# 监控视频内容分析

## 1.MOT多目标跟踪简介

MOT（Multi-Object Tracking），顾名思义，就是在一段视频中同时跟踪多个目标。在事先不知道目标数量的情况下，对视频中的行人、汽车、动物等多个目标进行检测并赋予ID进行轨迹跟踪。不同的目标拥有不同的ID，以便实现后续的轨迹预测、精准查找等工作。MOT主要应用场景是安防监控和自动驾驶等，这些场景中我们往往需要对众多目标同时进行追踪。这是仅用目标检测算法或单目标跟踪算法都无法做到的，人们就自然就开始了对多目标跟踪算法的。

在单目标跟踪中，我们往往会使用给定的初始框，在后续视频帧中对初始框内的物体进行位置预测。而多目标跟踪算法，大部分都是不考虑初始框的，原因就是上面的目标消失与产生问题。取而代之，在多目标跟踪领域常用的跟踪策略是TBD（Tracking-by-Detecton），又或者也可叫DBT（Detection-Based-Tracking）。即在每一帧进行目标检测，再利用目标检测的结果来进行目标跟踪，这一步我们一般称之为数据关联（Data Assoiation）。

多目标跟踪是在事先不知道目标数量的情况下，对视频中的行人、汽车、动物等多个目标进行检测并赋予ID进行轨迹跟踪。不同的目标拥有不同的ID，以便实现后续的轨迹预测、精准查找等工作。

## 2.MOT多目标跟踪任务难点

相对于单目标跟踪，多目标跟踪面临着更加复杂的问题包括频繁的遮挡、轨迹的管理、相似的外观和多目标间的相互影响。在实际的应用场景中，需要面对存在的各种复杂变化（以行人跟踪为例）。

### 2.1 目标自身的变化

目标的颜色变化（行人的衣服颜色变化），目标的尺度变化（离摄像头的远近）和目标的形态变化（行人的站立、蹲与躺）等。

### 2.2 外界环境的变化

光线明暗的变化、目标所处环境的多样性、目标的消失与出现和目标的遮挡问题。

这些复杂变化会影响跟踪对象与背景环境的区分度，从而进一步影响多目标跟踪算法的跟踪效果和结果的好坏，所以需要恰当地处理这些变化来提高多目标跟踪的准确性。

## 3.MOT与ReID的联系与区别

在MOT任务中，一般常用的特征模型有运动模型和表观模型，其中表观模型以行人重识别(ReID)类算法为主流。Re-ID任务主要解决的是跨摄像头场景下行人的识别与检索，其中存在给定了身份的图片序列query，需要为不同摄像头场景下的多组图片gallery的行人身份进行判定。

|  |  |  |
| --- | --- | --- |
| 指标 | ReID | MOT |
| 研究对象 | 行人/车辆 | 行人/车辆 |
| 场景 | 跨摄像头 | 单一摄像头 |
| 先验信息 | 表观信息 | 时空运动信息，表观信息 |
| 研究任务 | 行人匹配 | 数据关联 |
| 目标顺序是否严格对称 | 是 | 否 |

**目标序列严格对称：是指的待比对的两个视频序列中，目标的id是否一致，如果不一致，则表明两个视频序列各自可能存在新的目标。**

ReID更像是一个图像检索的问题，把检测出来的行人图片提取一个feature，根据feature来判断两张图片的相似性，实现检索的目的，而[tracking](https://www.zhihu.com/search?q=tracking&search_source=Entity&hybrid_search_source=Entity&hybrid_search_extra=%7B%22sourceType%22%3A%22answer%22%2C%22sourceId%22%3A265070848%7D)更多的像一个数据关联的问题，利用ReID特征、时空信息、运动信息等来把两个目标给匹配关联起来，跟踪方法其中有一类方法是tracking by detect（TBD），而ReID做tracking就属于这一类，主要思想就是先把行人目标检测出来，然后判断detect的bounding boxes是否属于同一个行人，把同一个行人的检测框给关联起来就是tracklet。

ReID的难点在于是否能够保证相同的人员在不同位置角度摄像机下能被识别出来。

Tracking 的问题是在于估计物体的state，它的难点在于如何基于detection的信息对物体的未来或者过去进行预测，来解决detection不可靠的问题。

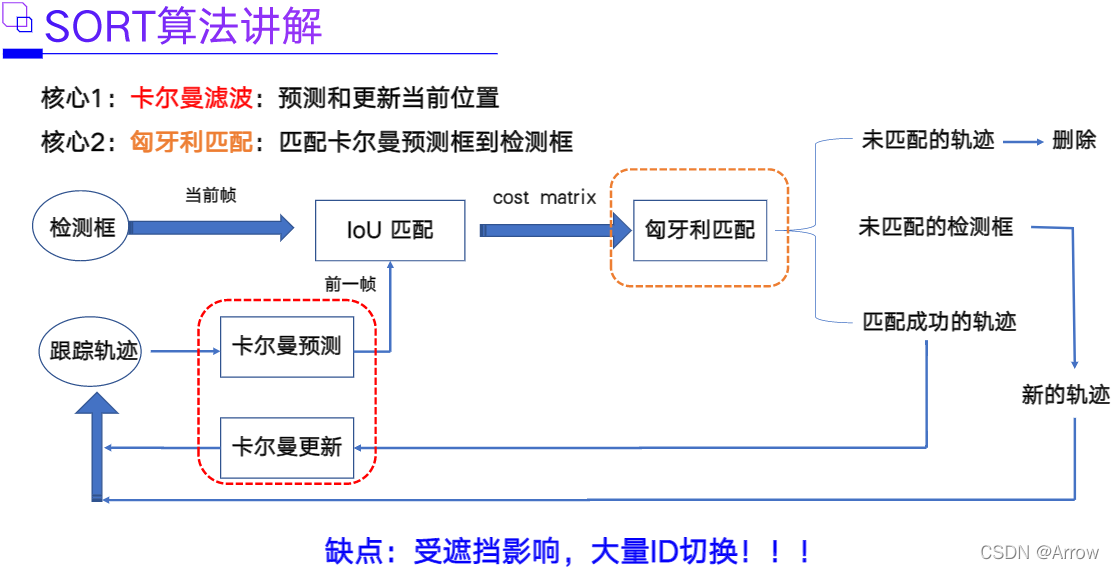
ReID可以不基于tracking来做，通过detection后就可以了，但是因为遮挡的问题，所以基于tracking的结果做re-ID可以提升短时遮挡问题的准确性。

Tracking 也可以不基于re-ID 来做，比RFS，即使同时考虑所有association，做monte carlo的概率分析，从而在不考虑ID的前提下预测目标数量和估计目标状态。也可以使用labelled RFS 从而预测ID，并把ID反馈给系统，来提升整体系统的正确率。

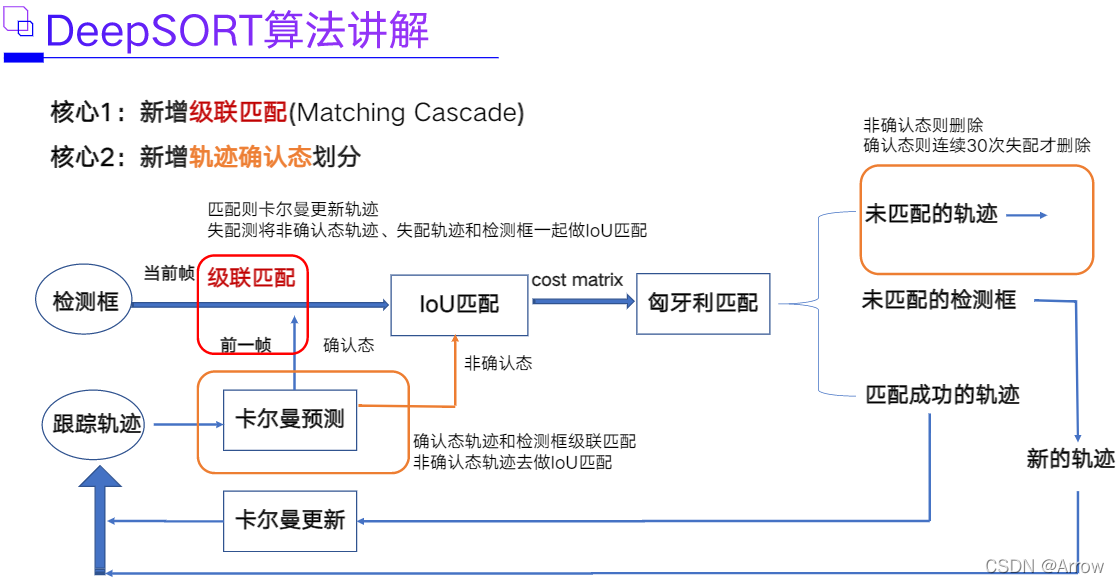


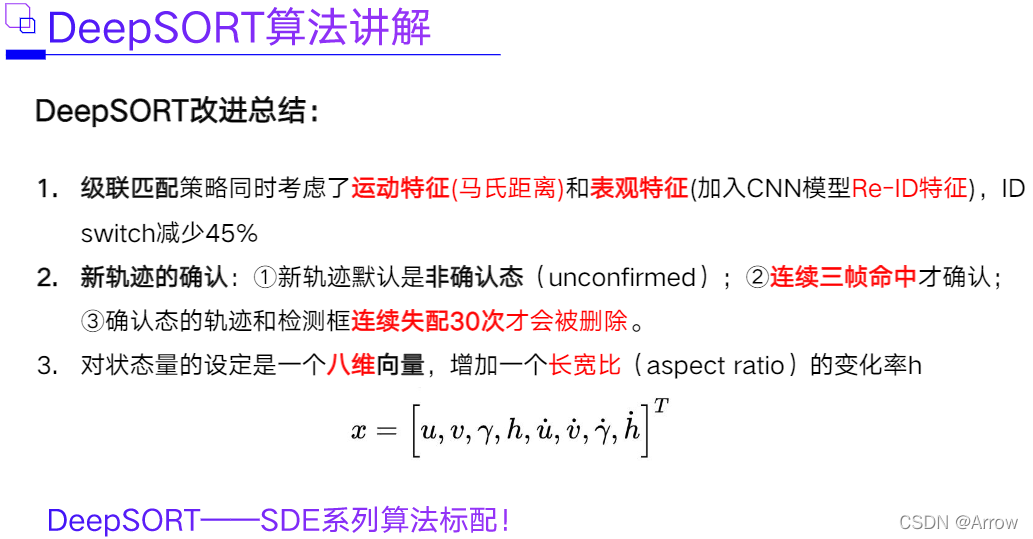
## 4.MOT主流算法

### 4.1.基于Tracking-by-detection的MOT

基于Tracking-by-detaction框架的MOT算法是先对视频序列的每一帧进行目标检测，根据包围框对目标进行裁剪，得到图像中的所有目标。然后，转化为前后两帧之间的目标关联问题，通过IoU、外观等构建相似度矩阵，并通过匈牙利算法、贪婪算法等方法进行求解。

首先检测目标，然后链接到轨迹中。这种策略也通常被称为“tracking-by-detection”。给定一个序列，在每帧中进行特定类型的目标检测或运动检测（基于背景建模，得到目标假设， 然后进行顺序或批量跟踪，将检测假设连接到轨迹中。有两个问题值得注意：第一，由于提前训练目标检测器，TBD大部分关注特定的目标类型，如行人、车辆或人脸。第二，TBD的性能非常依赖于所采用的目标检测器的性能代表方法：SORT、DeepSORT





### 4.2基于检测和跟踪联合的MOT

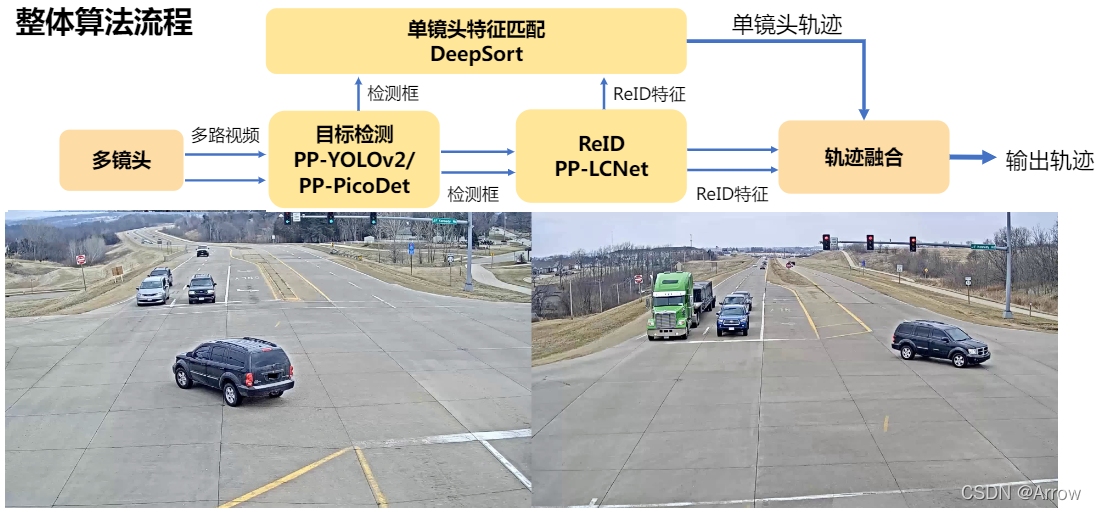
代表方法：JDE、FairMOT、CenterTrack、ChainedTracker

FairMO：它采用了一个非常简单的包含两个分别检测物体和提取ReID特征的均匀分支的网络框架。检测分支是以无锚的方式实现的，它估计物体的中心和大小，以位置感知图来表示。同样，re-ID分支为每个像素估计一个re-ID特征，来描述以该像素为中心的物体。注意到两个完全平等的分支和之前将检测和re-ID级联的方法完全不同。所以FairMOT消除了检测分支不公平的优势，有效地学习高质量的re-ID特征并且对于更好的MOT结果取得了检测和re-ID的平衡。

### **4.3**基于注意力机制的MOT

**代表方法：基于transformer的TransTrack,TrackFormor**

## 5.MTMC跨摄像头多目标跟踪



跨摄像头多目标追踪整体框架如图，主要分为目标检测，ReID,单镜头特征匹配，轨迹融合四部分。