URL: http://openaccess.thecvf.com/content_cvpr_2017/papers/Zhao_Spindle_Net_Person_CVPR_2017_paper.pdf

Main idea

针对ReID数据集中检测框和姿态变换造成的图片间的特征匹配错位问题，论文设计body region proposal network（RPN）在不同阶段提取出使用人体骨骼关键点提出相应人体部位，进而提取出相应部位的特征之后进行特征融合。特别地，论文采用树状融合的方法，根据语义重要性分多个阶段融合特征，使得具有鉴别力的重要语义特征被保存下来。

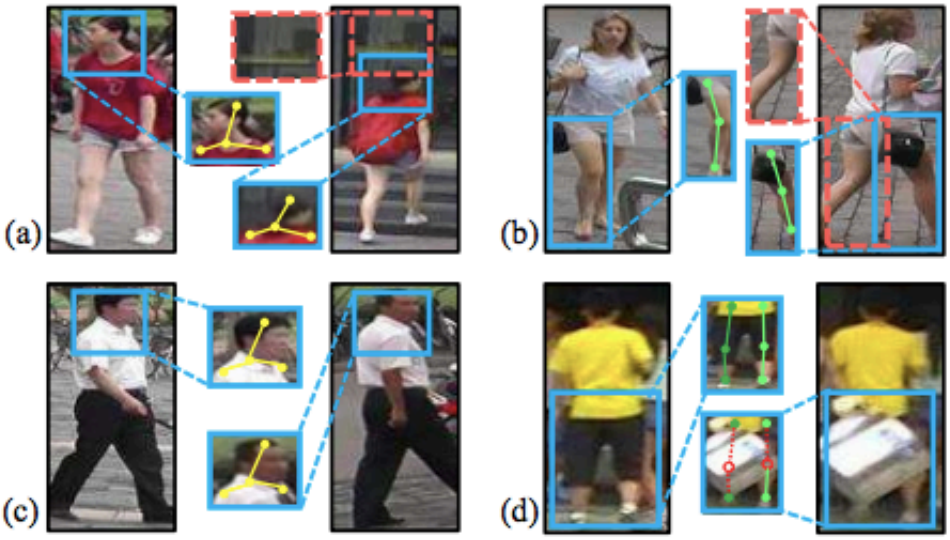

Figure 1. Challenges of person ReID. (a-b) Body region alignment across images. With human landmark information, the body regions shown in the blue boxes can be well aligned across images. However, without such information, directly matching based on location may results in ambiguities (the red boxes). (c) Detail information. With the local region features extracted from the blue boxes, detailed information can be captured and the two persons can be easily distinguished, even though their overall appearance are quite similar. (d) Occlusion. The lower-body region of the right image is occluded and the influence of the corresponding features should be weakened. All the examples are selected from the Market-1501 dataset [33].

Mode

（1）RPN

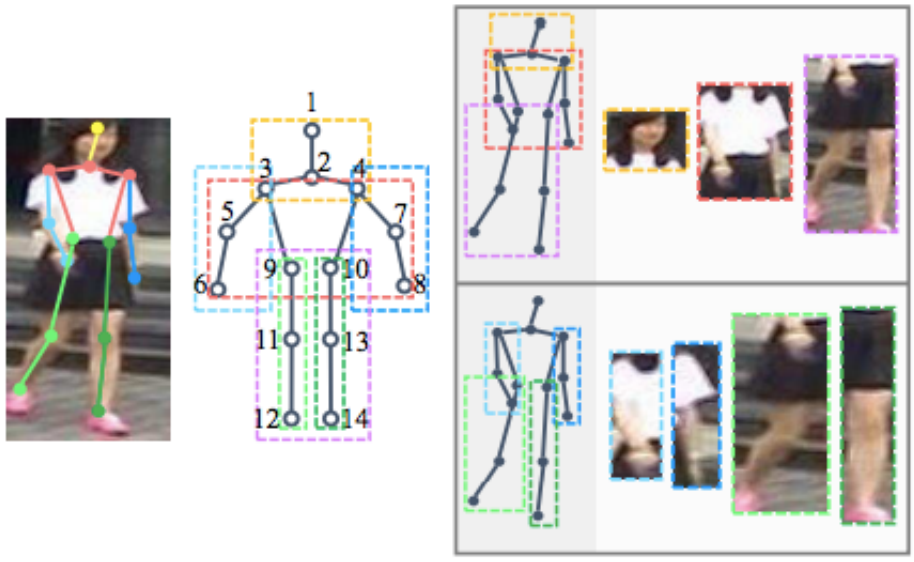

Figure 2. Illustration of the Region Proposal Network. (a) One sample image and the fourteen body joints. (b) The fourteen body joints are assigned to seven sets. (c) The seven body sub-regions proposed by the RPN from the corresponding body joint sets.

（2）model分为两个部分，特征提取部分和特征融合部分

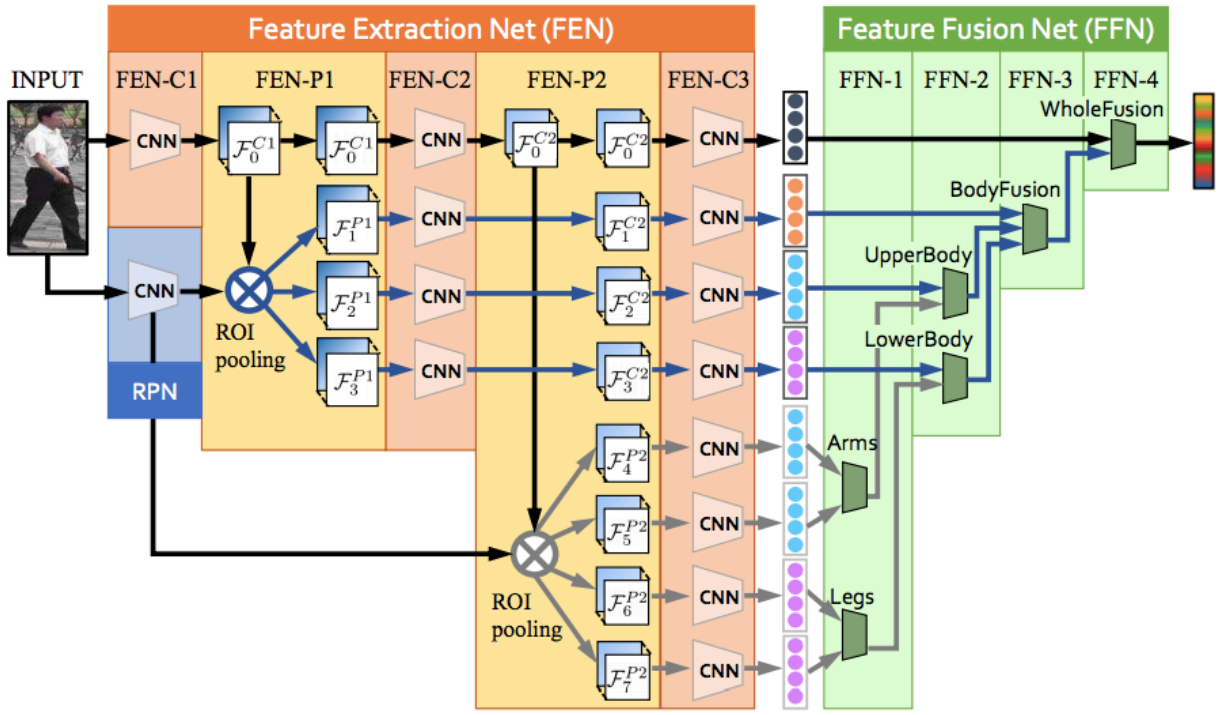

Figure 3. Flowchart of the proposed Spindle Net, including the Feature Extraction Net (FEN) and the Feature Fusion Net (FFN).

Experiment details

（1）RPN采用改进的CPM进行关键点定位，然后组合关键点得到四肢、头、上身、下身七个部分，采用和检测一样的ROI pooling提取出相应部分的特征。关键点定位网络在MPII人体姿态数据集上训练得来。 （2）语义信息相对重要的人体部件在第一阶段就开始提取特征，相对次要的四肢在第二阶段才提取特征。 （3）特征融合部分采用竞争机制，采用element-wise maximizaition操作选取元素级别大的

值，且四肢在第一阶段进行融合，头、上身、下身在后续阶段逐步融合。这样可以保留更多的重要的语义信息。

Visualization

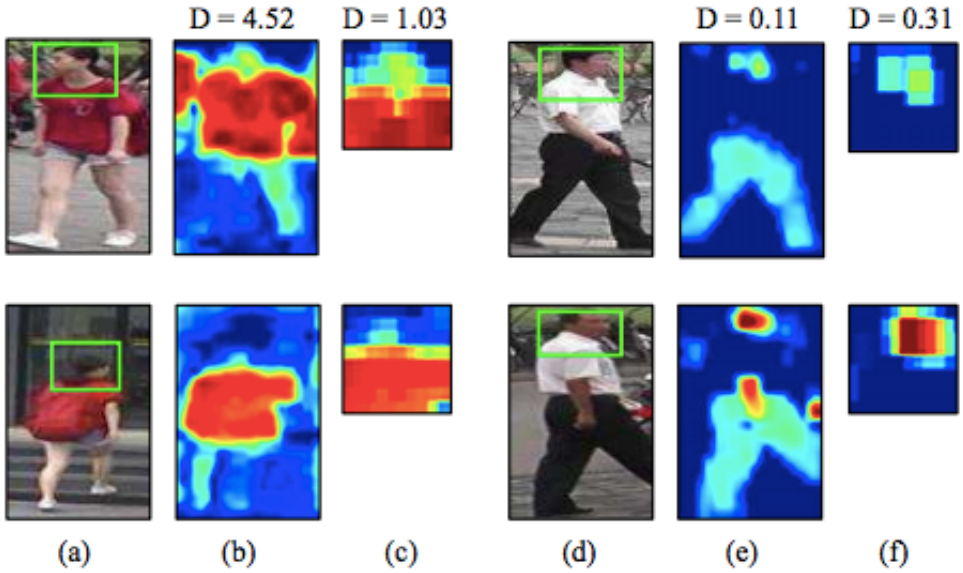

Figure 4. Two examples to demonstrate the effectiveness of the proposed sub-region features. (a) Two images of the same person. (b) Corresponding feature maps after FEN-C1. (c) Feature maps after FEN-P1. (D) Two similar persons. (b) Corresponding feature maps after FEN-C1. (c) Feature maps after FEN-P1. The average L_2 distances between feature maps are also listed.

由上图可以看出，通过人体部件的对齐，相同ID的图片的feature间的距离降低，不同ID的图片的feature间的距离增大，且更加关注差别大的部分（头部）。

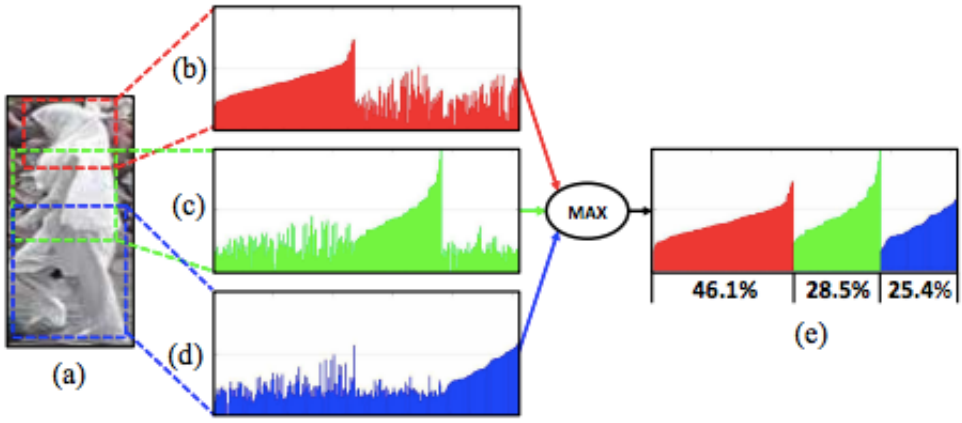

Figure 5. Illustration of feature fusion. Feature entries are sorted for better visualization. (a) Input image. (b-d) Three input feature vectors of the body fusion unit. The features of the head-shoulder region, the upper body region, and lower body region are marked in red, green and blue, respectively. (f) Result of the max operation. The head-shoulder features win 46.1% of the competition, much more than the other two region features in green and blue.

由上图可以看出最大竞争机制的特征融合的效果。（在图片中，头部占据更多的特征比例，因而在match时发挥更大的作用）

Thoughts

该论文融合了skeleton网络辅助特征对齐，且采用树状特征融合保留更多重要语义信息，具有非常不错的指导意义。