### 目标跟踪

### 数据格式

训练数据:

每一个目标应包含以下信息:

frame\_index: 当前目标位于视频第几帧

track\_id: 目标唯一识别标识, 不同帧间同一目标拥有相同track\_id

x1, y1, w, h: 目标所在位置

x:

y: 目标类别

z: 目标可见比例, 用于衡量遮挡情况

测试数据:

测试数据需要给出跟踪目标的位置(使用目标检测模型获得)

每一个目标应包含以下信息:

frame\_index: 当前目标位于视频第几帧

x1, y1, w, h: 目标所在位置

confidence: 检测目标置信度

(可以加入目标类别, data association阶段在类别一致的基础上进行跟踪)

### 算法流程

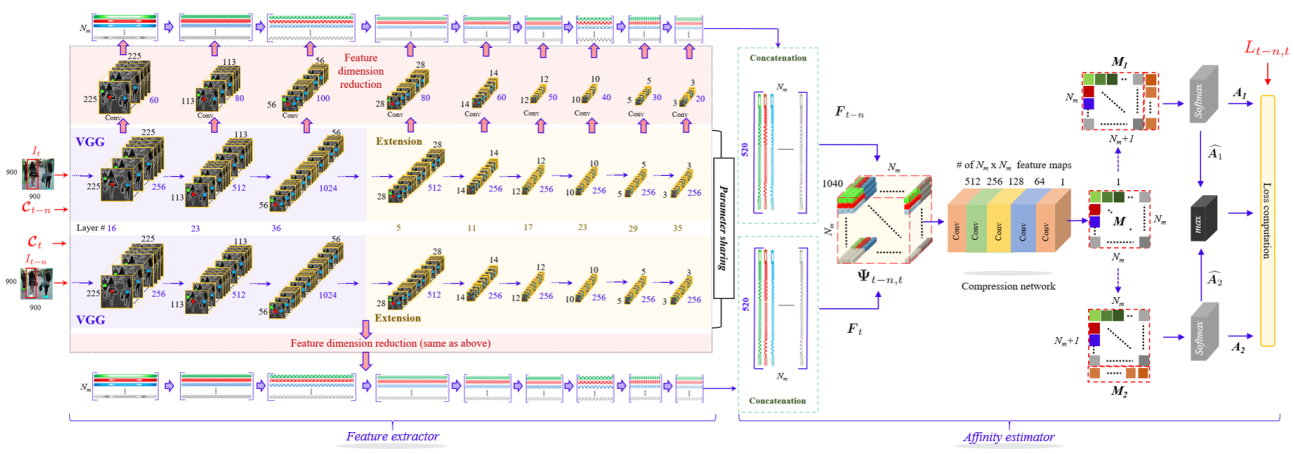
基于Siamese网络的目标跟踪方法.

**训练过程:**

模型输入为成对图像, 包括img1, bbox1, img2, bbox2, target(T). 其中target为根据track\_id生成的二值矩阵, 表示两帧图像中目标的对应关系, 最后一行一列表示没有目标与其对应.

1) 首先输入图像经VGG网络提取不同维度的特征, 然后拼接到一起得到feature

2) 然后根据输入bbox提取对应位置的特征信息, 得到每个目标的特征表示 N \* feature



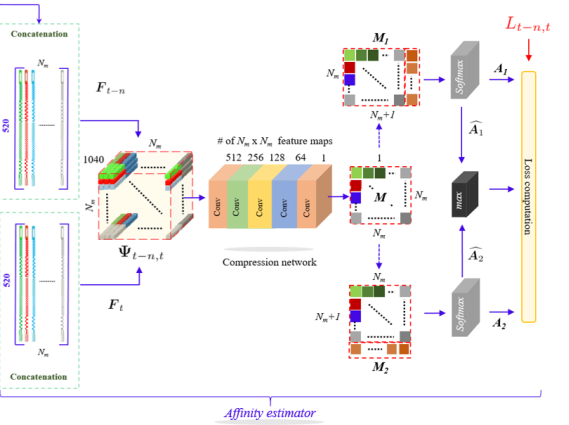
3) 计算两帧图像之间特征相似度和, 生成相似度矩阵, 考虑到存在目标没有对应的情况, 添加一行一列表示不存在对应关系

4) 计算target和相似矩阵之间loss

对M的行与列分别做softmax函数映射, 相似度归一化, 得到A

预测矩阵A与目标矩阵T之间的差异由4部分组成:

其中表示前向后的映射关系为的矩阵, 表示后向前的映射关系为的矩阵, 为的矩阵



**预测过程:**

预测过程与训练前向基本一致, 为了提升跟踪精度, 避免track\_id频繁切换, 设置记忆池机制, 对于位于记忆区间的每一个目标, 存储M个样本, 新目标与记忆池中目标相似度超过阈值则认为是同一目标.

在预测阶段, 按照视频顺序向模型中输入同一视频中帧图像image和检测结果bbox

1) 首先输入图像经过VGG网络提取特征, 并根据bbox获得对应位置特征信息, 与训练过程1) 2)过程一致

2) 当前得到的目标feature和记忆池中的feature计算相似度, 与训练过程3)相似,

得到相似度矩阵 其中表示记忆池中目标数, 表示当前帧检测目标个数.

3) 对于当前相似度矩阵, 将-similarity看做cost, 使用匈牙利算法进行目标匹配, 并在匹配结果的基础上使用IOU进行过滤(假设连续两帧之间同一目标位置变化不大), 得到当前帧的跟踪结果.

4) 将bbox和feature加入记忆池, 并删除超过记忆范围的目标信息.

5) 最终该视频中每一个检测目标分配一个track\_id, 同一目标拥有相同的track\_id.

### 实验结果

使用MOT17数据集进行模型性能评估

训练集:

sequence-02 600帧

sequence-04 1020帧

sequence-05 837帧

sequence-09 525帧

sequence-10 654帧

sequence-11 900帧

sequence-13 750帧

测试集:

sequence-01 450帧

sequence-03 1500帧

sequence-06 1194帧

sequence-07 500帧

sequence-08 625帧

sequence-12 900帧

sequence-14 750帧

经过40个epoch训练, 使用FRCNN的检测结果, 最终实验结果为

avg mota: 46.044 %

avg motp: 88.709 %

total fn: 52031.0

total fp: 7373.0

total idsw: 1187.0

total\_num\_objects: 112297.0

当前处理速度为7fps(GPU), 图像中目标较少时处理速度会稍快.

对于遮挡目标, 在遮挡过程中还不能做到很好跟踪, 但是目标再次出现时, 可以重新识别目标.

|  |  |  |
| --- | --- | --- |
|  |  |  |

### 不足与改进:

注意到经过简单训练(5个epoch左右), 模型即可达到较高精度, 后续训练性能提升较小, 考虑是hard example没有得到有效学习, 可以尝试控制正负样本比例或者使用focal loss等balance方法.

training阶段, 提取目标特征可以参考Roi-Pooling的方法, 将bbox映射到特征图, 并截取对应位置的特征信息, 然后缩放到相同尺寸.

inference阶段记忆机制的引入提升了Recall, 但是也增加了FP, 使用iou过滤部分目标可以得到改善; 得到相似度矩阵为, 对于unmatched情况判断不够灵活, 考虑如何生成的相关矩阵.

### Inference

参考项目github: <https://github.com/shijieS/SST>

使用数据集: MOT17 <https://motchallenge.net/data/MOT17/>