose variation prob- lem and improving feature learning capability. [40] H. Zhao, M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang. Spindle net: Person re-identification with hu- man body region guided feature decomposition and fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1077–1085, 2017. 2, 7, 8	

14. **Transferable Joint Attribute-Identity Deep Learning for Unsupervised Person Re-Identification**

(2) Using a pre-learned deep model on labelled source data but lacking an effective domain adaptation mechanism [59]; [59] H.-X. Yu, A. Wu, and W.-S. Zheng. Cross-view asymmetric metric learning for unsupervised person re-identification. In *ICCV*, 2017. 1, 2, 6, 7	

15. **Harmonious Attention Network for Person Re-Identification**

One common strategy is local patch calibra- tion and saliency weighting in pairwise image matching [48, 28, 51, 39]. However, these methods rely on hand- crafted features without deep learning jointly more expres- sive feature representations and matching metric holistically (end-to-end). A small number of attention deep learning models for re-id have been recently developed for reduc- ing the negative effect from poor detection and human pose change [19, 47, 30, 2]. Nevertheless, these deep methods implicitly assume the availability of large labelled training data by simply adopting existing deep architectures with high complexity in model design. [48] R. Zhao, W. Ouyang, and X. Wang.	[19] “*we propose to learn and localize deformable pedes- trian parts using Spatial Transformer Networks (STN) with novel spatial constraints. The learned body parts can re- lease some difficulties,* e.g. *pose variations and background clutters, in part-based representation. Finally, we inte- grate the representation learning processes of full body and body parts into a unified framework for person ReI- D through multi-class person identification tasks*“

Unsupervised salience learning for person re-identification. In *CVPR*, 2013. 1, 2 [19] D. Li, X. Chen, Z. Zhang, and K. Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *CVPR*, 2017. 1, 2, 6	
Recently, a few attention deep learning methods have been proposed to handle the matching misalignment chal- lenge in re-id [19, 47, 30, 18]. The common strategy of these methods is to incorporate a regional attention selec- tion sub-network into a deep re-id model. For example, Su et al. [30] integrate a separately trained pose detection model (from additional labelled pose ground-truth) into a part-based re-id model. Li et al. [19] design an end-to-end trainable part-aligning CNN network for locating latent dis- criminative regions (i.e. hard attention) and subsequently extract and exploit these regional features for performing re-id. Zhao et al. [47] exploit the Spatial Transformer Net- work [13] as the hard attention model for searching re-id discriminative parts given a pre-defined spatial constraint. [30] C. Su, J. Li, S. Zhang, J. Xing, W. Gao, and Q. Tian. Pose- driven deep convolutional model for person re-identification. In *ICCV*, 2017. 1, 2, 6	[30]"*we propose a Pose-driven Deep Convolutional (PDC) model to learn improved feature extraction and matching models from end to end. Our deep architecture explicitly leverages the hu- man part cues to alleviate the pose variations and learn robust feature representations from both the global image and different local parts. To match the features from glob- al human body and local body parts, a pose driven feature weighting sub-network is further designed to learn adaptive feature fusions. "*

16. **Efficient and Deep Person Re-Identification Using Multi-Level Similarity**

With the great success achieved by deep convolutional nets (ConvNets) in computer vision [18, 33, 12, 9, 13], some works have been proposed to address Person ReID with deep neural networks in an end-to-end fashion. One approach is to classify the images into different identities [7, 19, 30]. During testing, each image is represented by the output of final fully connected layer before the classifier. [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Annual Conference on Advances in Neural Information Pro- cessing Systems (NIPS)*, December 2012. [7] Y. Chen, X. Zhu, and S. Gong. Person re-identification by deep learning multi-scale representations. In *The IEEE In- ternational Conference on Computer Vision (ICCV)*, October 2017.	[18]"The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make train- ing faster, we used non-saturating neurons and a very efficient GPU implemen- tation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective " [7]"*In this work, we demonstrate the benefits of learn- ing multi-scale person appearance features using Convolu- tional Neural Networks (CNN) by aiming to jointly learn discriminative scale-specific*

	features and maximise multi- scale feature fusion selections in image pyramid inputs. Specifically, we formulate a novel Deep Pyramid Feature Learning (DPFL) CNN architecture for multi-scale appear- ance feature fusion optimised simultaneously by concurrent per-scale re-id losses and interactive cross-scale consen- sus regularisation in a closed-loop design. “
Recently, some published work show the promising power of deep Con- vNets in person ReID. [22] proposed a Siamese network that takes a pair of images to be compared. Convolutional layers are used to extracted visual features and product is used to indicate the similarity. [2] proposed an improved architecture where neighbor difference were used to mea- sure the similarity. [31] further extends this architecture by enlarging the neighbor search region and normalize the elements before computing product. All the above works formulate the Person ReID task as a binary classification problem. [2] E. Ahmed, M. Jones, and T. K. Marks. An improved deep learning architecture for person re-identification. In *The IEEE Conference on Computer Vision and Pattern Recog- nition (CVPR)*, June 2015. [31] A. Subramaniam, M. Chatterjee, and A. Mittal. Deep neural networks with inexact matching for person re-identification. In *Annual Conference on Advances in Neural Information Processing Systems (NIPS)*, December 2016.	[2] “*We present a deep convolutional architecture with layers specially designed to address the problem of re-identification. Given a pair of images as input, our network outputs a similarity value indicating whether the two input images depict the same person. Novel elements of our architecture include a layer that computes cross-input neighborhood differences, which capture local relationships between the two input images based on mid- level features from each input image. A high-level summary of the outputs of this layer is computed by a layer of patch summary features, which are then spatially integrated in subsequent layers.“* *[31] “In this work, we propose two CNN-based architectures for Person Re-Identification. In the first, given a pair of images, we extract feature maps from these images via multiple stages of convolution and pooling. A novel inexact matching technique then matches pixels in the first representation with those of the second. Furthermore, we search across a wider region in the second representation for matching. Our novel matching technique allows us to tackle the challenges posed by large viewpoint variations, illumination changes or partial occlusions.”*
Another interesting approach treats ReID as recognition problem and classifies the images to different identities di- rectly. [41, 30] extracted different body parts by human pose and combined local and global features for classifi- cation. [7, 27], on the other hand, considered the features at different scale. Among these works, [19] proposing to use STNs to find the meaningful local parts is similar to ours. However, our method has a different goal for the usage of STNs: we want to compute the similarity explicitly. The network structure is also distinct since we build a Siamese network and the final object is binary classification.	