**一 作者介绍的可以用来做异常检测的算法：（待深入了解）**

1 搭建具体的识别器(detector) 比如vioence detector 和 traffic accident detector . 局限性：只能检测特定的异常，无法检测其它的异常。无法应用到实践中

2 使用基于稀疏编码的方法。用正常的视频的小部分组成一个超字典，用这些字典组成一个正常的时间，如果不能组成正常的时间就认为它是异常的。 局限性：由于真实世界中，变化很巨大，所以该方法存在局限性

3 一些其它有关方法 ： rank-SVM、 successive linear programming 、deep ranking networks，d deep auto-encoder based approach

**二 作者论文里的一些重要观点和贡献**

1 提出了一个 multiple instance learning (MIL) 模型 2 提出了一个有挑战性的dataset

2 本问题目标是、在偏离正常情况的时候发出信号，确定发出异常的时间窗口。重点是确定异常发生，一旦检测到了异常可以用分类技术区分特定的异常。

3 作者认为只检测一种特定的异常（如交通事故）无法适用于全局的检测

4 作者提出，由于真实世界是复杂的，所以异常检测不能依赖于事故的先验知识，也就是说现实中的异常检测应该做到尽量少的监督 甚至是无监督

5 本文作者对 低监督的 data进行学习，就是视频级别的label ,只知道哪个视频有问题但不知道视频哪里有问题

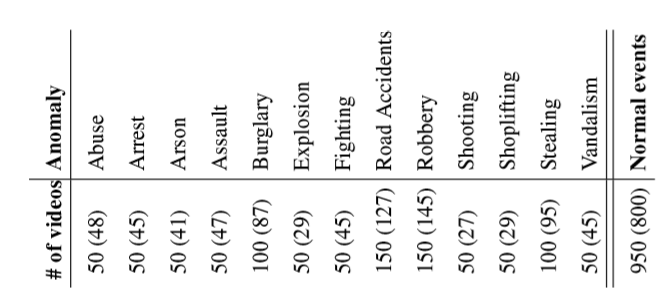
7 本文作者使用了MIL模型，把视频看作包，把视频片段看作包里的物体（instance）训练了一个异常排序模型

8 作者把这个问题看成一个回归问题，作者在框架内把特征向量映射为 异常的分数,作者强调了他们使用了弱标签的视频文件（只在视频级别进行标准而不是片段级别）

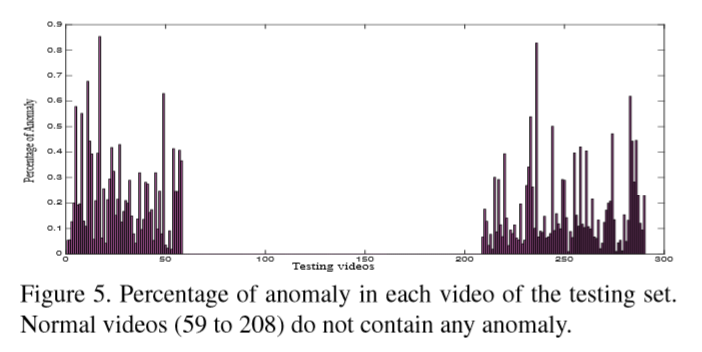
9 模型 在训练的时候将监控视频划分为固定数目的的片段，把他们当作物体放在包里 使用了positive包和 negative包 ，使用了deep MIL ranking损失函数

**三 模型**

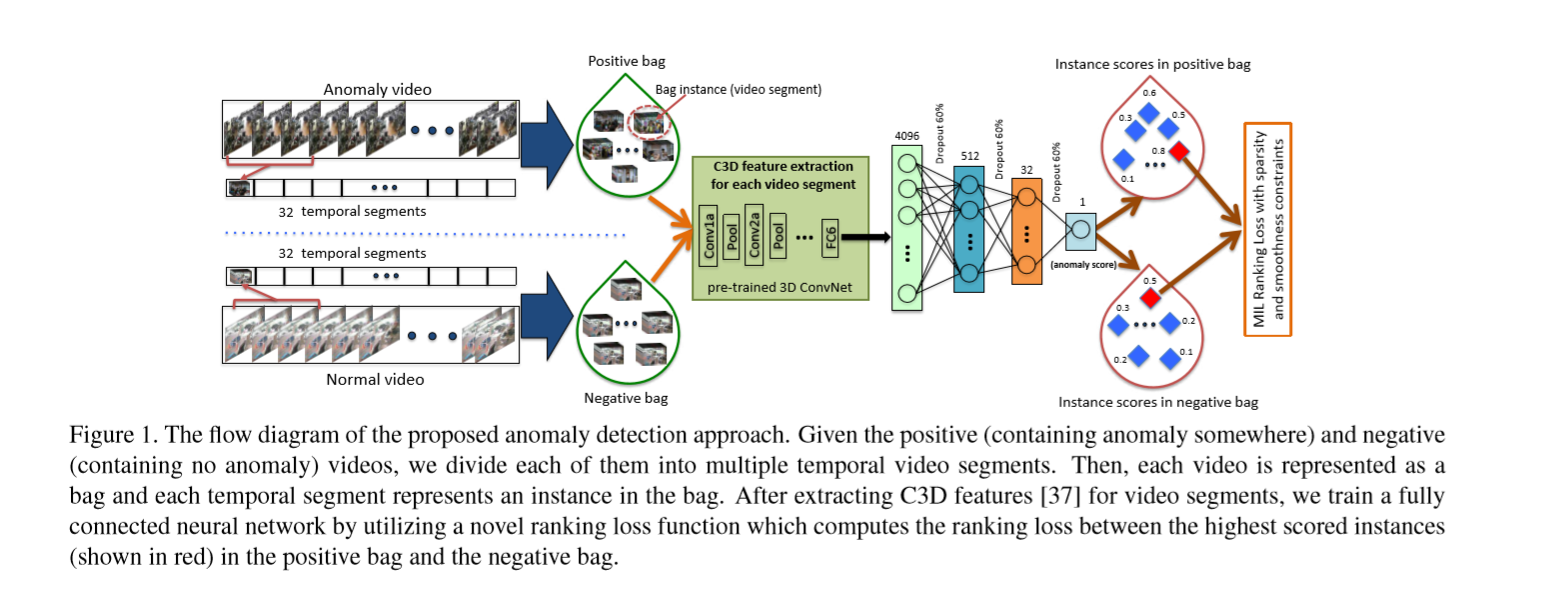
数据集：该模型的dataset是一个弱监督的数据集，只有video-level的 label，就是只知道那个视频里含有异常,但是不知道 视频里的哪一个时间窗口有异常。



（数据集视频分布）



模型结构图：



模型和问题分析：

1 数据结构

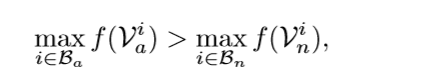
由于只采用在video-level，作者定义了新的结构。把video看成一个bag，把视频里的片段看成bag里的 instance，含有异常的video是 positive bag，相反没有异常的video 是 negative bag。作者将一个 video 分割为**无重叠**的片段，这些片段就是instance，所有的**视频片段数目相同。**

2 特征表示/特征提取

作者针对视频片段的特征进行学习，所以首先需要提取视频片段的特征，作者使用conv3D,3D的卷积层进行特征提取。

3 目标

因为真实世界中对于异常的标定是比较主观的。所以该模型的任务是 回归（regression ）而非分类。回归的目标让**异常的视频**的得分 和**正常视频**的得分差距尽可能大，



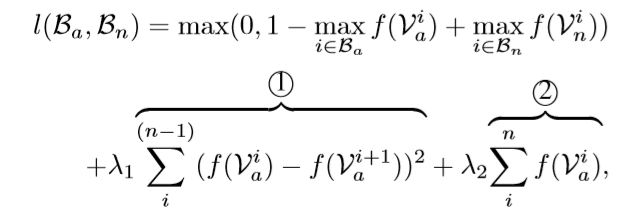
（这里 和是bag，a下标代表异常，n下标代表正常），  和是 bag里的instance， 表示片段的得分）

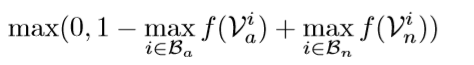
由于只有video-level的 label，我们不知道具体的instance的label，因此选择每个bag里的最大得分的instance进行比较。同时值得注意的是在这个式子中，异常包里的片段很有可能的能够得到很高的分数，而选择正常bag里得分最大的instance，能够避免误判一些正常的情况。（这些正常的情况和异常的很像）

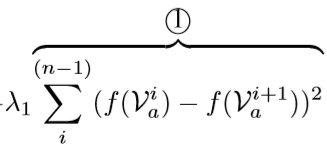
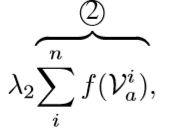
4 损失函数



其中：



是回归目标项

、、是惩罚项

上述三项中：

1. 让权值尽可能低，因为在神经网络中，我们认为权值低的模型，模型复杂度低。
2. 是**平滑项，**为什么需要平滑项？因为视频中，在时间上片段之间是连续的，因此相邻的片段之间的**差距不会很大**，换言之片段之间应该是**平滑的，**所以需要平滑项。
3. 是**离散项，**因为在真实世界中，正常情况下，正常的片段的数目要比异常片段的数目多，在本数据集中，异常片段在异常视频里也只有一部分，所以异常视频是**离散的，**所以这里添加离散项让所有片段得分之和变低，从而突出异常视频的得分。

平滑项和离散项（平滑约束和离散约束是论文的创新点）

模型流程总结：

模型的输入（batch =1）是一个异常的video和一个正常的video，将两个video分割成数目相同的片段。将这些片段送入3D CNN做特征提取，提取出4096维的特征，将所有的instance特征组合起来，送入3层全连接层，预测得分，得到每个片段的得分，最后在对这些得分计算损失，用反向传播算法进行训练。

实现细节：

超参数，学习率等详情看论文。

**总结**

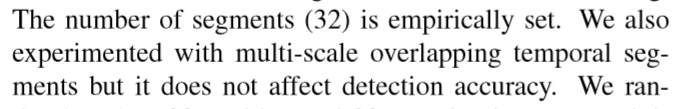
1 对于真实世界中的监控视频的异常学习，作者采用 多实例排序 （MIR）模型，很好的避免了视频片段标定难的问题。使得模型是一个弱监督模型。并且能够比较好的实现对真实世界的视频的异常检测。

2 本文除了提出这个模型之外，还提出了一个完备的数据集（很多方法在这个数据集上表现不好），本篇总结里没有总结数据集，但是这个数据集是一个有挑战性的数据集，同时这个数据集和现实生活更加接近（从市场，异常分布种类等等）

3 作者的工作主要是检测是否有异常，对于具体的异常，我们可以通过训练具体的识别器来进行下一步工作，所以作者的工作类似于 提取proposal的工作

**疑问**

1 作者将一个视频分割成多个片段，作者在论文里提出，分割片段的数目不会很大地影响检测精度，但是这个值得怀疑，需要用进一步的实验验证。



也就是说 片段的数目，以及片段是否重叠。同时每一个片段的时间差异相差太大（也就是两个时间长度相差比较大的视频，将视频分割成数目相同的片段进行下一步的工作）这些因素对实验结果对模型性能是否有影响

2 作者这里没有提到视频里是不是只有一个异常片段，从损失函数的结构来看，这个模型对于只有一个异常的视频效果应该较好。但是如果一个视频里有多个异常片段的话，这个模型的性能如何呢？当然这个问题可以通过将一段视频分割成多个视频，再针对这多个视频进行分割和下一步的工作来解决。但是是否有更好的解决方案？