

行人重识别研究综述(三)

ReID的关键是如何识别和匹配两个人的特征信息和语义结构特点。这个过程主要分为两个步骤。

第一个步骤是特征提取，主流方法是CNN。

第二个步骤是特征Matching，核心方法有两种：一种是基于预先定义的位置，例如global, local stripes和grid patches，是比较启发式的；另一种是基于semantic region，借助Person parts, salient regions和attention regions，具有一定的语义含义。

从四个比较重要的领域，stripes方法，grids方法，attention方法以及pose方法来回顾ReID的进展。

(一)

Stripes角度主要有三个方面的工作：DeepMetric, DeepReID和AlignedReID。

DeepMetric把一幅图片强行分成三大块，每一大块做一个SCNN，再将各个部分整合。想法非常简单，实际应用过程中相对有效。

DeepReID更复杂，相当于把一个人的结构分成很多小块，每一个小块进行操作。这个方法比较直接，更加细致。缺陷是在识别较为复杂的情况时，或者任何人之间特征区分较差时会受到噪音干扰。

Stripes效果最好的是孙剑老师的工作AlignedReID，他们通过动态规划计算距离，需要动态匹配的过程，比较复杂，但效果不错。过程分为两个部分，一个是水平的pooling，一个是global pooling，再将两部分融合，得到local distance和global distance，再加入hard sample mining。

(二)

第二种思路是基于网格的方法，主要有两个工作。

首先是发表在CVPR2015的工作IDLA。它将两个图片转化匹配，认为在另一个图像的邻域网格总能找到匹配。在难以匹配的情况下，可以到邻域寻找匹配，所以性能提高很多。

下一个是李玺老师的工作。他们认为之前Pre-define的过程有问题，匹配过程很难适应大场景的变换，所以基于这个不足做了一些改进：将网络结构转化成一个空间金字塔，在一层匹配不了的情况下，到上一层匹配。

(三)

第三个是Attention方法，它借助自然语言和图像语言做特征选择。

第一个是李玺老师的工作，发表在ICCV2017。这个工作是简单但是非常有效的。在将一个人进行匹配时不是所有区域都参与到匹配中，他们加入attention map，来自动发现适合做re-identify的pattern，再做triplet loss，能够在性能上提高7到8个点。这个模型不需要改变任何网络结构，只需要加入attention map，很适合工程应用。

还有工作基于这个思想做了一些改进。比如很复杂的HPNet，有多层attention，attention map有多个layer，还有遗忘skip的功能，需要把很多过程整合起来得到一个结果。最后HPNet相比上述的模型有进一步的提高，但是在market数据集上比上述模型的效果差一点。方法越来越复杂，可能在某个数据集上表现越来越好，但是可能泛化能力越来越差。而上述模型的模型简单，泛化能力强。

另一个对于attention regions learning的方法进一步深化，定义了两种attention，一种是hard attention，有主

干道，一种是soft attention，加入一些分支，然后把soft和hard枝干融合。最后只放出market数据集的结果，相比HA-CNN提高很多，但没有放出CHUK03的结果，无法重复实验。

(四)

最后是基于Pose的方法，让ReID方法具有更强可解释性。

PDC发表在ICCV2017，它将Pose信息嵌入到结构网络中，生成一个modified结构图像，然后对这个结构图像进行识别，效果会有极大提高。

PSE引入视角关系，将多视角结构进行整合，最后得到的结果也还不错。