基于自监督和多任务学习的视频异常检测

1. 要解决什么问题

通过单个代理任务建模异常事件检测，如未来帧预测，由于代理任务和实际任务（异常检测）之间缺乏完美的对齐，是次优的。

1. 用了什么方法

为了减少模型对于异常检测任务的不对齐，提出通过多个代理任务联合优化模型。

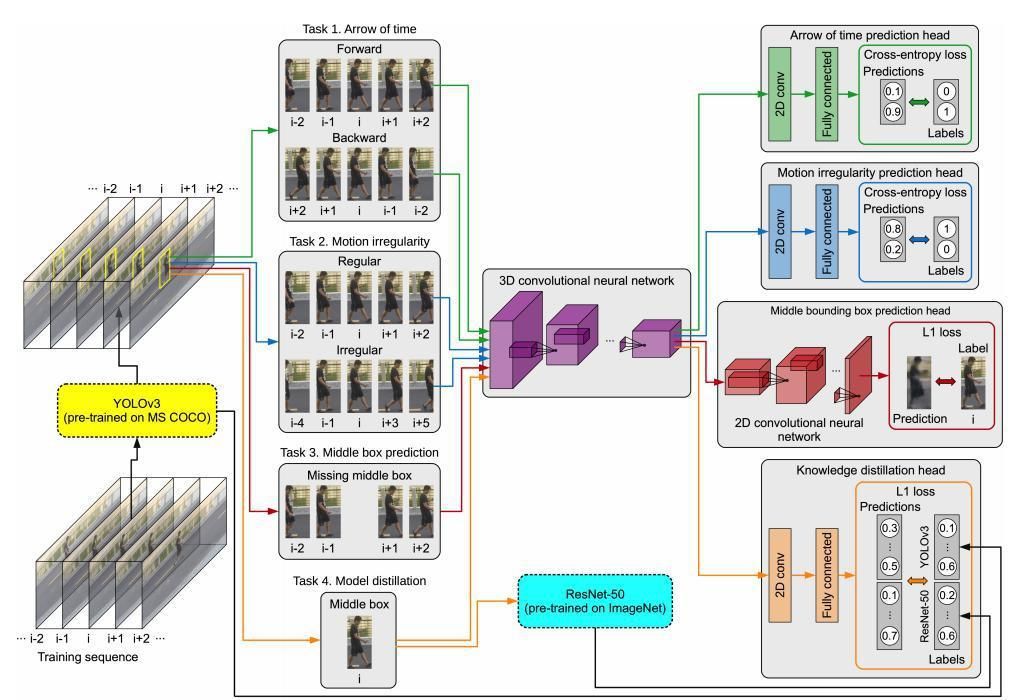


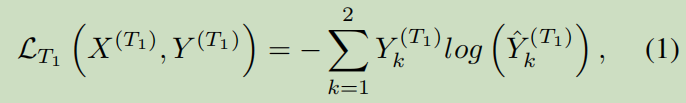
图1、异常检测框架是基于自监督和多任务学习。首先，使用目标检测器YOLOV3检测视频中的目标。对于每个目标，设计了三个自监督任务（学习时间的方向，预测运动的不规律和预测中间位置的物体外观）和一个知识蒸馏任务（使用YOLOV3和ResNet-50作为老师）。联合四个任务训练一个3D卷积网络。用虚线表示的模型是预先训练的。

**任务1：时间的方向**

对每个以对象为中心的序列，生成两个标注的训练样本。一个是按照时间顺序取出这些帧，标记为正向运动。第二个是以相反的顺序取出这些帧，标记为反向运动。

在推理时，本文假设对于异常行为的目标，预测时间的方向更难。

损失函数为

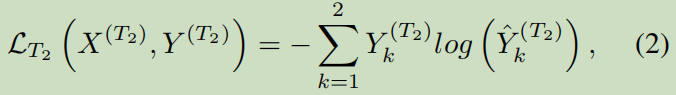


**任务2：运动不规律**

本文假设一些异常可以通过不规律的运动模式识别，所以训练一个模型来预测一个以对象为中心的序列是连续的或断断续续的（一些帧被跳过）。

对每个以对象为中心的序列，生成两个标注的训练样本。第一个样本获取从i-t到i+t的连续帧。通过保留视频帧i，然后从过去帧中随机增加t帧和从未来帧中随机增加t帧来生成间歇帧。随机跳帧的范围为{1，2，3，4}。

损失函数为

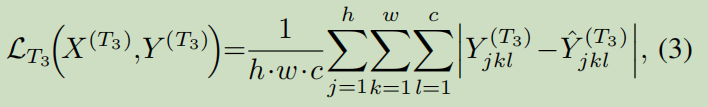


**任务3：中间帧的边界框预测**

通过前后帧重构出当前帧的图像。

当遇到异常的运动时，如一个人在跑动，则以那个人为中心的输入序列没有足够的信息提供给模型来准确的重构出中间画面，这样被标记为异常。

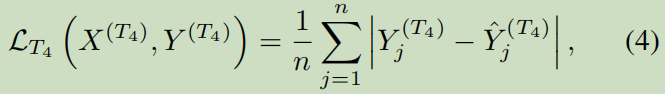
损失函数为



**任务4：模型蒸馏**

在推理时，本文假设当遇到一个外观不寻常的物体或者训练中没见过的物体类别的时候，在学生模型和YOLOV3老师模型之间出现较高的预测差异。

损失函数为



**联合损失**



其中λє(0,1]，是正则化知识蒸馏任务重要性的权重。在实验中观察到Lt4的值一般比其它的损失值有更高的数量级，在没有正则化的情况下主导着联合损失。在将我们的框架应用于异常检测任务前，根据联合损失的验证值微调λ的值。

**推理阶段**

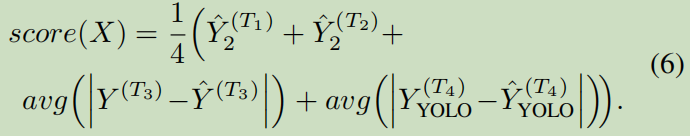
对时间方向的代理任务，采用时间序列向后移动的概率作为异常得分。

对运动不规律的任务，无间隔的测试序列X是断断续续的概率作为好的异常指标。

对于中间边界框预测的重建任务，将重建和真实目标的平均绝对误差作为异常分数。

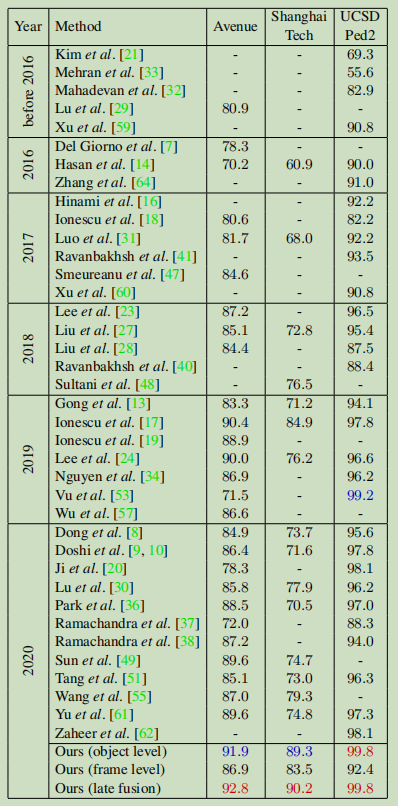
对于知识蒸馏任务，将YOLOV3预测的类别概率和该模型预测的类别概率的绝对误差作为异常指标。

最终一个目标的异常得分为每个预测头的异常得分的平均值



接着，重新装配检测到的目标到每一帧的像素级别的异常映射中。因此，在任意给定帧中可以容易地定位异常区域。为了建立一个平滑的像素级别的异常映射，使用3D均值滤波。某一帧的异常得分通过对应异常映射中的最大得分给出。最终视频帧级别的异常得分通过应用了一个时间上的高斯滤波后获取。

1. 效果如何



在Avenue\Ped2\Shanghai Tech数据集上都取得了AUC超过90%的成绩，可以说是碾压了之前的异常检测算法。

轻量的模型推理单个目标的异常得分耗时6ms。YOLOV3模型每帧耗时26ms检测目标。从目标级别的异常得分重新分配成异常映射耗时1ms。所有组件就位，框架以23FPS平均每帧5个目标的速度运行。在单卡11GB的1080Ti上测得。

1. 还存在什么问题和可借鉴的地方

该框架可以看成四个模型集成的结果。

对异常行为的判定应该是根据开源数据集经过精心的设计，泛化能力存疑。

模型的可解释性基于人工设定的条件，超出该范围的异常行为检测结果未知。

对于不同的异常检测场景，异常的类别也不同，根据不同场景设置不同的代理任务或者添加更多能反映异常的代理任务应该都能提升模型的性能。