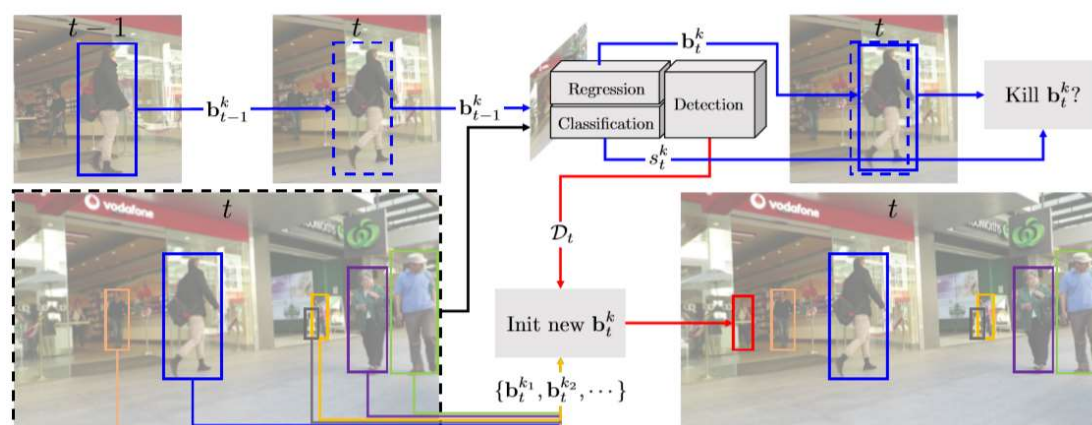


# CVPR2019 Tracktor: 《Tracking without bells and whistles》



Method	MOTA $\uparrow$	IDF1 $\uparrow$	MT $\uparrow$	ML $\downarrow$	FP $\downarrow$	FN $\downarrow$	ID Sw. $\downarrow$
D&T [18]	50.1	24.9	23.1	27.1	3561	52481	2715
Tracktor-no-FPN	57.4	58.7	30.2	22.5	2821	45042	1981
Tracktor	61.5	61.1	33.5	<b>20.7</b>	367	42903	1747
Tracktor+reID	61.5	62.8	33.5	<b>20.7</b>	367	42903	921
Tracktor+CMC	<b>61.9</b>	64.1	<b>35.3</b>	21.4	<b>323</b>	<b>42454</b>	458
Tracktor++ (reID + CMC)	<b>61.9</b>	<b>64.7</b>	<b>35.3</b>	21.4	<b>323</b>	<b>42454</b>	<b>326</b>

## 本文核心思想：

本文提出的 Tracktor 仅通过一个目标检测器即可完成 MOT 任务。对于给定的帧  $t$  包含两个主要的处理步骤，如上图，用蓝色和红色表示。

首先（蓝色部分，对应轨迹 bbox 回归更新），将  $t-1$  帧已有的轨迹边界框作为第  $t$  帧该轨迹的起始边界框，进行回归对齐即可得到当前帧对应轨迹的 bbox（这里的理论支持前提是高帧率视频下，前一帧和后一帧位置变化不大）。然后，将新 bbox 的位置相应对象分类分数用于 kill 掉可能被遮挡的轨道。

其次（红色部分，新轨迹的加入），对于新轨迹的出现，检测器仍然提供  $t$  帧的一组检测  $D_t$ ，如果其中某些检测结果没有与已存在的轨迹集合  $B_t$  的任何边界框的 IOU 超过一定阈值，则初始化这些检测为新轨迹。

以上是 Tracktor 的核心内容，Tracktor++ 就是加入 ReID 和运动模型。

**总结：**Tracktor 模型简单，通过用前帧目标的 bbox 作为当前帧的初始 bbox 来回归，大大减少了 FP，且省略了 data association 步骤，执行速度快，而且可以从目标检测领域的进步中受益，无需对跟踪进行专门优化训练，但无法解决遮挡问题引发的 IDSW，所以作者在 Tracktor++ 模型中加入 ReID，这样就解决了遮挡问题，但是由于计算可能带来的帧率下降，导致前后两帧结果差距较大，因此加入运动模型。

## 本文的贡献：

- 1) 介绍了 Tracktor，该 Tracktor 通过利用检测器的回归头来执行对象边界框的时间重新对齐，从而解决了多对象跟踪问题。
- 2) 介绍了对 Tracktor 的两个简单扩展，ReID Siamese 网络和运动模型。最终的跟踪器在三个具有挑战性的多对象跟踪基准测试中表现出最先进的性能。

- 3) 对失败案例和具有挑战性的跟踪方案进行了详细的分析,并显示没有任何一种专用的跟踪方法比本文的回归方法具有更好的性能。
- 4) 提出了一种作为检测器的新跟踪范式的方法,该方法可以利用检测器并使研究人员专注于剩余的复杂跟踪挑战。这包括对有前途的未来研究方向的广泛研究。

## 1. Background and Motivation

本文提出观点: **A detector is all you need for Multi-Object Tracking**。MOT 的常规思路都是要先检测,在做数据关联。但是,作者发现,最近两年,虽然有很多方法被提出,但是,在几个公共的 MOT 数据集上,并没有很明显的提升(两年才提升了 2 点多)。作者发现通过仅仅利用物体检测算法,如 Faster rcnn,就可以达到 SOTA 的效果。这也引出了一个很有意思的问题:如果一个检测器就可以很好地完成 MOT 的任务,那么,tracking algorithm 有什么用呢?这还是必要的吗?

## 2. A detector is all you need:

有一些物体检测算法中包含通过 regression 的方式进行 bounding box refinement 的模块。作者提出使用这种 regressor 来进行 MOT。这种方式有如下两种优势:

- 1). 不需要任何关于 tracking 的训练;
- 2). 在测试阶段不进行任何复杂的优化,因此该算法是 online 的。

此外,本文的方法也可以达到 SOTA 的效果。

### 2.1. Object detector

在 MOT17Det 行人检测数据集上使用 ResNet-101 和特征金字塔网络(FPN)训练了更快的 R-CNN。

### 2.2. Tracktor

MOT 的挑战在于:提取给定的视频帧中的多个物体的时间和空间上位置信息,即:轨迹。这种轨迹信息被定义为:一系列有序的物体包围盒的集合。

在时刻  $t=0$ , 作者的 tracker 用第一组检测的结果进行初始化,即:  $D_0 = d_0^1, d_0^2, \dots = B_0$ 。在图 1 中展示了两个随后的步骤: **the bounding box regression and track initialization**。

#### Bounding box regression:

第一步就是,如下图蓝色箭头所示,探索 bounding box regression 来更新已有的轨迹。通过将  $t-1$  帧的 bounding box  $b_{t-1}^k$  进行回归,得到第  $t$  帧新的位置  $b_t^k$ 。在 Faster RCNN 中,这就对应了在当前帧的 feature map 上进行 RoI Pooling 操作,但是用的是前一帧的 BBox。作者提出这种做法的一个假设就是:两帧之间的运动不是很明显,特别是在高帧率的视频上。身份自动从先前的边界框转移到回归边界框,从而有效地创建了轨迹。对于所有后续帧重复此操作。

在 BBox 回归以后,作者的跟踪器考虑两种情况来删除一个轨迹:

- 1). 一个物体在视频帧中消失了,或者被其他物体被遮挡了,判断依据即:如果新的 bbox 的位置 classification score 小于某一阈值;
- 2). 目标之间的遮挡,可以通过采用 NMS 来处理。

#### Bounding Box Initialization

为了处理新出现的目标，目标检测器也提供了整个视频帧的检测结果  $D_t$ 。第二步，即图中红色箭头部分，类似于第一帧的初始化。但是，从  $D_t$  开始的检测，当且仅当  $IoU$  与任何已有的轨迹  $b_t^k$  不超过某一阈值。即，将匹配不到的检测认为一个目标为新的 id。

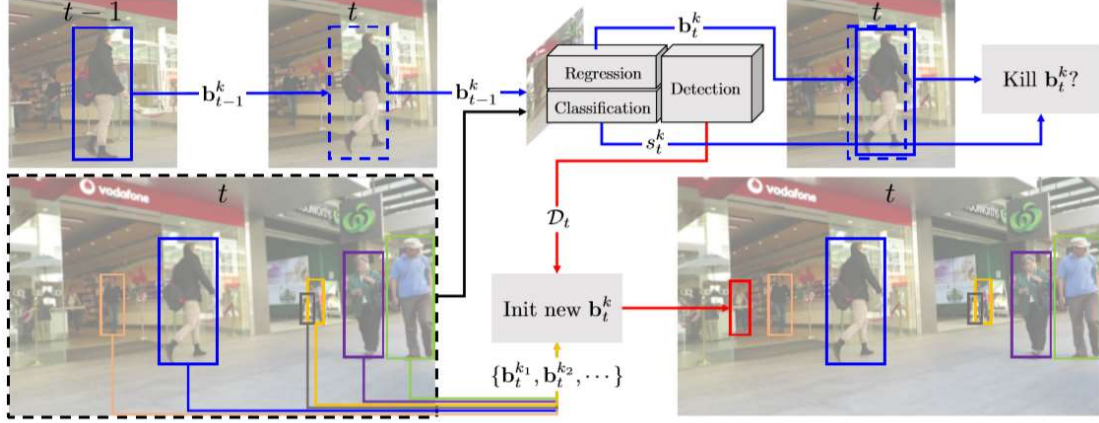


Figure 1: The presented Tracktor accomplishes multi-object tracking only with an object detector and consists of two primary processing steps, indicated in blue and red, for a given frame  $t$ . First, the regression of the object detector aligns already existing track bounding boxes  $b_{t-1}^k$  of frame  $t-1$  to the object's new position at frame  $t$ . The corresponding object classification scores  $s_t^k$  of the new bounding box positions are then used to kill potentially occluded tracks. Second, the object detector (or a given set of public detections) provides a set of detections  $D_t$  of frame  $t$ . Finally, a new track is initialized if a detection has no substantial Intersection over Union with any bounding box of the set of active tracks  $B_t = \{b_t^{k_1}, b_t^{k_2}, \dots\}$ .

### 2.3. Tracking extensions

作者将该模型进行了拓展，即：结合了 motion model 和 re-identification model。

**Motion model.** 作者之前的假设：两帧之间的变化不是很大，在有些情况下并不成立：

**large camera motion and low video frame rates.** 在极端的情况下，BBox 从 frame  $t-1$  在第  $t$  帧中可能根本不包含目标物体了。所以，作者设计了两种 motion model 来改善 BBox 在将来帧中的位置。**对于运动相机**，作者采用相机运动补偿（camera motion compensation, CMC）的方式进行缓解。作者采用图像配准的方式来对齐视频帧，用的是 增强型相关系数 (ECC) 最大化。**对于低帧率的视频**，作者采用等速假设(CVA)。

**Re-identification.** 为了让 tracker 能够保持 online，作者提出利用 short-term re-ID 的方式（借助 Siamese Network 来进行 appearance feature 的匹配）来改善效果。为了达到这个目标，作者将删除了的目标，存储固定帧数的样本。然后将这些样本和新检测的目标在 embedding space 进行重识别。为了最大程度地降低错误 reID 的风险，我们仅考虑  $IoU$  足够大的成对的停用边界框和新边界框。

	Method	MOTA ↑	IDF1 ↑	MT ↑	ML ↓	FP ↓	FN ↓	ID Sw. ↓
MOT17	Tracktor++	<b>53.5</b>	52.3	19.5	36.6	<b>12201</b>	248047	2072
	eHAF [58]	51.8	<b>54.7</b>	<b>23.4</b>	37.9	33212	<b>236772</b>	1834
	FWT [23]	51.3	47.6	21.4	35.2	24101	247921	2648
	jCC [30]	51.2	54.5	20.9	37.0	25937	247822	<b>1802</b>
	MOTDT17 [9]	50.9	52.7	17.5	35.7	24069	250768	2474
	MHT-DAM [32]	50.7	47.2	20.8	36.9	22875	252889	2314
MOT16	Tracktor++	<b>54.4</b>	<b>52.5</b>	19.0	<b>36.9</b>	<b>3280</b>	<b>79149</b>	682
	HCC [44]	49.3	50.7	17.8	39.9	5333	86795	<b>391</b>
	LMP [59]	48.8	51.3	18.2	40.1	6654	86245	481
	GCRA [43]	48.2	48.6	12.9	41.1	5104	88586	821
	FWT [23]	47.8	44.3	<b>19.1</b>	38.2	8886	85487	852
	MOTDT [9]	47.6	50.9	15.2	38.3	9253	85431	792
2D MOT-2015	Tracktor++	<b>44.1</b>	46.7	18.0	<b>26.2</b>	6477	<b>26577</b>	1318
	AP-IHWDPL-p [8]	38.5	<b>47.1</b>	8.7	37.4	<b>4005</b>	33203	586
	AMIR15 [56]	37.6	46.0	15.8	26.8	7933	29397	1026
	JointMC [30]	35.6	45.1	<b>23.2</b>	39.3	10580	28508	457
	RAR15pub [17]	35.1	45.4	13.0	42.3	6771	32717	<b>381</b>

Table 2: We compare our online multi-object tracker Tracktor++ with other modern tracking methods. As a result, we achieve a new state-of-the-art in terms of MOTA for public detections on all three MOTChallenge benchmarks. The arrows indicate low or high optimal metric values.

实验部分机翻：

### 3. Experiments

我们在针对行人跟踪的多个数据集上演示了我们提出的 Tracktor 跟踪器及其扩展 Tracktor ++ 的跟踪性能。此外，我们对上述扩展进行了消融研究，进一步表明，我们的跟踪器在跟踪准确性方面优于最新方法，并且在身份保存方面表现出色。

**MOTChallenge.** 多目标跟踪基准 MOTChallenge 由几个具有挑战性的行人跟踪序列组成，具有频繁的遮挡和拥挤的场景。序列的视角，物体大小，相机运动和帧频各不相同。挑战包含三个单独的跟踪基准，即 2D MOT 2015 [37]，MOT16 和 MOT17 [45]。MOT17 测试套件包括总共 7 个序列，每个序列都提供了三组公共检测。这些检测源自不同的对象检测器，每个检测器的性能均得到提高，即 DPM [19]，Faster R-CNN [52] 和 SDP [63]。我们的物体检测器在 MOT17Det [45] 检测基准上进行了训练，该基准包含与 MOT17 相同的图像。MOT16 基准还包含与 MOT17 相同的序列，但仅提供 DPM 公共检测。2D MOT 2015 基准提供了 11 个序列的 ACF [14] 检测。跟踪问题的复杂性需要几个指标来衡量跟踪器性能的不同方面。多目标跟踪准确性 (MOTA) [29] 和 ID F1 分数 (IDF1) [53] 量化了两个主要方面，即对象覆盖率和身份。

**Public detections.** 为了与其他跟踪方法进行公平比较，我们使用 MOTChallenge 提供的公开检测结果执行了所有实验。也就是说，本文中比较的所有方法，包括我们的方法及其扩展，都通过帧检测来处理相同的预先计算的帧。对于我们的方法，仅从公共检测边界框初始化新轨迹，即，我们从不使用对象检测器来检测新边界框。我们仅应用边界框回归器和分类器分别获得新的  $b_t^k$  和  $s_t^k$ 。MOTChallenge 公开基准包含多种方法 [30、9、13]，这些方法通过训练有素的神经网络将给定的检测分类，因此，我们将给定的检测处理也视为公开的。

#### 3.1. Ablation study

Method	MOTA ↑	IDF1 ↑	MT ↑	ML ↓	FP ↓	FN ↓	ID Sw. ↓
D&T [18]	50.1	24.9	23.1	27.1	3561	52481	2715
Tracktor-no-FPN	57.4	58.7	30.2	22.5	2821	45042	1981
Tracktor	61.5	61.1	33.5	20.7	367	42903	1747
Tracktor+reID	61.5	62.8	33.5	20.7	367	42903	921
Tracktor+CMC	61.9	64.1	35.3	21.4	323	42454	458
Tracktor++ (reID + CMC)	61.9	64.7	35.3	21.4	323	42454	326

Table 1: This ablation study illustrates multiple aspects on the performance of our Tracktor. In particular the improvements from extending it with tracking specific methods, i.e., a short-term bounding box re-identification and camera motion compensation by frame alignment. The combination yields the Tracktor++ tracker. We evaluated only on the Faster R-CNN set of MOT17 public detections. The arrows indicate low or high optimal metric values.

表 1 中对 MOT17 [45] 训练集的消融研究旨在显示三个方面：(i) 我们的方法在应用检测器进行跟踪时的优势；(ii) 改进的物体检测方法的潜力；以及 (iii) 通过使用跟踪方法（即重新识别 (reID) 和摄像机运动补偿 (CMC)）扩展我们的香草 Tracktor 进行了改进。应该指出的是，尽管 MOT17Det 和 MOT17 包含相同的图像，但我们并未对训练集进行交叉验证，因为我们的香草跟踪器从未接受过跟踪 GT 数据的训练。视频对象检测器和跟踪器 D&T [18] 在跟踪 GT 数据时训练检测器，该数据会生成两帧的跟踪小波。但是，尽管随后进行了动态编程，但是基于检测器的跟踪器却不如我们基于在线回归的多帧跟踪生成器。此外，我们通过显示 Tracktor-no-FPN 的跟踪性能（即我们的方法和一种不带特征金字塔网络 (FPN) 的 Faster R-CNN）来证明我们框架在改进检测方法方面的潜力 [41]。尽管我们对 Tracktor ++ 的扩展具有简单性，但它们的贡献对于身份开关的大幅减少和 IDF1 度量的增加意义重大。在下一部分中，我们将证明该效果可以成功地与测试集上的其他最新方法进行比较。



### 3.2. Benchmark evaluation

	Method	MOTA $\uparrow$	IDF1 $\uparrow$	MT $\uparrow$	ML $\downarrow$	FP $\downarrow$	FN $\downarrow$	ID Sw. $\downarrow$
MOT17	Tracktor++	<b>53.5</b>	52.3	19.5	36.6	<b>12201</b>	248047	2072
	eHAF [38]	51.8	<b>54.7</b>	<b>23.4</b>	37.9	33212	<b>236772</b>	1834
	FWT [23]	51.3	47.6	21.4	35.2	24101	247921	2648
	jCC [30]	51.2	54.5	20.9	37.0	25937	247822	<b>1802</b>
	MOTDT17 [9]	50.9	52.7	17.5	35.7	24069	250768	2474
	MHT-DAM [32]	50.7	47.2	20.8	36.9	22875	252889	2314
MOT16	Tracktor++	<b>54.4</b>	<b>52.5</b>	19.0	<b>36.9</b>	<b>3280</b>	<b>79149</b>	682
	HCC [44]	49.3	50.7	17.8	39.9	5333	86795	<b>391</b>
	LMP [59]	48.8	51.3	18.2	40.1	6654	86245	481
	GCRA [43]	48.2	48.6	12.9	41.1	5104	88586	821
	FWT [23]	47.8	44.3	<b>19.1</b>	38.2	8886	85487	852
	MOTDT [9]	47.6	50.9	15.2	38.3	9253	85431	792
2D MOT 2015	Tracktor++	<b>44.1</b>	46.7	18.0	<b>26.2</b>	6477	<b>26577</b>	1318
	AP-HWDPL-p [8]	38.5	<b>47.1</b>	8.7	37.4	<b>4005</b>	33203	586
	AMIR15 [36]	37.6	46.0	15.8	26.8	7933	29397	1026
	JointMC [30]	35.6	45.1	<b>23.2</b>	39.3	10580	28508	457
	RAR15pub [17]	35.1	45.4	13.0	42.3	6771	32717	<b>381</b>

Table 2: We compare our online multi-object tracker Tracktor++ with other modern tracking methods. As a result, we achieve a new state-of-the-art in terms of MOTA for public detections on all three MOTChallenge benchmarks. The arrows indicate low or high optimal metric values.

我们在各自基准测试集上评估 Tracktor++ 的性能，而无需对训练集进行任何训练或优化。表 2 列出了所有序列以及 MOT17 在所有三组公共检测中积累的总体结果。为了进行比较，我们只考虑在 MOTChallenge 基准中正式发布和同行评审的条目。我们的补充材料详细列出了各个序列的所有结果。对于所有序列，都使用摄像机运动补偿 (CMC) 和 reID。唯一的低帧速率序列是 2D MOT 2015 AVGTownCentre，为此我们应用了上述恒速假设 (CVA)。对于最初来自 KITTI [20] 基准的两个自动驾驶序列，我们应用了旋转以及平移相机运动补偿。注意，对于所有基准测试，我们使用相同的 Tracktor++ 跟踪器，该跟踪器经过 MOT17Det 对象检测训练。正如我们所展示的，它能够在 MOTA 方面实现所有三个挑战的最新技术。

特别是我们的 MOT16 结果表明我们的跟踪器能够应对性能相对较差的检测。由于跟踪器的性质和逐帧边界框回归的鲁棒性，我们以较大的余量，特别是互为假名 (FN) 和身份保留 (IDF1) 优于所有其他跟踪器 MOT16。值得注意的是，即使场景的特点与 MOT17 有所不同，我们也提供了有关 2D MOT 2015 的最新技术。我们不使用 MOT15 训练序列，这进一步说明了我们的跟踪器的通用性。

## 4. Analysis

我们的跟踪器的卓越性能，无需任何跟踪专门培训或优化，就需要进行更彻底的分析。如果没有复杂的跟踪方法，就不会在拥挤和被遮挡的地方表现出色，而只能在仁慈的跟踪场景中表现出色。这就引出了一个问题，即更常见的跟踪方法是否也不能专门解决这些复杂的情况。我们的实验和随后的分析应证明我们在轻松跟踪场景中的方法的优势，并激发未来的研究重点关注剩余的复杂跟踪问题。特别是，我们对通过检测进行跟踪的通用执行提出了质疑，并提出了新的跟踪范式。随后对 MOT17 训练数据进行了分析，我们将所有表现最佳的方法与公开共享的数据进行了比较。

### 4.1. Tracking challenges

Method	Online	Graph	reID	Appearance model	Motion model	Other
Tracktor	×					
Tracktor++	×		×		Camera	
FWT [23]		Dense				Face detection
jCC [30]		Dense				Point trajectories
MOTDT17 [9]	×		×	×	Kalman	
MHT-DAM [32]		Sparse		×	Kalman	

Table 3: A summary of the fundamental characteristics of our methods and other state-of-the-art trackers.

为了更好地了解我们的跟踪器，我们想分析具有挑战性的跟踪方案，并将其优缺点与其他跟踪器进行比较。为此，我们在表 3 中总结了它们的基本特性。FWT [23] 和 jCC

[30]均在所有检测中均应用了图谱优化来弥补所有检测结果。相反, MHT\_DAM [32]将其优化限制在假设轨迹的稀疏前视图。

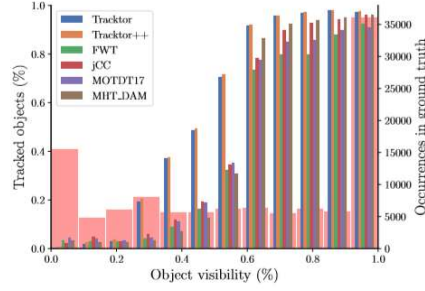


Figure 2: We illustrate the ratio of tracked objects with respect to their visibility evaluated on the Faster R-CNN public detections. The results clearly demonstrate that none of the presented more sophisticated methods achieves superior performance to our approach. This is especially noticeable for highly occluded boxes. The transparent red bars indicate the ground truth distribution of visibilities.

**Object visibility.** 直观地, 我们期望对象-对象或对象-非对象遮挡的跟踪性能会下降, 即可见性下降的目标的跟踪性能会下降。在图 2 中, 我们比较了成功跟踪的边界框与其可见性的比率。透明的红色条表示每种可见性的 gt 边界框的出现, 并说明了对跟踪器总体性能的成比例影响。即使对于可见性低至 0.3 的部分封闭边界框, 我们的方法也能实现出色的性能。MHT DAM 和 MOTDT17 的识别保留方面[9]或 MHT DAM 和 jCC 的精细插值功能似乎都无法成功解决高度遮挡的物体。表 2 中较高的 MOTA 值很大程度上是由于 gt 可见性的不平衡分布所致。不出所料, 我们的扩展版本仅对我们的香草 Tracktor 进行了较小的改进。

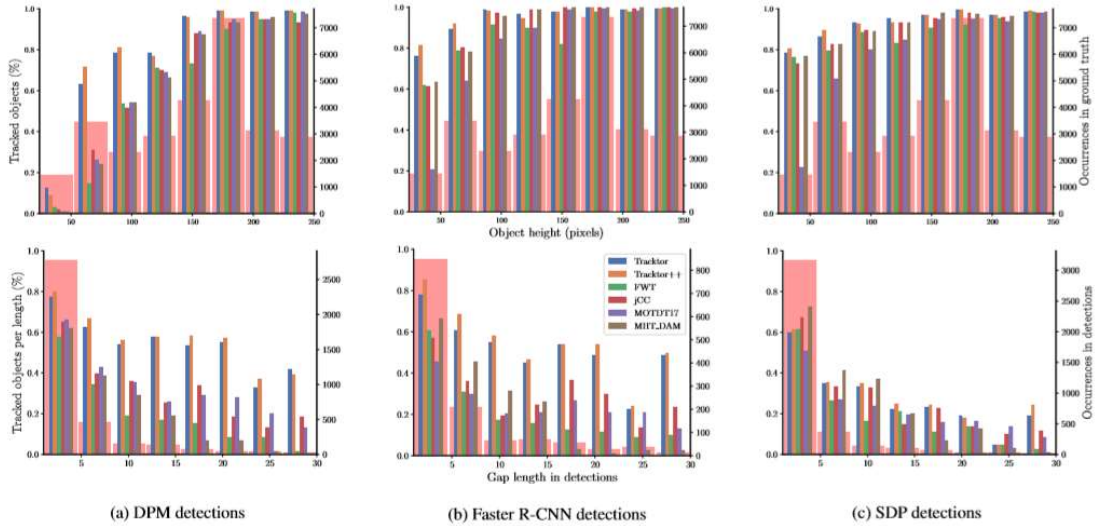


Figure 3: The two rows illustrate the ratio of tracked objects with respect to: (i) object heights and (ii) the length of gaps in the provided public detections. The transparent red bars indicate the ground truth distribution of heights and gap lengths in the detections, respectively. To demonstrate the shortcomings of the presented trackers we limited the height comparison to objects with visibility greater or equal than 0.9. Tracks that are not detected at all are not considered as a gap. Hence, SPD generates the most gaps. For it also provides the most detections.

**Object size.** 考虑到图 2 中大部分可见但未跟踪的对象, 我们认为对象的可跟踪性不仅取决于其可见性, 还取决于其大小。因此, 我们对可见性进行相同的比较, 但是对对象的大小进行比较。在图 3 的第一行中, 我们假设行人的高度与其大小和所有三个 MOT17 公共检测集的大小成正比。对于大于 250 像素的物体高度, 所有方法的效果均相似。为了展示它们甚至对于高度可见的对象的缺点, 我们仅比较可见度大于 0.9 的对

象。正如预期的那样，在所有三个检测集中，对象的可跟踪性随其大小而急剧下降。我们的跟踪器显示出在补偿所有对象尺寸的 DPM 和 Faster R-CNN 检测不足方面的优势。除 MOTDT17 以外的所有方法均受益于 SDP 提供的其他小检测。对于我们的跟踪器来说，这在很大程度上归功于我们的 **Faster-RCNN 检测器的功能金字塔网络扩展**。但是，在线 MOTDT17 方法的学习的外观模型和 reID 似乎通常容易受到小检测。外观模型通常具有较小的对象尺寸和很少观察到的像素。总之，除了我们对劣质检测的补偿外，对于不同的物体尺寸，追踪器都没有表现出更好的性能。

**Robustness to detections.** 跟踪检测方法相对于可见性和大小的性能本质上受基础检测方法的鲁棒性限制。但是，正如观察到的物体大小一样，跟踪器在应对或受益于变化的检测质量方面的能力也有所不同。在图 3 的第二行中，我们根据跟踪器覆盖它们的检测差距来量化此功能。我们将检测间隙定义为至少一次被检测到的地面真实轨迹的一部分，并比较每个间隙的覆盖范围与间隙长度。凭直觉，较长的间隙更难弥补，因为在线或精细跟踪器必须分别执行更长的幻觉或插值。我们用透明红色指示了在相应的检测组上出现的缺口长度。对于 DPM 和 Faster R-CNN 检测，有两种解决方案可以显着地覆盖间隙：(i) 诸如 jCC 的精细内插，或者 (ii) 像 MOTDT 中那样使用 Kalman 滤波器和 reID 进行运动预测。与基于图的 jCC 方法相比，在线 MOTDT17 方法在覆盖特别长的间隙方面表现出色。但是，这些专用的跟踪方法都无法产生与逐帧回归跟踪器相似的鲁棒性，从而可以实现更好的覆盖范围。对于超过 15 帧的长检测间隙尤其如此。改进的 SDP 检测最能使大多数方法受益，而且我们和 MOTDT17 跟踪仪都没有以明显的间隙长度鲁棒性说服。

**Identity preservation.** 表 2 中汇总的 Tracktor ++ 结果表明，在 IDF1 和身份切换方面，身份保存性能可与专用跟踪方法相比。这无需在 jCC [30] 或 HAF [58] 中进行任何精细的图形优化即可实现。特别是，采用复杂外观模型和 reID 的 MOTDT17 并没有明显优于我们的回归跟踪器及其相对简单的扩展。但是，我们的方法在减少 MOT17 以及 MOT16 中的假阳性数方面表现出色。此外，我们还显示了合并其他身份保留扩展的跟踪能力。

## 4.2. Oracle trackers

Method	MOTA ↑	IDF1 ↑	FP ↓	FN ↓	ID Sw. ↓
Tracktor	61.5	61.1	367	42903	1747
Tracktor++	+0.4	+3.6	-44	-449	-1421
Oracle-Kill	+0.7	-0.7	-178	-694	+129
Oracle-REG	+1.4	+5.6	-218	-1401	-1463
Oracle-MM	+0.9	+5.2	-168	-898	-1332
Oracle-reID	0.0	+10.0	0	0	-1094
Oracle-MM-reID	+0.9	+13.9	-168	-898	-1706
Oracle-MM-reID-INTER	+2.6	+15.9	+3774	-6769	-1680
Oracle-ALL	+10.7	+22.5	-360	-11745	-1743

Table 4: To show the potential of Tracktor and indicate promising future research directions, we present multiple oracle trackers. Each oracle exploits ground truth data for a specific task, simulating, e.g., a perfect re-identification (reID) or motion model (MM). We evaluate only on the Faster R-CNN set of MOT17 public detections and highlight performance gains and losses with respect to the vanilla Tracktor in green and red, respectively. The arrows indicate low or high optimal metric values.

我们已经表明，没有一种专用的跟踪方法专门针对具有挑战性的跟踪方案，即重度遮挡下的物体或小的物体。因此，我们希望激励 Tracktor 作为一种新的跟踪范例。为此，我们对性能进行了双重分析：(i) 对象检测器对终止策略和边界框回归的影响；(ii) 确定可能扩展到 Tracktor 的性能上限。在表 4 中，通过用 gt 信息替换我们的算法部分，

我们展示了几种 oracle 跟踪器。除非另有说明，否则所有其他跟踪方面均由我们的香草跟踪器处理。他们的分析应该为研究人员提供有关 Tracktor 最有前途的研究方向和扩展范围的有用见解。

**Detector oracles.** 为了模拟可能完美的对象检测器，我们引入了两个预言机：

- Oracle-Kill：我们不是使用 NMS 或分类评分来杀死，而是使用 gt 信息。
- Oracle-REG：我们将边界框放置在其 gt 位置上，而不是进行回归。

两种预言在 MOTA 和 FP 方面均取得了实质性的改进。然而，由于回归与其他看不见的边界框作斗争，因此用 gt 而不是分数杀死会破坏身份保存。

**Extension oracles.** 应该注意的是，具有非完美扩展的 Tracktor ++ 已经弥补了检测器的某些不足。reID 和运动模型（MM）预言机模拟了潜在的附加性能提升。为了保持在线状态，这些排除了任何形式的事后跟踪间隙插入。

- Oracle-MM：运动模型将每个边界框放在下一帧中 gt 的中心。
- Oracle-reID：使用基础真实身份执行重新识别。

不出所料，这两个预言都大大改进了 IDF1 和身份切换。组合的 Oracle-MM-reID 代表 Tracktor ++ 的扩展上限。

**Omniscient oracle.** Oracle-ALL 执行地面真相查杀，回归和 reID。我们认为它的最高 MOTA 为 72.2%，结合较高的 IDF1 和几乎没有假阳性，是具有更快 R-CNN 和 FPN 对象检测器的 Tracktor 的绝对上限。

运动模型，尤其是，Oracle-MM-reIDINTER 提出了一种预测运动模型，该模型通过长时间遮挡使物体的位置产生幻觉。这种运动模型避免了由于后期检测间隙和相机移动而导致的错误后处理以及来自错误线性遮挡路径的其他误报。

#### 4.3. Towards a new tracking paradigm

总结我们的分析，我们就如何利用 Tracktor 作为未来研究方向的起点提出了两种方法：

**Tracktor with extensions.** 将 Tracktor 应用于给定的一组检测，并使用跟踪特定方法对其进行扩展。框到框的边界框回归将涵盖具有大型且高度可见的对象的场景。在这方面，考虑到对象的单个运动，它似乎最有可能实现一种透明的运动模型。另外，这种运动预测器减少了高级杀人策略的必要性。

**Tracklet generation.** 类似于通过检测进行跟踪，我们提出了一种通过小跟踪进行跟踪的方法。确实，许多算法已经将小轨迹用作输入[24、65]，因为它们在计算运动或外观模型方面的信息更为丰富。但是，通常使用特定的跟踪方法来创建这些小轨迹。我们提倡利用检测器本身，不仅要创建稀疏检测，而且要逐帧创建小轨迹。其余复杂的跟踪案例应通过后续的跟踪方法来解决。

在这项工作中，我们正式定义了那些困难的案例，分析了不仅我们的方法而且其他专用跟踪解决方案均失败的情况。通过这样做，我们质疑了当前多目标跟踪研究的重点，特别是与挑战性跟踪方案之间缺少的对抗。

## 5. Conclusions

我们已经证明，训练有素的 Faster-RCNN 检测器的边界框回归器足以解决当前基准中存在的大多数跟踪情况。转换为 Tracktor 的探测器无需进行跟踪地面真实数据的专门培训，就可以在线工作。此外，我们还展示了我们的轨迹可以通过重新识别和摄像机运动补偿来扩展，从而为 MOT 挑战提供了实质性的最新技术。我们分析了多种专用跟踪方法在挑战性跟踪场景中的性能，与基于回归的 Tracktor 相比，没有一种方法产生了更好的性能。我们希望这项工作能够建立新的跟踪范例，并利用对象检测器的全



部功能。