Learning Context Graph for Person Search

Yichao Yan Qiang Zhang Bingbing Ni Wendong Zhang Minghao Xu Xiaokang Yang

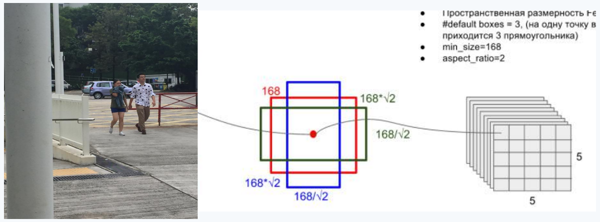
CVPR19 Oral

https://arxiv.org/pdf/1904.01830.pdf

Intelligent Information Fusion Research Group

行人重识别需要两步：

1. 目标检测，将图像块抠出来存储，



2. 之后抠出的图像块再去重识别。

这是两个阶段的，会浪费一定的存储空间和时间。

行人检索 将上面两个阶段合为一体

目标检测时抽的feature，之后再经过reid的相关网络进行重识别。速度和空间上还是比较实惠一点。

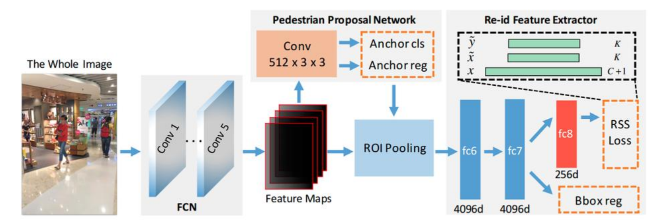


Figure 4. Overview of our framework. Given a whole image, we first utilize a fully convolutional network to extract feature maps. Then we deploy a convolution layer with 512 3 × 3 filters on the top of the feature maps, followed by sibling anchor classification (denoted by Anchor cls) and regression layers (denoted by Anchor reg) to predict pedestrian ROIs. These ROIs are then used to pool the feature vector for each candidate box on the convolutional feature maps. Three fully connected layers are utilized to produce the final feature vector (fc8) for computing distances. The boxes with dashed orange borders represent the loss layers

有趣的引入

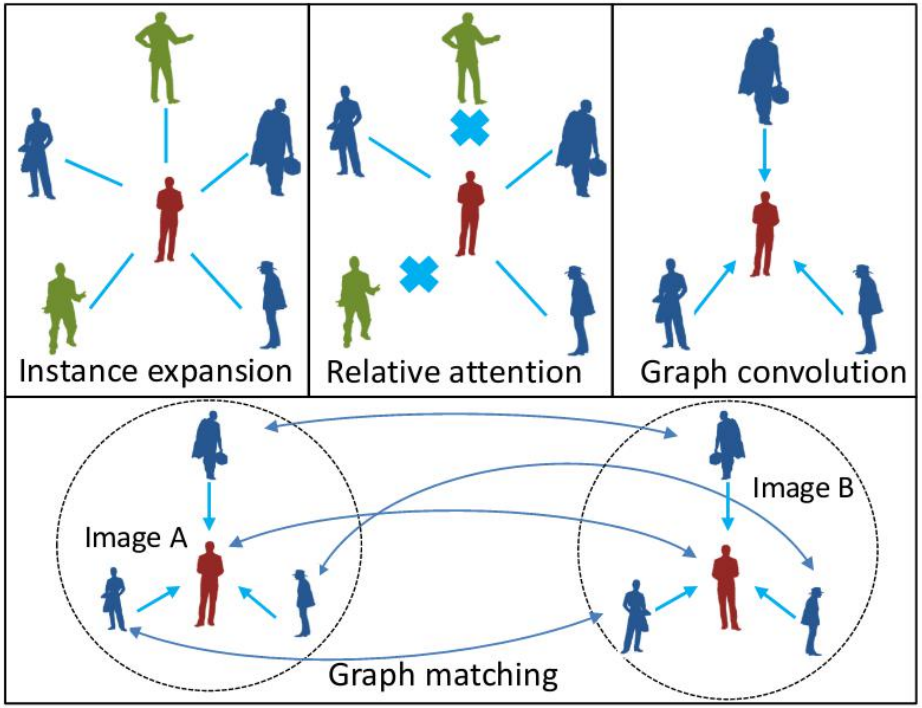


Figure 1. Illustration of the proposed framework.

一个人会与其他人一起出现在一个摄像头下面，那么这些人也会有概率跟你一起走向另外一个摄像头。

Then: **同行人，可以提供reid帮助?**
不同摄像头下的内部结构

有趣的引入

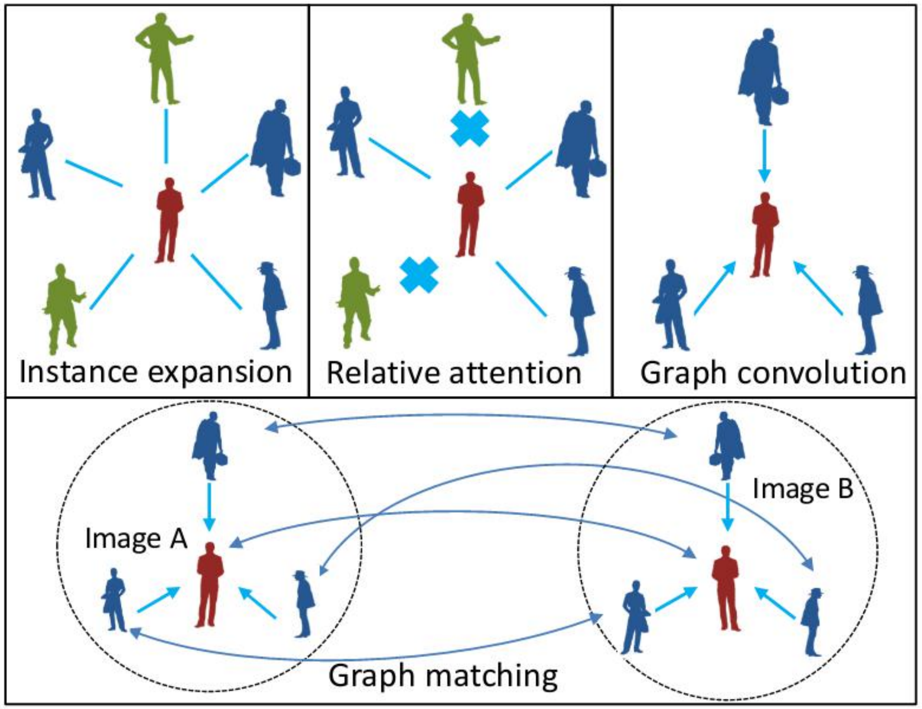


Figure 1. Illustration of the proposed framework.

→ 肯定会引入一些噪声，需要过滤

Conclusion: **同行人，是可以提供reid帮助。**

有趣的引入

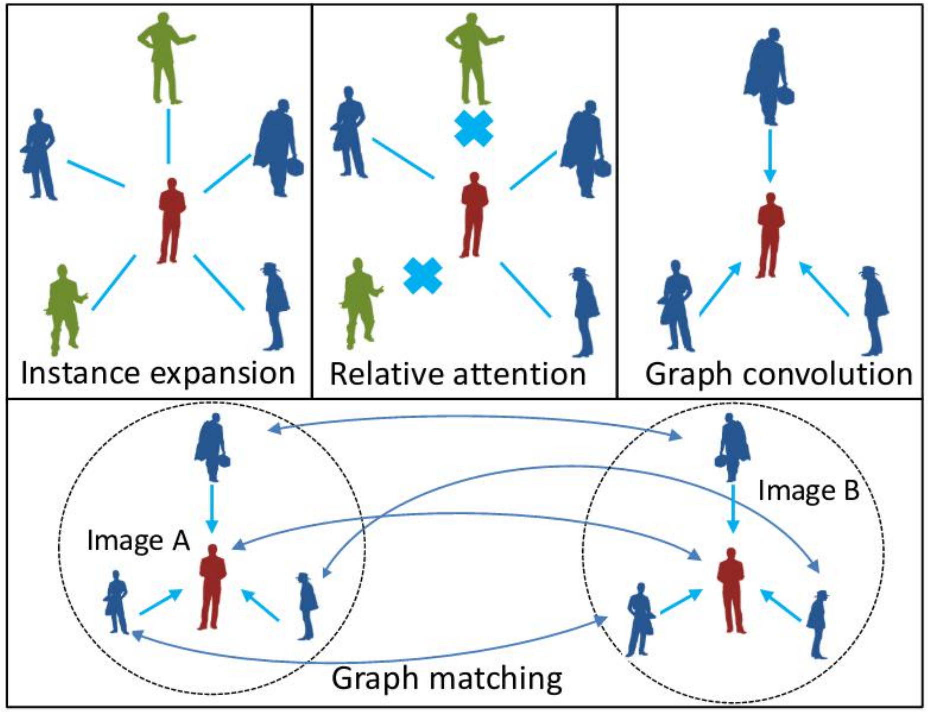


Figure 1. Illustration of the proposed framework.

所以，对于这个引入，我们知道作者对要解决**行人检索**问题的两个难点就是：

1. 怎么筛出与一个人同行的其他人?

2. 怎么利用这些同行的人来进行reid?

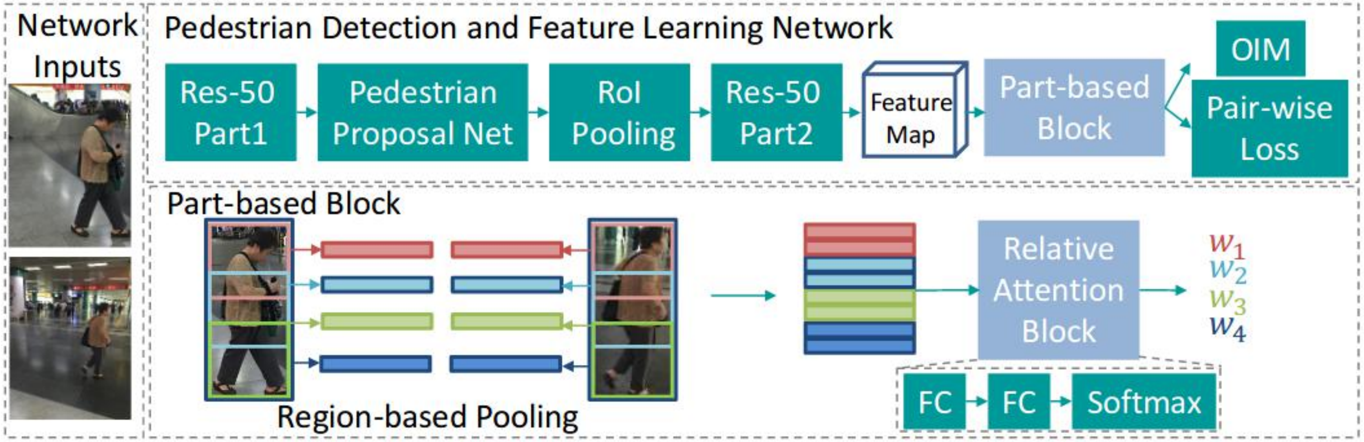


Figure 2. Architecture of the detection and part-based feature learning framework.

单实例训练
CVPR2017

多实例训练

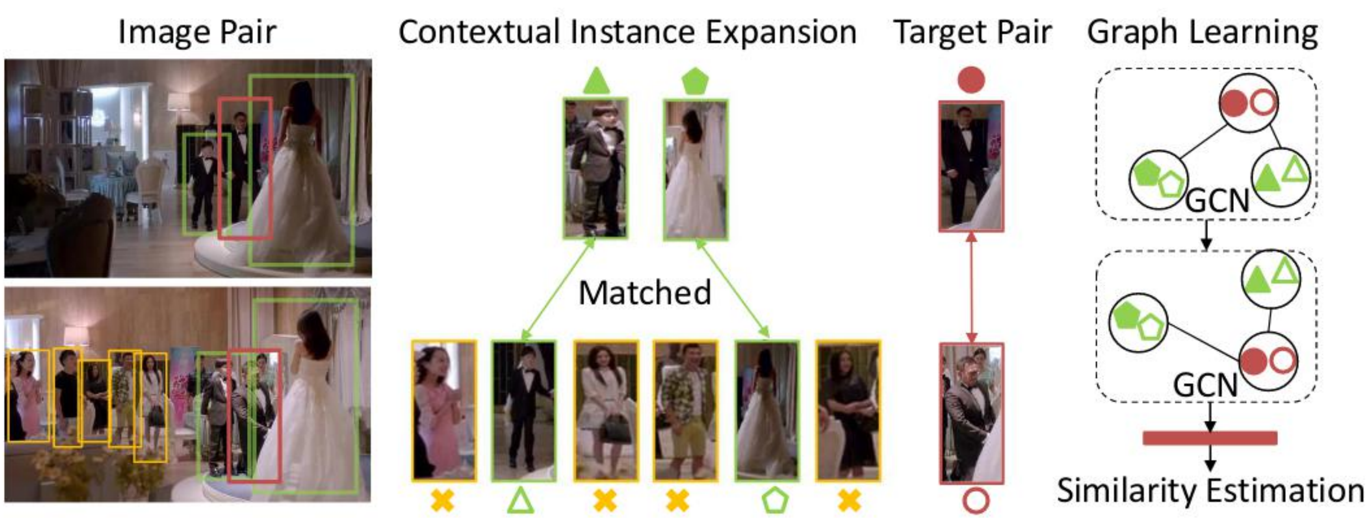


Figure 3. Architecture of the detection and part-based feature learning framework.

多实例训练

总体框架

由于之后需要取top-k对图片。
而最近块的方法又不错

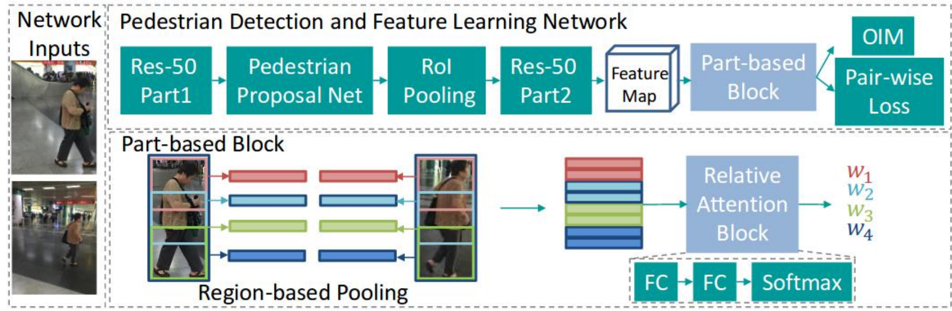
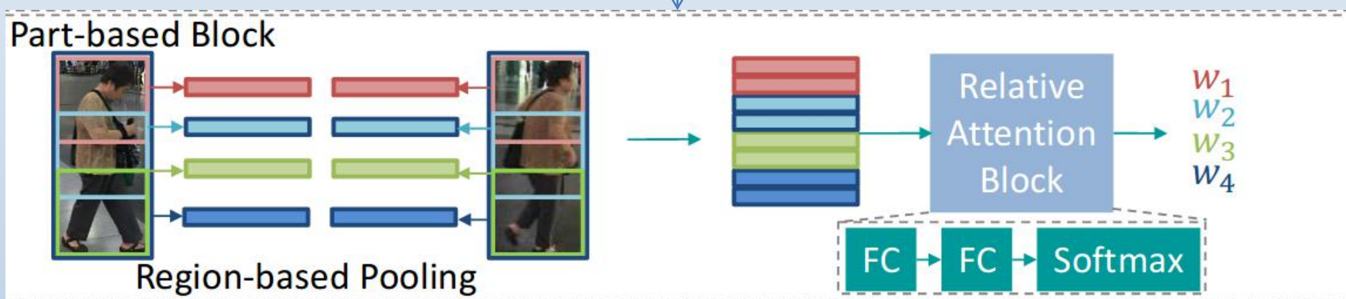


Figure 2. Architecture of the detection and part-based feature learning framework.



$$s(i,j) = \sum_{r=1}^{R} w_r cos(\mathbf{x}_i^r, \mathbf{x}_j^r),$$

权重需要综合考虑

权重需要综合考虑

positive pair

negative pair

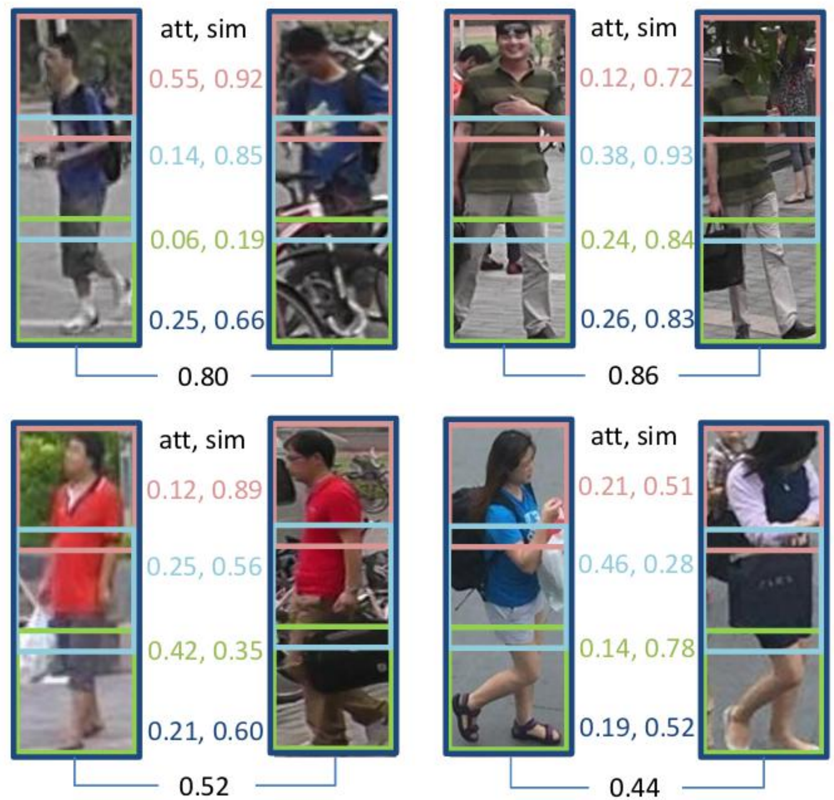


Figure 4. Attention examples. “att” denotes the attention wights of different body parts, “sim” denotes similarity between body parts. The values under image pairs denote the overall similarity.

Table 2. Component analysis results.

Dataset	CUHK-SYSU		PRW	
Methods	**top-1(%)**	**mAP(%)**	**top-1(%)**	**mAP(%)**
uniform	79.2	75.9	65.4	24.5
attention	82.7	80.2	67.8	27.8

Figure 3. Architecture of the detection and part-based feature learning framework.

$$\mathbf{Z}^{(l+1)} = \sigma(\hat{\mathbf{A}}\mathbf{Z}^{(l)}\mathbf{W}^{(l)}),$$

if use Siamese GCN

prevents the contextual information to propagate between graphs

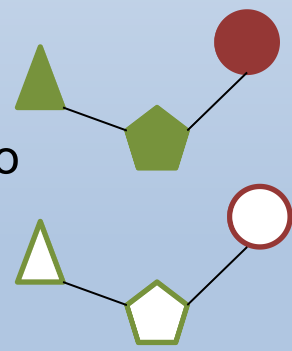


Table 5. Results of different graph structures

Dataset	CUHK-SYSU		PRW	
Methods	**top-1**	**mAP**	**top-1**	**mAP**
DisGCN [20]	83.4	81.3	69.8	29.5
Ours	86.5	84.1	73.6	33.4

Table 2. Component analysis results.

Dataset	CUHK-SYSU		PRW	
Methods	**top-1(%)**	**mAP(%)**	**top-1(%)**	**mAP(%)**
uniform	79.2	75.9	65.4	24.5
attention	82.7	80.2	67.8	27.8
graph	86.5	84.1	73.6	33.4

Table 3. Comparison of results on CUHK-SYSU with gallery size of 100

Method	mAP(%)	top-1(%)
CNN + DSIFT + Euclidean [50]	34.5	39.4
CNN + DSIFT + KISSME [50][18]	47.8	53.6
CNN + BoW + Cosine [53]	56.9	62.3
CNN + LOMO + XQDA[22]	68.9	74.1
OIM [42]	75.5	78.7
IAN (ResNet-50) [41]	76.3	80.1
IAN (ResNet-101) [41]	77.2	80.5
NPSM [26]	77.9	81.2
I-Net [15]	79.5	81.5
CNN_v + MGTS[7]	83.0	83.7
Ours	**84.1**	**86.5**

Table 4. Comparison of results on PRW

Method	mAP(%)	top-1(%)
OIM [42]	21.3	49.9
IAN (ResNet-101) [41]	23.0	61.9
NPSM [26]	24.2	53.1
CNN_v + MGTS[7]	32.6	72.1
Ours	**33.4**	**73.6**

实验

Thank You~~!!