URL: https://arxiv.org/abs/1707.00408

Main idea

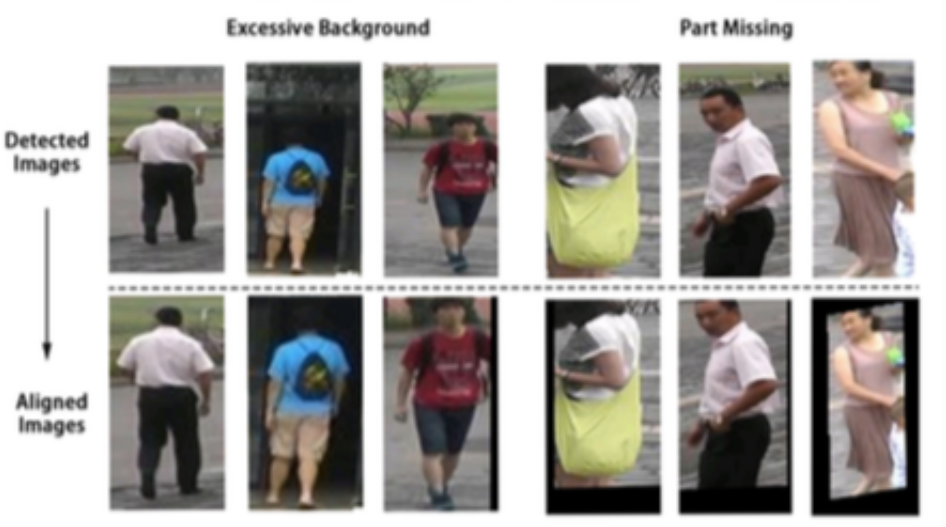

Fig. 1 Sample images influenced by detector errors (the first row) which are aligned by the proposed method (the second row). Two types of errors are shown: excessive background and part missing. We show that the pedestrian alignment network (PAN) corrects the misalignment problem by 1) removing extra background or 2) padding zeros to the image borders. PAN reduces the scale and position variance, and the aligned output thus benefit the subsequent matching step.

这篇论文提出了一个观察结果：网络学习过程中具有更加关注人体的注意力机制，相较于背景，它们在人体上具有更高的激活值。因此论文提出采用STN对图片进行空间变化对人体特征进行对齐，以解决ReID中行人在图片中的尺度不一和人体部件缺失的问题。

Model

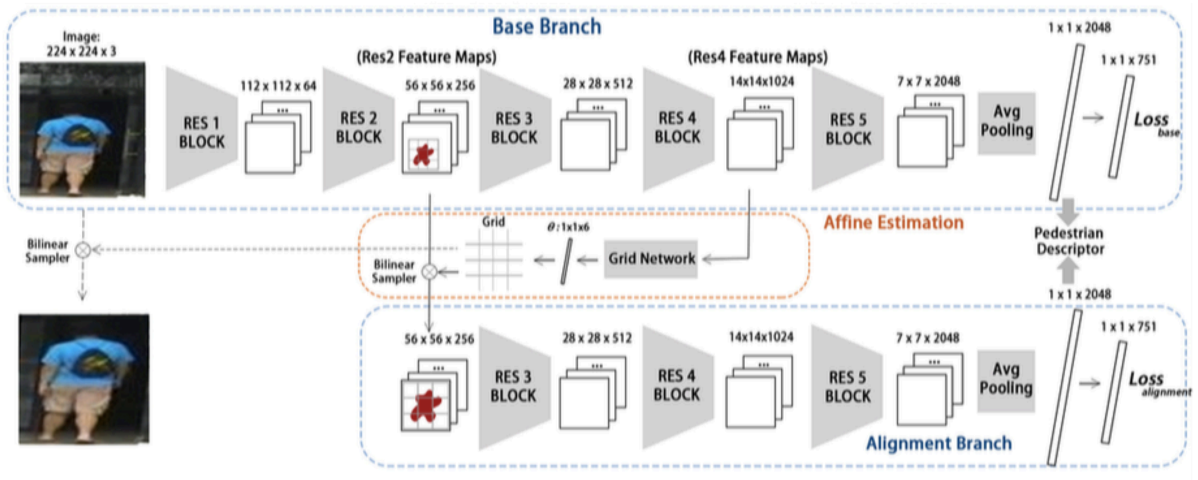

Fig. 2 Architecture of the pedestrian alignment network (PAN). It consists of two identification networks (blue) and an affine estimation network (orange). The base branch predicts the identities from the original image. We use the high-level feature maps of the base branch (Res4 Feature Maps) to predict the grid. Then the grid is applied to the low-level feature maps (Res2 Feature Maps) to re-localize the pedestrian (red star). The alignment stream then receives the aligned feature maps to identify the person again. Note that we do not perform alignment on the original images (dotted arrow) as previously done in [Jaderberg et al., 2015] but directly on the feature maps. In the training phase, the model minimizes two identification losses. In the test phase, we concatenate two 1 × 1 × 2048 FC embeddings to form a 4096-dim pedestrian descriptor for retrieval.

- 整个模型分为三个部分：base分支，alignment分支和affine estimation分支。
- base和alignment分支都有各自的分类loss。
- alignment分支和base网络共享参数，且alignment分支的输入是由affine estimation分支生成的对齐后的feature map。
- affine estimation分支是一个STN去除大多数背景，并对丢失的人体部件补0，进而减小人在图片中的尺度差异和检测错误导致的姿态差异。
- STN使用的是第二和第四个block的feature map，且并非对原图片进行对齐而是对feature map进行特征对齐，以减小网络参数和运行时间。

Thoughts

这篇文章融合了STN的思想对行人进行空间对齐，减小了person scale和part missing的影响，具有一定的借鉴意义。