arXiv:2301.00493v1 [cs.CV] 2 Jan 2023

Argoverse 2：面向自动驾驶感知与预测的新一代数据集

Benjamin Wilson[*†, 1], **William Qi**[*†], **Tanmay Agarwal**[*†], **John Lambert**[†], **Jagjeet Singh**[†],
Siddhesh Khandelwal[2], **Bowen Pan**[†, 3], **Ratnesh Kumar**[†], **Andrew Hartnett**[†],
JhonyKaesemodel Pontes[†], **Deva Ramanan**[†, 4], **Peter Carr**[†], **James Hays**[†, 1]

[1]佐治亚理工学院， [2]不列颠哥伦比亚大学， [3]麻省理工学院， [4]CMU

摘要

我们介绍 Argoverse 2（AV2）——由三个数据集组成的数据集集合，旨在支持自动驾驶领域的感知与预测研究。带标注的传感器数据集 包含 1,000 个多模态数据序列，其中包括来自七个环视摄像头和两个立体摄像头的高分辨率图像，以及激光雷达点云和与地图对齐的六自由度位姿。这些序列包含 26 个物体类别的三维长方体标注，且所有类别的采样都足够充分，可用于训练和评估三维感知模型。激光雷达数据集 包含 20,000 个未标注的激光雷达点云序列及与地图对齐的位姿。该数据集是迄今规模最大的激光雷达传感器数据集合，可支持自监督学习以及新兴的点云预测任务。最后，运动预测数据集包含 250,000 个场景，这些场景经过筛选，涵盖了自动驾驶车辆与每个局部场景中的其他参与者之间有趣且具有挑战性的交互。模型需要预测每个场景中“评分对象”的未来运动，并获得能够记录物体位置、航向、速度和类别的轨迹历史。在这三个数据集中，每个场景都有各自的高精地图，其中包含三维车道和人行横道几何信息，数据来源于六座不同城市采集的数据。我们相信，这些数据集将以现有数据集无法实现的方式，支持新的和已有的机器学习研究问题。所有数据集均依据知识共享署名-非商业性使用-相同方式共享 4.0 许可协议发布。

1 引言

为了实现安全、可靠的自动驾驶目标，必须解决一系列机器学习任务，从立体深度估计、运动预测到三维目标检测。近年来，已有大量高质量的自动驾驶数据集发布，用于支持对这些以及其他重要机器学习任务的研究。许多数据集都是经过标注的“传感器”数据集 [4, 45, 39, 40, 24, 33, 18, 14, 41, 36]，其理念源自具有重要影响力的 KITTI 数据集 [17]。Argoverse 三维跟踪数据集 [6] 是首个包含“高精地图”的此类数据集——即包含车道级几何信息的地图。同样具有重要影响力的还有自动驾驶“运动预测”数据集 [12, 22, 34, 4, 52] ——其中包含抽象化的目标轨迹，而非原始传感器数据；Argoverse 运动预测数据集 [6]便是其中首个。

在过去两年里，Argoverse 团队举办了六项竞赛，涵盖三维跟踪、立体深度估计和运动预测。我们为这些任务维护评估服务器和排行榜，

*贡献相同。
†在 Argo AI 任职期间完成的工作。

以及三维检测。这些排行榜累计收录了来自四百个团队的数千项提交[1]。我们还维护 Argoverse API，并处理了超过一百个问题[2]。通过这些经验，我们总结出以下指导原则，以指导新一代 Argoverse 数据集的创建。

1. **越大并不总是越好**。自动驾驶车辆会采集海量传感器数据，在实际操作中难以处理。即使经过压缩，传感器数据集的规模也有数太字节。如果标准基准进一步扩大，我们就有可能疏远学术界的大部分群体，让进展落在资源雄厚的行业团体手中。因此，我们的数据规模与 *nuScenes* 的传感器数据相当，但不会超过它 *[4]*，也不会超过 *Waymo Open [45]*。
2. **让每个实例都发挥作用**。驾驶过程中的大部分时间都很枯燥。数据集应聚焦于当前预测和感知系统难以应对的复杂、有趣场景。因此，我们会挖掘特别拥挤、动态且运动学异常的场景。

3. **多样性至关重要**。仅使用冬季底特律的数据进行训练，不足以检测迈阿密的物体——迈阿密的摩托车和轻便摩托车出现频率是底特律的 15 倍。行为也各不相同，因此学到的行人运动行为可能无法泛化。因此，我们的每个数据集都取自六座各具特色的城市——奥斯汀、底特律、迈阿密、帕洛阿尔托、匹兹堡和华盛顿特区——也涵盖从多雪到晴朗的不同季节。
4. **绘制世界地图**。 高精地图是感知和预测的重要先验。通过巧妙编码地图信息的学习方法，在 Argoverse 竞赛中表现出色。因此，我们扩展了高精地图表示，加入三维车道几何、道路标线、人行横道、更高分辨率的地面高度等内容。

5. **进行自监督学习**。 近年来，其他机器学习领域也因自监督学习而取得了巨大成功。来自动态场景的大规模激光雷达数据与高精地图配对，有望带来优于当前监督式方法的表示。因此，我们构建了规模最大的激光雷达传感器数据集。

6. **应对长尾问题**。 乘用车十分常见，因此我们可以评估汽车的预测和检测准确率。然而，利用现有数据集，我们无法评估公交车和摩托车这类具有独特行为的目标的预测准确率，也无法评估婴儿车和轮椅的检测效果。因此，我们推出了迄今为止面向传感器和预测数据集的最大分类体系，并确保有足够的稀有目标样本用于训练和评估模型。

秉持这些指导原则，我们构建了 Argoverse 2（AV2）的三个数据集。下文将重点介绍其中的一些贡献。

1. 这个包含 1,000 个场景的传感器数据集 拥有迄今最大的自动驾驶分类体系，共 30 个类别。其中 26 个类别各包含至少 6,000 个立方体标注框，可支持多样化的分类体系训练与测试。与近期的自动驾驶数据集不同，该数据集还包含立体图像。2. 这个包含 20,000 个场景的激光雷达数据集 是目前最大的激光雷达自监督学习数据集。唯一类似的数据集 ONCE [36],由其他团队同期开发，但不包含高精地图。

3. 这个包含 250,000 个场景的运动预测数据集 拥有最全面的分类体系——涵盖 5 种动态参与者和 5 种静态参与者——并且覆盖了同类数据集中最大的制图区域。

我们相信，这些数据集将助力研究诸多问题，例如三维检测、三维跟踪、单目与立体深度估计、运动预测、视觉里程计、姿态估计、车道线检测、地图自动化、自监督学习、运动恢复结构、场景流、光流、接触时间估计以及点云预测。

2 相关工作

过去几年，在众多高质量数据集的推动下，自动驾驶感知和预测研究取得了快速进展。

传感器数据集与三维目标检测与跟踪。 新的三维目标检测传感器数据集 [4, 45, 39, 40, 24, 33, 18, 14, 41, 36] 催生了许多具有影响力的检测方法，例如

[1]此统计包含未发布到公开排行榜的私人提交。[2] https://github.com/argoverse/argoverse-api

基于锚框的方法，如 PointPillars [27], ，以及较新的无锚框方法，如 AFDet [16] 和 CenterPoint [51]。这些方法显著提升了所有数据集上的准确率。反过来，这些改进也使得分离特定目标点云成为可能，这已被证明对离板检测与跟踪 [42],以及仿真 [8],极具价值；此前，这些任务都需要人工标注的三维边界框 [35]。新的方法还探索了其他点云表示形式，例如距离图像 [5, 2, 46]。流式感知 [29, 21]提出了一种新范式，用于探索准确率与延迟之间的权衡。表 1 对 AV2 传感器数据集 与近期三维目标检测数据集进行了详细比较。

运动预测。 在运动预测领域，进展同样显著。转向基于注意力的方法 [28, 38, 37] 催生了多种用于地图和轨迹数据的新型基于向量的表示 [15, 31]。新数据集也为新算法铺平了道路：nuScenes [4], Lyft L5 [22], 和 Waymo 开放运动数据集 [12] 都在证明车道图对 Argoverse 至关重要后发布了车道图 1 [6]。Lyft 还引入了交通/速度控制数据，而 Waymo 则在地图中加入了人行横道多边形、车道边界（含标线类型）、限速和停止标志。近期，YANDEX 发布了 Shifts [34] 数据集，这是迄今为止规模最大的预测数据集合（按场景小时数计算）。这些数据集共同推动了对多主体长时域运动预测的探索，并利用了静态地图和动态地图。

继 Argoverse 1.1 取得成功后，我们将 AV2 定位为一个大规模代码仓库，收录高质量的运动预测场景，并保证数据频率（恰好为 10 赫兹）和多样性（覆盖 6 座城市、包含 >2000 公里的独特道路）。这与 nuScenes（报告的数据频率仅为 2 赫兹）和 Lyft（采集自一段仅 10 公里的道路）形成对比，但与 Waymo 开放运动数据集互为补充（采用类似的场景挖掘和数据配置方法）。对于这些安全关键问题，互补的数据集至关重要，因为它们为评估泛化能力和探索迁移学习提供了机会。为提升易用性，我们还对 AV2 进行了设计，使其在数据规模和格式方面都易于广泛使用——表2详细比较了本数据集与其他近期预测数据集。

自动驾驶感知面临的更广泛问题。除了目标检测和运动预测任务外，新的大规模自动驾驶传感器数据集也为探索数十种新的感知问题创造了机会，尤其是那些有望通过自监督解决的问题。近期已有一些新问题被提出；得益于 SemanticKITTI [1]，视频中的实时三维语义分割受到了关注。高清地图自动化[54, 30] 和高清地图变化检测 [26] 也受到更多关注，此外还有三维场景流和像素级场景仿真 [50, 8]。现有数据集还包含热成像图像 [10, 9]等独特模态。我们的激光雷达数据集支持针对自由空间预测 [23] 或点云预测 [48, 49]的新方法开展大规模自监督训练。

3 Argoverse 2 数据集

3.1 传感器数据集

Argoverse 2 传感器数据集是*Argoverse 1* 三维跟踪数据集的后继版本。AV2 的规模更大，包含 1,000 个场景，而 Argoverse 1 只有 113 个；此外，AV2 中的每个场景也更加丰富 —— 非车辆、非行人立方体标注框的数量是 Argoverse 1 的 23 倍。Argoverse 2 传感器数据集中的 30 秒场景由作者手动挑选，旨在包含拥挤场景、代表性不足的物体、值得关注的天气状况以及有趣的行为，例如加塞和乱穿马路。每个场景时长为 15 秒。表 1 将 AV2 传感器数据集与若干自动驾驶数据集进行了比较。图 1、2 和 3 展示了 AV2 的场景在标注范围、物体多样性、物体密度和场景动态性方面优于其他数据集。

与我们的数据集最相似的是影响力深远的 nuScenes [4] ——这两个数据集都包含 1,000 个场景和高精地图，不过 Argoverse 的独特之处在于还提供地面高度图。nuScenes 包含雷达数据，而 AV2 包含立体图像。nuScenes 的分类体系庞大，包含 23 种物体类别，其中 10 种具备适用于训练与评估的数据。我们的数据集包含 30 种物体类别，其中 26 种的数据采样充分，适用于训练与评估。nuScenes 覆盖两座城市，而我们提出的数据集覆盖六座城市。

表 1：Argoverse 2传感器和激光雷达数据集与其他传感器数据集的比较 s.

Name	# 场景	城市	激光雷达?	# 摄像头	立体	高精地图?	# 类别	# 评估类别
Argoverse 1 [6]	113	2	✓	7	✓	✓	15	3
KITTI [17]	22	1	✓	2	✓		3	3
nuScenes [4]	1,000	2	✓	6		✓	23	10
ONCE [36]	581	–	✓	7			5	3
Waymo Open [45]	1,150	3	✓	5			4	4
Argoverse 2 传感器	1,000	6	✓	9	✓	✓	30	26
Argoverse 2 激光雷达	20,000	6	✓	-		✓	-	-

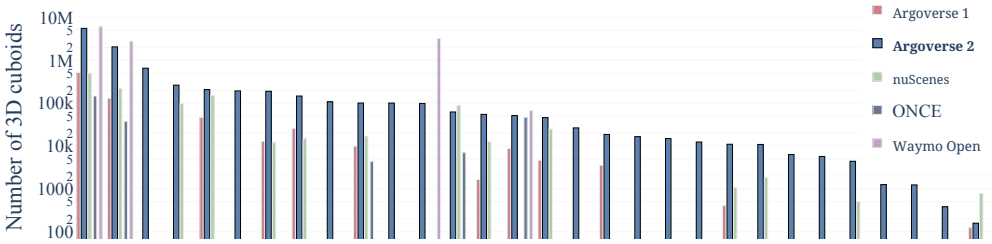

图1：Argoverse 1 三维跟踪、Argoverse 2 传感器数据集、nuScenes、ONCE 和 Waymo Open 中各类别标注的3D长方体数量。nuScenes 的标注频率为 2 赫兹，而 Argoverse 为 10赫兹，但这并未考虑 Argoverse 2 中物体多样性的相对增长。

传感器套件。激光雷达扫描以 10 赫兹的频率采集，同时由 7 个摄像头以 20 帧/秒的速度拍摄图像，这些摄像头的位置经过设计，可提供完整的全景视野。此外，还提供相机内参、外参以及全局坐标系中的 6 自由度自车位姿。激光雷达回波由两台 32线激光雷达采集，它们以 10 赫兹的频率朝同一方向旋转，但朝向相差 180°。摄像头与两台激光雷达同步触发，使帧率达到 20 赫兹。七个全局快门摄像头与激光雷达同步，其曝光时间的中心与激光雷达扫描经过各自视场的时刻对齐。附录中提供了一张示意图，展示车辆的传感器套件及其坐标系。

激光雷达同步精度。在 AV2 中，我们显著提升了摄像头和激光雷达相较于 Argoverse 1 的同步效果。我们的同步精度在 $[-1.39, 1.39]$ 毫秒以内，优于 Waymo Open 数据集（其报告值为 $[-6, 7]$ 毫秒 [45]）。

标注。 AV2 传感器数据集包含针对我们 30 类类别分类体系中的物体所进行的 10 赫兹三维长方体标注（图1）。立方体标注框具有轨迹标识符，对于

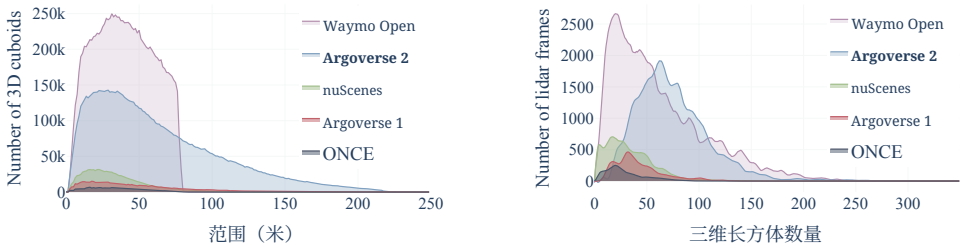

图2：**左：** Argoverse 2传感器数据集中按距离划分的已标注三维长方体数量。Argoverse 2 的立方体标注框中，约有 14% 位于 75 米以外；Waymo Open、nuScenes 和 ONCE 的这一比例均低于 1%。**右：** 每个激光雷达帧中的三维长方体数量。Argoverse 2 的每个激光雷达帧平均包含 75 个三维长方体；Waymo Open 平均为 61 个，nuScenes 为 33 个，ONCE 为 30 个。

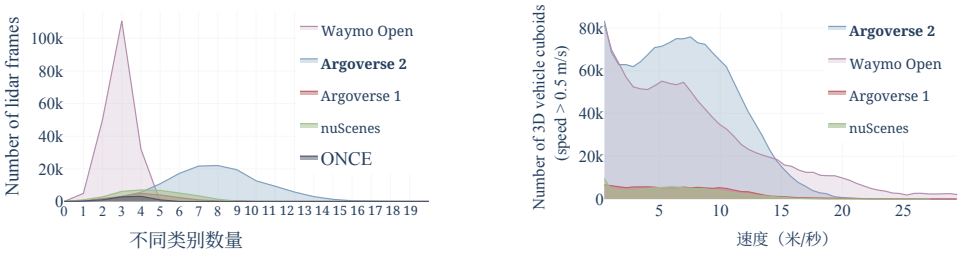

图 3：**左：** Argoverse 2 传感器数据集中每个激光雷达帧标注的类别数量。每个场景中，Argoverse 2 的多样性约为 Argoverse 1 的 2× 倍，比 Waymo Open 多样 3× 倍。**右：** 车辆类别的速度分布。我们仅考虑速度大于 0.5 米/秒的行驶中的车辆。Argoverse 2 中行驶中的车辆数量约为 Waymo Open 的 1.3× 倍。Argoverse 2 中约有 28% 的车辆处于行驶状态，平均速度为 7.27 米 / 秒。我们没有与 ONCE 数据集进行比较，因为它不提供三维长方体的跟踪信息。

同一对象实例而言，这些标识符在时间上保持一致。位于“感兴趣区域”（ROI）内的物体会被标注；该区域指距离地图所标注的“可行驶”区域五米以内的范围。

隐私。 无论是在车辆内部，还是在可行驶区域之外，所有人脸和车牌都经过大幅模糊处理，以保护隐私。

SensorDataset 划分。我们随机将数据集划分为训练集、验证集和测试集，分别包含 700、150 和 150 个场景。

3.2 激光雷达数据集

Argoverse 2 激光雷达数据集 旨在支持激光雷达领域的自监督学习研究，以及点云预测 [48, 49]。由于激光雷达数据比完整的传感器套件更加紧凑，我们可以纳入时长加倍的场景（30 秒而非 15 秒），并将场景数量大幅增加——从 1,000 个增至 20,000 个——相当于驾驶时长约增加 40 倍，而所需存储空间仅为原来的 5 倍。AV2 激光雷达数据集按照与预测数据集（第 3.3.2 节）相同的标准进行筛选，以确保每个场景都具有研究价值。虽然激光雷达数据集没有 3D 目标标注，但每个场景都附带一张高精地图，其中包含有关该场景的丰富三维信息。

我们的数据集包含 20,000 段、每段 30 秒的序列，是迄今为止规模最大的同类数据集。同期发布的唯一类似数据集 ONCE [36], 包含 100 万帧激光雷达帧，而我们的数据集包含 600 万帧激光雷达帧。我们的数据集以 10 赫兹的频率采样，而 ONCE 的采样频率为 2 赫兹，因此更适用于点云预测或自监督任务，因为这些任务重视点云随时间的演变。

激光雷达数据集划分。我们将数据集随机划分为训练集、验证集和测试集，分别包含 16,000、2,000 和 2,000 个场景。

3.3 运动预测数据集

运动预测旨在预测局部环境中动态参与者的未来状态（或占用图）。自动驾驶中相关参与者的示例包括：车辆（停放和行驶中的车辆）、行人、骑行者、滑板车和宠物。预测系统生成的未来状态预测是运动规划的主要输入，运动规划会根据这些预测进行轨迹选择。生成这些预测是一项复杂的多模态任务，涉及众多不同、部分可观测且彼此存在社会交互的智能体。然而，借助利用观测到的真实未来状态为数据“自标注”的能力，运动预测成为机器学习的理想应用领域。

在 Argoverse 1 成功的基础上，Argoverse 2 运动预测数据集提供了一组更新的预测场景，这些场景采集自自动驾驶车队。以下列出的设计决策总结了我们从内部研究与开发中汲取的经验，

表2：Argoverse 2 运动预测数据集与其他近期运动预测数据集的比较。连字符“-”表示属性不适用或不可用。如果在数据收集之后按有趣程度挖掘有趣场景/参与者，而不是获取全部/随机样本，则我们将“按有趣程度挖掘”定义为真。† 公开排行榜数据截至2021年8月27日获取。

	ARGOVERSE [6]	INTER [52]	LYFT [22]	WAYMO [12]	NUSCENES [4]	YANDEX [34]	OURS
# 场景	324k	-	170k	104k	41k	600k	250k
# 独立曲目	11.7M	40k	53.4M	7.6M	-	17.4M	13.9M
平均轨迹长度	2.48 秒	19.8 秒	1.8 秒	7.04 秒	-	-	5.16 秒
总时间	320 小时	16.5 小时	1118 小时	574 小时	5.5 小时	1667 小时	763 小时
场景持续时间	5 s	-	25 s	9.1 秒	8 s	10 s	11 s
测试预测时域	3 s	3 s	5 s	8 s	6 s	5 s	6 s
采样率	10 赫兹	10 赫兹	10 赫兹	10 赫兹	2 Hz	5 Hz	10 赫兹
# 城市	2	6	1	6	2	6	6
独特道路	290 千米	2 km	10 千米	1750 千米	-	-	2220 千米
每种情景的平均轨迹数	50	-	79	-	75	29	73
# 评估对象类别	1	1	3	3	1	2	5
多智能体评估	×	✓	✓	✓	×	✓	✓
按有趣程度挖掘	✓	×	-	✓	×	×	✓
向量图	✓	×	×	✓	✓	×	✓
下载大小	4.8 GB	-	22 GB	1.4 TB	48 GB	120 GB	58 GB
# 公开排行榜条目†	194	-	935	23	18	3	-

以及近 260 个不同团队[3] 通过 3 项竞赛 [43]提交的 2,700 多份作品所带来的反馈：

1. **运动预测是长尾领域中的安全关键系统**。 因此，我们的数据集侧重于包含不同类型焦点智能体的多样化有趣场景（见第 3.3.2 节）。我们的目标是鼓励开发能够在长尾事件期间确保安全的方法，而不是优化在“轻松里程”上的预期性能。

2. **任务难度存在一个“恰到好处”的区间**。 如附录中的图10所示，Argoverse 1 测试集上的性能已开始趋于稳定。Argoverse 2 旨在逐步提高预测难度，从而推动未来几年富有成效的专项研究。这些变化意在鼓励开发能够在较长预测时域（3 s $\rightarrow$ 6s）上表现良好、处理多种动态物体（1 $\rightarrow$ 5）并确保长尾场景安全的方法。未来的 Argoverse版本可以通过缩短观测窗口、延长预测时域，继续提高问题难度。

3. **易用性至关重要**。 Argoverse 1 得益于规模庞大且活跃的研究社区，这在很大程度上要归功于其设置和使用简单。因此，我们特别注意确保现有的 Argoverse 模型能够轻松移植到 Argoverse 2 上运行。具体而言，我们优先考虑地图元素的直观访问方式，以鼓励采用将车道图作为强先验的方法。为改善训练和泛化，所有位姿也都经过插值和重采样，精确达到 10 赫兹（Argoverse 1 的采样频率则是近似值）。新数据集包含的场景更少，但单个场景更长、更复杂；这确保数据集总规模既足以训练复杂模型，又足够小，便于访问。

3.3.1 数据表示

该数据集由 250,000 个互不重叠的场景组成，这些场景取自美国六个不同的城市驾驶环境，并按 80/10/10 的比例随机划分为训练集/验证集/测试集。数据集共包含 10 种物体类型，动态类别和静态类别各占 5 种（见图 4）。每个场景都包含一幅局部矢量地图，以及自车在局部环境中观测到的所有轨迹的 11 秒（10 赫兹）轨迹数据（二維位置、速度和朝向）。每个场景的前 5 秒称为观测 窗口，随后 6 秒称为预测 时域。

在每个场景中，我们将一条轨迹标记为“焦点智能体”。焦点轨迹在整个场景持续期间均保证被完整观测，并经过专门筛选，以最大限度地呈现其与地图特征及附近其他参与者之间的有趣交互（见第 3.3.2 节）。为评估多智能体预测，我们还将部分轨迹标记为“评分对象”（如图 5 所示），并确保这些轨迹与场景相关且达到最低数据质量要求。

[3]此计数包括 未发布到公开排行榜的私人提交 。

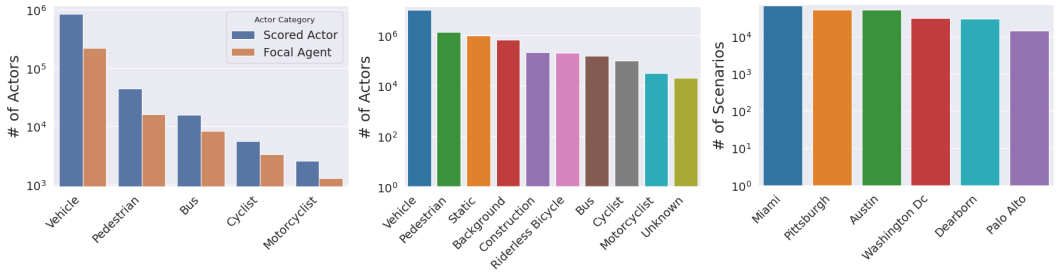

图 4：运动预测数据集中的物体类型和地理分布直方图。**左侧**：焦点和评分类别中各物体类型的直方图。**中间**：数据集中所有轨迹的物体类型直方图。不同静态物体类型之间的细粒度区分（例如，施工锥桶 与无人骑行的自行车）是预测数据集所独有的。**右侧**：数据集中各大都市区的直方图。

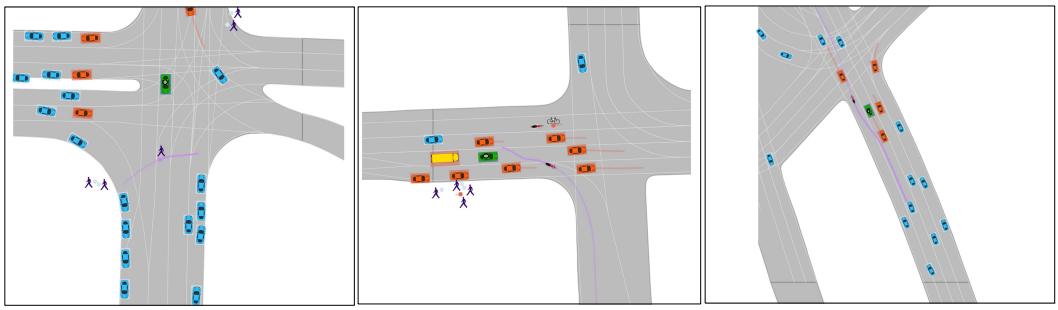

图 5：运动预测数据集中一些有趣场景的可视化。这些场景展示了多种物体类型（车辆、行人、公交车、骑自行车的人或摩托车骑手）。自车以绿色表示，焦点智能体以紫色表示，评分对象以橙色表示。其他未评分的轨迹以蓝色显示。物体位置记录于观测历史的最后一个时间步。为便于可视化，焦点智能体显示了完整的 5 秒历史和 6 秒未来，而其他评分对象仅显示 1.5 秒的未来。**左侧**展示行人从自车前方穿过的场景，**中间**和**右侧**则展示摩托车骑手在交通流中穿行的场景。

3.3.2 挖掘有趣场景

Argoverse 2 的源数据取自车队日志，这些日志带有标注，标记了有趣或难以预测的事件。每条日志都被截取为 30 秒，并经过趣味性 评分模块处理，以便在数据选择中提高自然分布长尾部分样本的占比。我们采用启发式方法，从五个维度对场景中的每条轨迹进行评分：物体类别、运动学、地图复杂度、社会上下文，以及与自车的关系（详见附录）。

最终场景通过提取互不重叠的 11 秒窗口生成，其中至少有一条候选轨迹在整个时长内均被完整观测。得分最高的候选轨迹称为“焦点智能体”；自车 30 米范围内所有其他被完整观测的轨迹称为“评分对象”。由此得到的数据集丰富多样、颇具挑战性，同时规模适中，便于广泛使用（参见表2中的下载大小）。图 6 显示，最终数据集比 Argoverse 1.1 有趣得多，也验证了我们的直觉：在启发式模块中得分较高的参与者，更难准确预测其轨迹。

3.4 高精地图

上述三个数据集中的每个场景都采用相同的高精地图表示。每个场景都包含自己的局部地图区域，类似于 Waymo Open Motion [12] 数据集。这与原始 Argoverse 数据集有所不同，后者将所有场景定位到两张城市级地图上——一张对应匹兹堡，另一张对应迈阿密。附录中提供了一些示例。

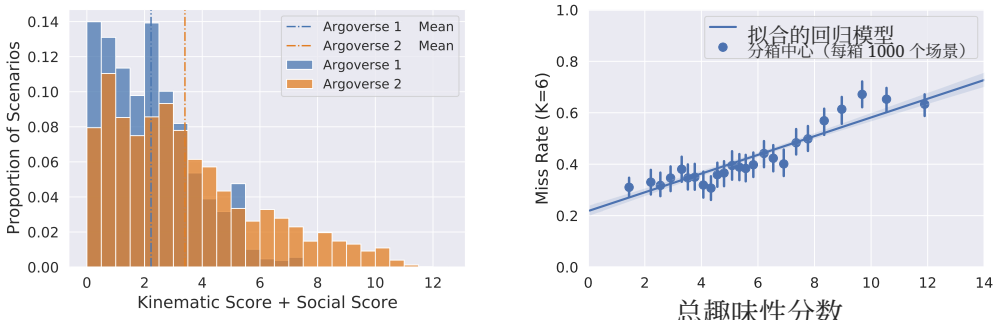

图 6：**左图：**直方图比较了 Argoverse 1.1 和 Argoverse 2 中为焦点智能体分配的趣味性分数的分布。**右图：**图示展示了 Argoverse 2 测试划分中总趣味性分数与预测难度之间的关系。我们对每个场景评估 WIMP [25]，并对计算出的漏检率（K=6，2 米阈值）拟合回归模型。

逐场景地图的优势包括查询效率更高，并且能够处理地图变化。在我们的数据集中，可能会多次观测到某个交叉路口，而在这段时间内，车道、人行横道甚至地面高度都可能发生变化。

车道图。高精地图的核心特征是车道图，它由一个图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 组成，其中 $\mathcal{V}$ 是各个车道段。在附录中，我们逐一列出并定义了每个车道段所提供的属性。与 Argoverse 1 不同，我们提供实际的 3D 车道边界，而不只是中心线。不过，我们的 API 提供了代码，可以按任意所需的采样分辨率快速推断中心线。折线的量化精度为 1 厘米。我们的表示方式比 nuScenes 更丰富，后者仅提供 2D 而非 3D 车道几何。

可行驶区域。 我们不再像 Argoverse 1 那样以栅格化格式提供可行驶区域分割，而是以矢量格式发布，即三维多边形。这带来了多项优势，尤其是压缩方面，使我们能够为数万个场景存储独立地图，同时也可以轻松推导出栅格格式。多边形顶点的量化精度为 1 厘米。

地表高度。 只有传感器数据集包含密集的地表高度图（不过，其他数据集仍在折线上提供稀疏的三维高度信息）。我们为距可行驶区域边界 5 米等值线以内的区域提供地表高度，并将其定义为感兴趣区域 (ROI) [6]。之所以如此，是因为对于建筑物内部和密集建成的城市街区内部，地表高度的概念并不明确；由于遮挡，地面车辆无法观测这些区域。栅格网格的量化分辨率为 30 厘米，高于 Argoverse 1 中的 1 米分辨率。

局部地图范围。每个场景的局部地图包括在自车轨迹的 l_2-范数方向上扩张 100 米范围内找到的所有实体。

4 实验

Argoverse 2 支持多种下游任务。本节重点介绍三种不同的学习问题：三维目标检测、点云预测和运动预测——分别由传感器、激光雷达和运动预测数据集支持。首先，我们通过在包含二十六个评估类别的数据上训练最先进的三维检测模型，展示 Argoverse 2 传感器数据集具有极具挑战性 且 多样化的 分类体系，其中包括婴儿车、轮椅和狗等“长尾”类别。其次，我们通过点云预测任务展示 Argoverse 2 激光雷达数据集在大规模自监督学习中的实用性。最后，我们展示运动预测实验，为广泛分类体系下的运动预测提供首个基线。

4.1 三维目标检测

表3：Argoverse 2 传感器数据集上的三维目标检测结果，取自2022年12月21日的排行榜。*Detectors* 是 CVPR 2022 自动驾驶研讨会Argoverse 2 三维目标检测挑战赛的获胜者。

方法	MCDS (↑)	MAP (↑)	MATE (↓)	MASE (↓)	MAOE (↓)
CENTERPOINT（我们的）	0.14	0.18	0.49	0.34	0.72
DETECTORS [13]	0.34	0.41	**0.40**	**0.30**	0.54
BEVFUSION [32]	**0.37**	**0.46**	**0.40**	**0.30**	**0.50**

我们使用最先进的无锚三维目标检测模型 CenterPoint [51] 提供三维检测结果基线。我们的 CenterPoint 实现以点云作为输入，并将其裁剪为 200 m × 200 m 的网格，体素分辨率为 [0.1m，在.1 m] 的xy（鸟瞰图）平面内，以及在 z- 轴上为 0.2 m。为适应更大的分类体系，我们加入六个检测头，以促进特征专门化。图 7 使用 nuScenes [4]平均精度指标准确刻画了我们的三维检测基线性能。广泛的分类体系使我们能够评估“轮式设备”（电动滑板车）、“婴儿车”、“狗”和“轮椅”等类别；我们发现，尽管拥有大量训练数据，采用强大基线方法时这些类别的性能仍然较差。

在表3中，我们展示了提交情况的一个快照至 Argoverse 2 三维目标检测排行榜。

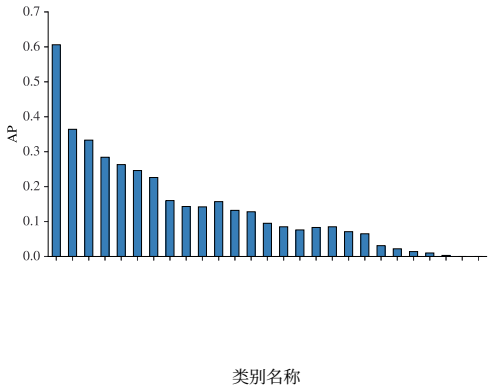

图7：我们的 3D 目标检测基线在验证集划分上的平均精度，数据来自传感器数据集（*Beta*）。我们的实验展示了丰富多样的分类体系，以及具有挑战性的“长尾”类别。

4.2 点云预测

我们使用 Argoverse 2 激光雷达数据集，按照 SPF2 [49] 的实验方案执行点云预测。给定一段过去的场景点云序列，模型需要预测一段未来的场景点云序列。我们将过去 1 秒（10 赫兹）的场景点云以距离图像格式作为输入，并预测接下来 1 秒的距离图像。SPFNet 在每个时间步预测两张输出图：第一张输出图是预测的距离值，第二张输出图是有效性掩码。以往的点云预测模型是在 KITTI 或 nuScenes 等较小的数据集上评估的。为探究训练数据量对性能的影响，我们逐步增加用于训练同一模型架构的数据量，最多使用包含 16,000 个序列的完整训练集。

评估。 我们使用三个指标评估预测模型的性能：平均交并比、l_1-范数和倒角距离。平均交并比 用于评估预测的距离掩码。 l_1-范数用于衡量预测距离图像与真实值图像的像素集合之间的平均 l_1 距离；这两者均由真实值距离掩码进行掩码处理。倒角距离 通过将真实值点云与预测场景点云之间两个方向（前向和反向）的倒角距离相加得到，其中预测场景点云是通过对预测距离图像进行反投影得到的。

表4：在测试集划分的激光雷达数据集上进行点云预测的结果。

	# 训练日志						
	125	250	500	1k	2k	4k	16k
平均交并比 (%) (↑)	55.5	63.4	61.7	65.1	68.0	68.4	**70.9**
l_1-归一化 (↓)	13.5	12.5	11.8	9.9	8.9	8.7	**7.4**
倒角距离 (↓)	31.1	25.9	22.4	22.9	20.5	18.2	**14.0**

SPF2 的结果与讨论。 表4列出了点云预测实验的结果。随着训练数据的增加，模型在三个指标上的性能均稳步提升。这些结果以及自监督学习领域的相关研究 [3, 7] 表明，大量训练数据能够带来显著差异。另一个观察结果是，我们的数据集上的预测结果，其倒角距离显著高于 KITTI [49]上的预测结果。我们推测这可能有两个原因：（1）Argoverse 2 激光雷达数据集的感知范围大得多（超过 200 米，而 KITTI 激光雷达传感器的感知范围为 120 米），这往往会显著增大倒角距离。（2）与 KITTI 数据集相比，Argoverse 2 激光雷达数据集中的动态场景占比更高。

4.3 运动预测

我们提出了几种预测基线方法 [6]，尝试利用数据的不同方面。仅使用焦点智能体进行训练且不捕捉任何社会交互的方法包括：恒定速度、最近邻以及 LSTM 编码器-解码器模型（包括使用和不使用地图先验的版本）。我们还评估了 WIMP [25]，将其作为一种能够捕捉社会交互的基于图的注意力方法。所有超参数均取自参考实现。

评估。 基线方法根据标准指标进行评估。按照 [6],，我们使用最小平均位移误差（*minADE*）和最小最终位移误差（*minFDE*）作为指标；它们分别评估预测轨迹与真实值之间的平均 L2 距离和终点 L2 距离，其中预测轨迹取误差最小者。我们还使用漏检率（*MR*），该指标表示在测试样本中，按照终点误差计算，所有预测轨迹均未在真实值 2.0 米范围内的样本所占的比例。最终的性能结果既体现了整个社区在该问题上的进展，也说明与 Argoverse 1.1 相比，数据集难度显著增加。

表5：在车辆、公交车、摩托车骑行者等类车辆物体类型上，*Argoverse 2*运动预测（*Beta*）数据集的运动预测基线方法表现。使用地图先验表示可以访问地图信息，而利用社会上下文则意味着在特征表示中编码其他参与者的状态。在 K=1 时筛选交叉路口（多模态）场景，会导致所有方法的表现较差。由于数据集偏向运动学特征显著的轨迹，恒定速度模型的表现尤其不佳。请注意，WIMP 等现代深度方法在 K=6 时的漏检率仍为 0.42，这表明 Argoverse 2 数据集的难度有所增加。与最佳结果相差不超过 1% 的数值以粗体标出。

模型	MAP 地图先验	社会上下文	K=1 最小平均位移误差↓	K=1 最小最终位移误差↓	K=6 失误率↓	K=6 最小平均位移误差↓	K=6 最小最终位移误差↓	MR↓
恒定速度 [6]			7.75	17.44	0.89	-	-	-
最近邻 [6]			4.46	11.71	**0.81**	2.18	4.94	0.60
最近邻 [6]	✓		6.45	15.51	0.84	4.30	10.08	0.78
长短期记忆网络 [6]			**3.05**	8.28	0.85	-	-	-
长短期记忆网络 [6]	✓		5.07	12.71	0.90	3.73	9.09	0.85
WIMP [25]	✓	✓	**3.09**	**7.71**	0.84	**1.47**	**2.90**	**0.42**

表6：Argoverse 2 运动预测数据集上的运动预测结果，取自2022年12月21日的在线排行榜。*BANet* 是CVPR 2022年自动驾驶研讨会Argoverse 2运动预测挑战赛（第1名）的获胜者，QML和GANet获得了荣誉提名（HM）奖。条目按照*Brier*最小最终位移误差排序如下。

方法	K=1 最小平均位移误差↓	K=1 最小最终位移误差↓	K=6 失误率↓	K=6 最小平均位移误差↓	K=6 最小最终位移误差↓	K=6 失误率↓	K=6 Brier最小最终位移误差↓
THOMAS（GOHOME标量） [20]	1.95	4.71	0.64	0.88	1.51	0.20	2.16
GORELA（无集成） [11]	1.82	4.62	0.61	0.76	1.48	0.22	2.01
GANET（集成）（HM） [47]	1.81	4.57	0.61	0.73	1.36	**0.17**	1.98
GANET（无集成） [47]	**1.77**	**4.48**	**0.59**	0.72	**1.34**	**0.17**	1.96
QML (HM) [44]	1.84	4.98	0.62	**0.69**	1.39	0.19	1.95
BANet (OPPRED)（#1） [53]	1.79	4.61	0.60	0.71	1.36	0.19	**1.92**

基线结果。 表5总结了基线方法的结果。对于 K=1，Argoverse 1 [6] 表明，恒定速度模型（最小最终位移误差=7.89）的表现优于 NN+map(prior)（最小最终位移误差=8.12),

这里的情况并非如此。这进一步证明，Argoverse 2 在运动学方面更加多样化，无法通过恒定速度假设来解决。令人意外的是，采用地图先验的最近邻和长短期记忆网络变体，其表现不如未采用地图先验的变体，这表明这些基线方法在利用地图方面仍有很大的改进空间。对于 K=6，WIMP 的表现显著优于其他所有基线方法。这强调了训练具有表达能力的模型至关重要：此类模型不仅能够利用地图先验和社会上下文，还能做出多样化的预测。我们的前 3 届 Argoverse运动预测竞赛 [43]呈现出类似趋势：基于图的注意力方法（例如 [25, 31, 37]）继续在竞赛中占据主导地位，在 K=6时的准确度几乎是下一个最佳基线方法（最近邻）的两倍。尽管如此，一些基于栅格化的方法（例如 [19]）也展现出了令人鼓舞的结果。最后，我们还在迁移学习和不同物体类型的背景下评估了基线方法，相关结果总结于附录中。

In T表 6 展示了 Argoverse 2运动预测排行榜上提交内容的概况 .

5 结论

讨论。 在这项工作中，我们介绍了构成 Argoverse 2 的三个新数据集，并针对三项任务提供了基线探索：三维目标检测、点云预测和运动预测。我们的数据集为许多其他任务带来了新的机遇。我们认为，与现有数据集相比，我们的数据集具有优势，包括高精地图、丰富的分类体系、地理多样性以及有趣的场景。

局限性。 与任何人工标注的数据集一样，其中存在标签噪声，不过我们会在发布前尽量将其降至最低。运动预测数据集中不包含物体的三维边界框，但根据物体类型可以对目标范围作出合理假设。运动预测数据集也存在不完善的跟踪，这与最先进的三维跟踪器的表现一致。

参考文献

[1] Jens Behley、Martin Garbade、Andres Milioto、Jan Quenzel、Sven Behnke、Cyrill Stachniss 和 Jurgen Gall。SemanticKITTI：用于激光雷达序列语义场景理解的数据集。发表于*ICCV*，2019年10月。[2] Alex Bewley、Pei Sun、Thomas Mensink、Drago Anguelov 和 Cristian Sminchisescu。用于尺度不变三维目标检测的距离条件膨胀卷积。发表于机器人学习会议，2020年。[3] Tom B Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell 等。语言模型是少样本学习者。*arXiv*预印本 *arXiv:2005.14165*，2020年。[4] Holger Caesar、Varun Bankiti、Alex H. Lang、Sourabh Vora、Venice Erin Liong、Qiang Xu、Anush Krishnan、Yu Pan、Giancarlo Baldan 和 Oscar Beijbom。nuScenes：自动驾驶多模态数据集。发表于*CVPR*，2020年。[5] Yuning Chai、Pei Sun、Jiquan Ngiam、Weiyue Wang、Benjamin Caine、Vijay Vasudevan、Xiao Zhang 和 Dragomir Anguelov。直击要点：利用图卷积核在距离图像中实现高效三维目标检测。发表于*CVPR*，2021年6月。[6] Ming-Fang Chang、John Lambert、Patsorn Sangkloy、Jagjeet Singh、Slawomir Bak、Andrew Hartnett、De Wang、Peter Carr、Simon Lucey、Deva Ramanan 和 James Hays。Argoverse：基于丰富地图的三维跟踪与预测。发表于*CVPR*，2019年。[7] Ting Chen、Simon Kornblith、Mohammad Norouzi 和 Geoffrey Hinton。用于视觉表征对比学习的简单框架。发表于 *ICML*，2020年。[8] Yun Chen、Frieda Rong、Shivam Duggal、Shenlong Wang、Xinchen Yan、Sivabalan Mani-vasagam、Shangjie Xue、Ersin Yumer 和 Raquel Urtasun。GeoSim：通过几何感知合成实现逼真的自动驾驶视频仿真。发表于*CVPR*，2021年6月。[9] Yukyung Choi、Namil Kim、Kibaek Park、Soonmin Hwang、Jae Shin Yoon、Kyounghwan An 和 In So Kweon。用于自主驾驶与辅助驾驶的 KAIST 多光谱昼夜数据集。*IEEE*智能交通系统汇刊，19(3):934–948，2018年。

[10] Yukyung Choi、Namil Kim、Kibaek Park、Soonmin Hwang、Jae Shin Yoon、Yoon In 和 Inso Kweon。全天视觉地点识别：基准数据集与基线。发表于*IEEE*计算机视觉与模式识别会议研讨会。变化环境中的视觉地点识别研讨会，2015年。[11] Alexander Cui、Sergio Casas、Kelvin Wong、Simon Suo 和 Raquel Urtasun。GORELA：为实现视点不变运动预测而相对而行。*arXiv* 预印本 *arXiv:2211.02545*，2022年。[12] Scott Ettinger、Shuyang Cheng、Benjamin Caine、Chenxi Liu、Hang Zhao、Sabeek Pradhan、Yuning Chai、Benjamin Sapp、Charles Qi、Yi n Zhou、Zoey Yang、Aurelien Chouard、Pei Sun、Jiquan Ngiam、Vijay Vasudevan、Alexander McCauley、Jonathon Shlens 和 Dragomir Anguelov。面向自动驾驶的大规模交互式运动预测：Waymo 开放运动数据集。*CoRR*，abs/2104.10133，2021年。[13] Jin Fang、Qinghao Meng、Dingfu Zhou、Chulin Tang、Jianbing Shen、Cheng-Zhong Xu 和 Liangjun Zhang。CVPR 2022 自动驾驶研讨会 Argoverse 三维目标检测竞赛技术报告，2022年。[14] Nils Gählert、Nicolas Jourdan、Marius Cordts、Uwe Franke 和 Joachim Denzler。Cityscapes 3 D：九自由度车辆检测数据集与基准。*CoRR*，abs/2006.07864，2020年。[15] Jiyang Gao、Chen Sun、Hang Zhao、Yi Shen、Dragomir Anguelov、Congcong Li 和 Cordelia Schmid。VectorNet：从向量化表示中编码高精地图与智能体动态。发表于*CVPR*，2020年6月。[16] Runzhou Ge、Zhuangzhuang Ding、Yihan Hu、Yu Wang、Sijia Chen、Li Huang 和 Yuan Li。AFDet：无锚单阶段三维目标检测。发表于*CVPR*研讨会，2020年。[17] Andreas Geiger、Philip Lenz 和 Raquel Urtasun。我们准备好迎接自动驾驶了吗？KITTI视觉基准套件。发表于*CVPR*，2012年6月。[18] Jakob Geyer、Yohannes Kassahun、Mentar Mahmudi、Xavier Ricou、Rupesh Durgesh、Andrew S. Chung、Lorenz Hauswald、Viet Hoang Pham、Maximilian Mühlegg、Sebastian Dorn、Tiffany Fernandez、Martin Jänicke、Sudesh Mirashi、Chiragkumar Savani、Martin Sturm、Oleksandr Vorobiov、Martin Oelker、Sebastian Garreis 和 Peter Schuberth。A2D2：奥迪自动驾驶数据集。*CoRR*，abs/2004.06320，2020年。[19] Thomas Gilles、Stefano Sabatini、Dzmitry Tsishkou、Bogdan Stanciulescu 和 Fabien Moutarde。HOME：用于未来运动估计的热力图输出。*arXiv*预印本 *arXiv:2105.10968*，2021年。[20] Thomas Gilles、Stefano Sabatini、Dzmitry Tsishkou、Bogdan Stanciulescu 和 Fabien Moutarde。THOMAS：结合学习式多智能体采样的轨迹热力图输出。发表于国际学习表征会议，2022年。[21] Wei Han、Zhengdong Zhang、Benjamin Caine、Brandon Yang、Christoph Sprunk、Ouais Alsharif、Jiquan Ngiam、Vijay Vasudevan、Jonathon Shlens 和 Zhifeng Chen。三维点云的流式目标检测。发表于欧洲计算机视觉会议，2020年。[22] John Houston、Guido Zuidhof、Luca Bergamini、Yawei Ye、Long Chen、Ashesh Jain、Sammy Omari、Vladimir Iglovikov 和 Peter Ondruska。一千零一小时：自动驾驶运动预测数据集。*arXiv:2006。14480 [cs]*，2020年11月。备注：发表于 CoRL2020。[23] Peiyun Hu、Aaron Huang、John Dolan、David Held 和 Deva Ramanan。利用自监督自由空间预测进行安全局部运动规划。发表于*CVPR*，2021年6月。[24] R. Kesten、M. Usman、J. Houston、T. Pandya、K. Nadhamuni、A. Ferreira、M. Yuan、B. Low、A. Jain、P. Ondruska、S. Omari、S. Shah、A. Kulkarni、A. Kazakova、C. Tao、L. Platinsky、W. Jiang 和 V. Shet。Lyft Level 5自动驾驶数据集。*arXiv*，2019年。[25] Siddhesh Khandelwal、William Qi、Jagjeet Singh、Andrew Hartnett 和 Deva Ramanan。面向自动驾驶的假设情境运动预测。*arXiv*预印本 *arXiv:2008.10587*，2020年。[26] John W. Lambert 和 James Hays。信任，但要验证：用于高清地图变化检测的跨模态融合。发表于神经信息处理系统进展数据集与基准赛道，2021年。[27] Alex H. Lang、Sourabh Vora、Holger Caesar、Lubing Zhou、Jiong Yang 和 Oscar Beijbom。PointPillars：用于点云目标检测的快速编码器。发表于*CVPR*，2019年6月。[28] Edouard Leurent 和 Jean Mercat。密集交通中的自主决策社会注意力。*CoRR*，abs/1911.12250，2019年。

[29] Mengtian Li、Yu-Xiong Wang 和 Deva Ramanan。面向流式感知。发表于欧洲计算机视觉会议，2020年。[30] Qi Li、Yue Wang、Yilun Wang 和 Hang Zhao。HDMapNet：在线高精地图构建与评估框架。*CoRR*，abs/2107.06307，2021年。[31] Ming Liang、Bin Yang、Rui Hu、Yun Chen、Renjie Liao、Song Feng 和 Raquel Urtasun。学习用于运动预测的车道图表示。发表于欧洲计算机视觉会议，2020年。[32] Zhijian Liu、Haotian Tang、Alexander Amini、Xinyu Yang、Huizi Mao、Daniela Rus 和 Song Han。BEVFusion：采用统一鸟瞰视图表示的多任务多传感器融合。*arXiv* 预印本 *arXiv:2205.13542*，2022年。[33] Yuexin Ma、Xinge Zhu、Sibo Zhang、Ruigang Yang、Wenping Wang 和 Dinesh Manocha。TrafficPredict：异构交通智能体的轨迹预测。*CoRR*，abs/1811.02146，2018年。[34] Andrey Malinin、Neil Band、Alexander Ganshin、German Chesnokov、Yarin Gal、Mark J. F. Gales、Alexey Noskov、Andrey Ploskonosov、Liudmila Prokhorenkova、Ivan Provilkov、Vatsal Raina、Vyas Raina、Mariya Shmatova、Panos Tigas 和 Boris Yangel。Shifts：涵盖多个大规模任务中真实分布偏移的数据集。*CoRR*，abs/2107.07455，2021年。[35] Sivabalan Manivasagam、Shenlong Wang、Kelvin Wong、Wenyuan Zeng、Mikita Sazanovich、Shuhan Tan、Bin Yang、Wei-Chiu Ma 和 Raquel Urtasun。LiDARsim：利用真实世界数据实现逼真的激光雷达仿真。发表于*CVPR*，2020年6月。[36] Jiageng Mao、Minzhe Niu、Chenhan Jiang、Hanxue Liang、Jingheng Chen、Xiaodan Liang、Yamin Li、Chaoqiang Ye、Wei Zhang、Zhenguo Li、Jie Yu、Hang Xu 和 Chunjing Xu。自动驾驶的百万场景：ONCE数据集。*arXiv:2106.11037 [cs]*,2021年8月。备注：已被NeurIPS 2021数据集与基准专题接收。[37] Jean Mercat、Thomas Gilles、Nicole El Zoghby、Guillaume Sandou、Dominique Beauvois 和 Guillermo Pita Gil。用于多模态联合车辆运动预测的多头注意力。发表于*IEEE*国际机器人与自动化会议。IEEE，2020年。[38] Jean Mercat、Thomas Gilles、Nicole El Zoghby、Guillaume Sandou、Dominique Beauvois 和 Guillermo Pita Gil。用于多模态联合车辆运动预测的多头注意力，2019年。[39] Abhishek Patil、Srikanth Malla、Haiming Gang 和 Yi-Ting Chen。H3D数据集：用于拥挤城市场景中全周向三维多目标检测与跟踪。*CoRR*，abs/1903.01568，2019年。[40] Quang-Hieu Pham、Pierre Sevestre、Ramanpreet Singh Pahwa、Huijing Zhan、Chun Ho Pang、Yuda Chen、Armin Mustafa、Vijay Chandrasekhar 和 Jie Lin。A*3D数据集：迈向复杂环境下的自动驾驶。*CoRR*，abs/1909.07541，2019年。[41] Matthew Pitropov、Danson Evan Garcia、Jason Rebello、Michael Smart、Carlos Wang、Krzysztof Czarnecki 和 Steven Waslander。加拿大恶劣驾驶条件数据集。国际机器人研究杂志，40(4-5):681–690，2020年12月。[42] Charles R. Qi、Yin Zhou、Mahyar Najibi、Pei Sun、Khoa Vo、Boyang Deng 和 Dragomir Anguelov。基于点云序列的离车三维目标检测。发表于*CVPR*，2021年6月。[43] Jagjeet Singh、William Qi、Tanmay Agarwal 和 Andrew Hartnett。Argoverse运动预测竞赛。https://eval.ai/web/challenges/challenge-page/454/overview。访问日期：08-27-2021。[44] Tong Su、Xishun Wang 和 Xiaodong Yang。Argoverse 2运动预测挑战赛的QML，2022年。[45] Pei Sun、Henrik Kretzschmar、Xerxes Dotiwalla、Aurelien Chouard、Vijaysai Patnaik、Paul Tsui、James Guo、Yin Zhou、Yuning Chai、Benjamin Caine、Vijay Vasudevan、Wei Han、Jiquan Ngiam、Hang Zhao、Aleksei Timofeev、Scott Ettinger、Maxim Krivokon、Amy Gao、Aditya Joshi、Yu Zhang、Jonathon Shlens、Zhifeng Chen 和 Dragomir Anguelov。自动驾驶感知的可扩展性：Waymo Open 数据集。发表于*CVPR*，2020年。[46] Pei Sun、Weiyue Wang、Yuning Chai、Gamaleldin Elsayed、Alex Bewley、Xiao Zhang、Cristian Sminchisescu 和 Dragomir Anguelov。RSN：用于高效、精确激光雷达三维目标检测的距离稀疏网络。发表于*CVPR*，2021年6月。

[47] Mingkun Wang、Xinge Zhu、Changqian Yu、Wei Li、Yuexin Ma、Ruochun Jin、Xiaoguang Ren、Dongchun Ren、Mingxu Wang 和 Wenjing Yang。GANet：用于运动预测的目标区域网络，2022年。[48] Xinshuo Weng、Jianren Wang、Sergey Levine、Kris Kitani 和 Nicholas Rhinehart。4D预测：100,000点的序列预测。发表于 *ECCV '20*研讨会论文集，2020年8月。[49] Xinshuo Weng、Jianren Wang、Sergey Levine、Kris Kitani 和 Nick Rhinehart。利用SPF2反转预测流程：面向序列姿态预测的序列点云预测。发表于机器人学习会议（*CoRL*）论文集，2020年11月。[50] Zhenpei Yang、Yuning Chai、Dragomir Anguelov、Yin Zhou、Pei Sun、Dumitru Erhan、Sean Rafferty 和 Henrik Kretzschmar。SurfelGAN：合成用于自动驾驶的逼真传感器数据。发表于*CVPR*，2020年6月。[51] Tianwei Yin、Xingyi Zhou 和 Philipp Krahenbuhl。基于中心的三维目标检测与跟踪。发表于*CVPR*，2021年6月。[52] Wei Zhan、Liting Sun、Di Wang、Haojie Shi、Aubrey Clausse、Maximilian Naumann、Julius Kummerle、Hendrik Konigshof、Christoph Stiller、Arnaud de La Fortelle 等。交互数据集：包含语义地图、用于交互式驾驶场景的国际性对抗与协作运动数据集。*arXiv* 预印本 *arXiv:1910.03088*，2019年。[53] Chen Zhang、Honglin Sun、Chen Chen 和 Yandong Guo。BANet：采用边界感知网络的运动预测，2022年。[54] Jannik Zürn、Johan Vertens 和 Wolfram Burgard。面向城市驾驶场景理解的车道图估计。*CoRR*，abs/2105.00195，2021年。

6 附录

6.1 关于传感器套件的补充信息

图 8 展示了用于采集 Argoverse 2数据集的传感器套件示意图。图 9 展示了标注的行人3D立方体的速度分布和偏航角分布。

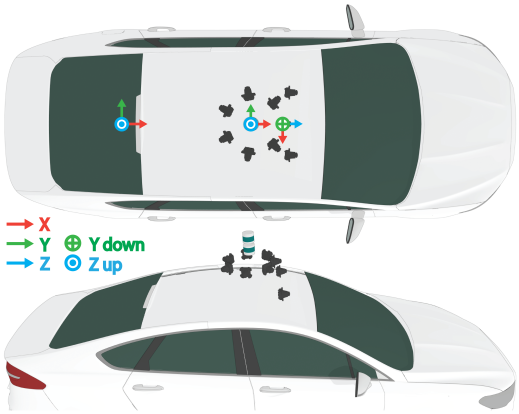

图 8：汽车传感器示意图，展示了三个坐标系：（1）后轴处的车辆坐标系；（2）相机坐标系；以及激光雷达帧。

6.2 关于运动预测数据集的补充信息

6.2.1 趣味性分数

运动学评分会筛选出执行急转弯或显著加速或减速的轨迹。地图复杂度程序会偏向选择穿越底层车道图的复杂路线的轨迹。具体而言，复杂地图区域、穿过交叉路口的路径以及变道都会获得较高评分。

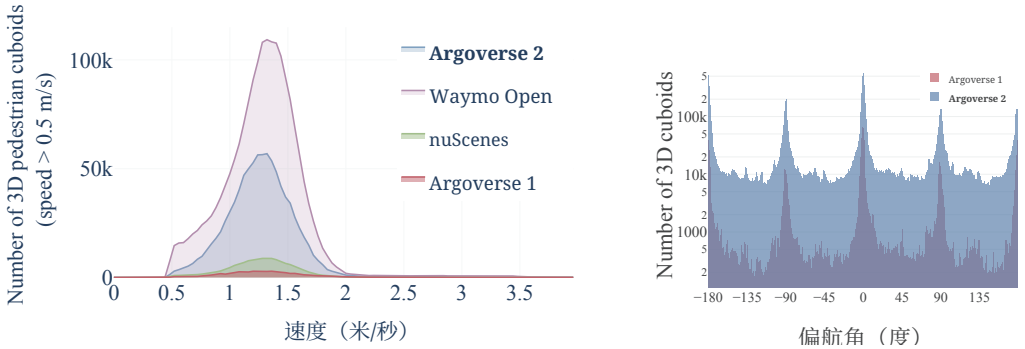

图9：**左侧**：按速度分布统计的行人运动三维长方体数量。当速度大于0.5米/秒时，我们将物体定义为运动物体。**右侧**：按偏航角分布统计的标注的3D长方体数量。

穿过其他参与者密集区域的轨迹会获得较高的社交评分。社交评分还会筛选非车辆物体类别，以确保训练与评估中有足够的稀有类别样本，例如摩托车。最后，自动驾驶车辆评分计划会鼓励选择与自车期望路线相交的轨迹。

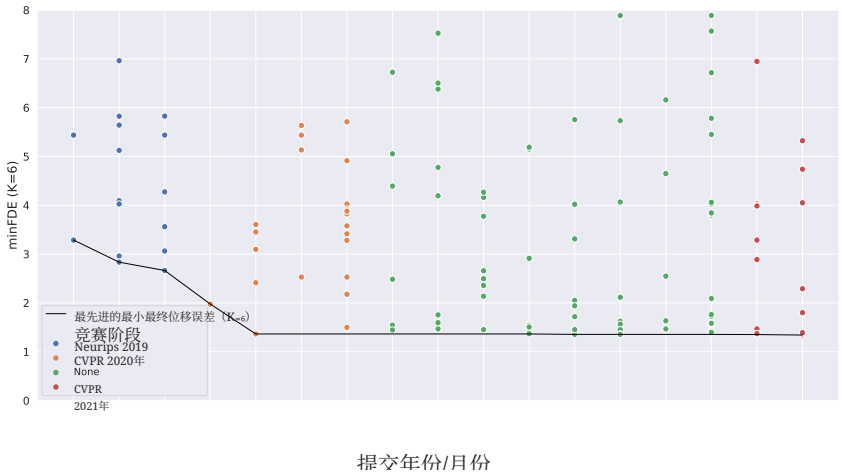

图10：Argoverse 1.1上提交结果的最小最终位移误差指标值随时间的变化。各个点表示向公开排行榜提交的结果。颜色表示具体的竞赛阶段。实心黑线表示最先进水平的性能。研究社区取得了巨大进展，但自2020年初以来已经趋于平稳。不过，我们注意到，性能达到或接近最先进水平的方法数量和多样性仍在持续增长。此外，后续竞赛按照“漏检率”和概率加权最终位移误差对排行榜进行排序，这些指标也显示出一定进展。不过，最小最终位移误差并未显著改善。

6.3 关于高精地图的其他信息

传感器数据集中的高精地图示例在图12中，我们展示了与各个日志/场景相关联的局部高精地图示例。

6.4 其他 3D 检测结果

图13中，我们展示了检测基线的其他评估指标。

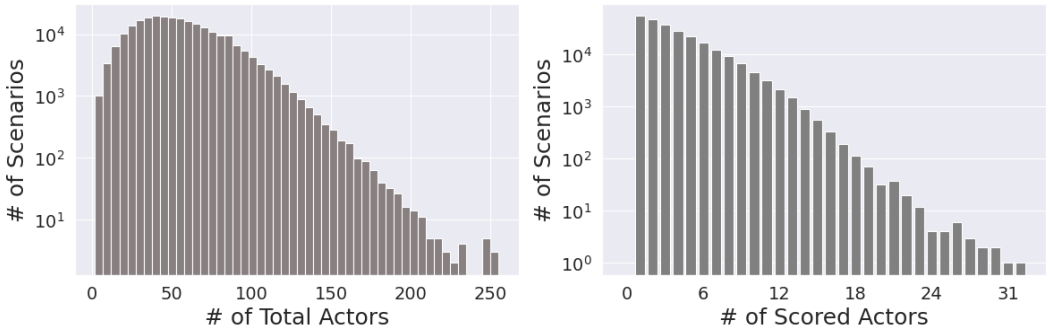

图11：运动预测数据集场景中参与者数量（评分参与者和所有类型参与者）的直方图。激光雷达数据集采用相同标准进行挖掘，因此分布也相同。

地图实体	提供的属性	TYPE	描述
LANE 路段	是否为交叉路口	布尔值	此车道段是否位于交叉路口内。
	车道类型	枚举类型	指定哪些车辆类型可以合法使用此车道通行。
	左侧车道边界	三维折线	城市地图坐标系中的左侧边界折线
	右侧车道边界	三维折线	城市地图坐标系中的右侧边界折线。
	左侧车道标线类型	枚举类型	道路上车道段左侧的道路车道标线类型。
	右侧车道标线类型	枚举类型	道路上车道段右侧的道路车道标线类型。
	左侧邻居	整数	紧邻该车道段左侧的唯一车道段，或无。
	右邻居	整数	紧邻该车道段右侧的唯一车道段，或无。
	后继标识符	整数列表	沿前方行驶可进入的车道段。
	ID	整数	唯一标识符
可行驶 AREA	区域边界	三维多边形	自动驾驶车辆可在不损坏自身的情况下行驶的区域
	ID	整数	唯一标识符
行人 过街通道	边1、边2	三维折线	沿主轴的两条边的端点，从而定义一个多边形。
	ID	整数	唯一标识符
地面高度	-	二维栅格数组	栅格网格量化至30 厘米分辨率。

表 7：每种 Argoverse 2 场景的高精地图属性。

平均精度（AP）

$$AP = \frac{1}{101} \sum_{t \in \mathcal{T}} \sum_{r \in \mathcal{R}} p_{interp}(r) \tag{1}$$

真正例指标 平均平移误差（ATE）

$$ATE = \|t_{det} - t_{gt}\|_2 \tag{2}$$

平均尺度误差（ASE）

$$ASE = 1 - \prod_{d \in \mathcal{D}} \frac{\min(d_{det}, d_{gt})}{max(d_{det}, d_{gt})} \tag{3}$$

平均方向误差（AOE）

$$AOE = |\theta_{det} - \theta_{gt}| \tag{4}$$

综合检测得分（CDS）

$$CDS = mAP \cdot \sum_{x \in \mathcal{X}} (1 - x) \tag{5}$$

其中 $\mathcal{X} = \{mATE_{unit}, mASE_{unit}, mAOE_{unit}\}$

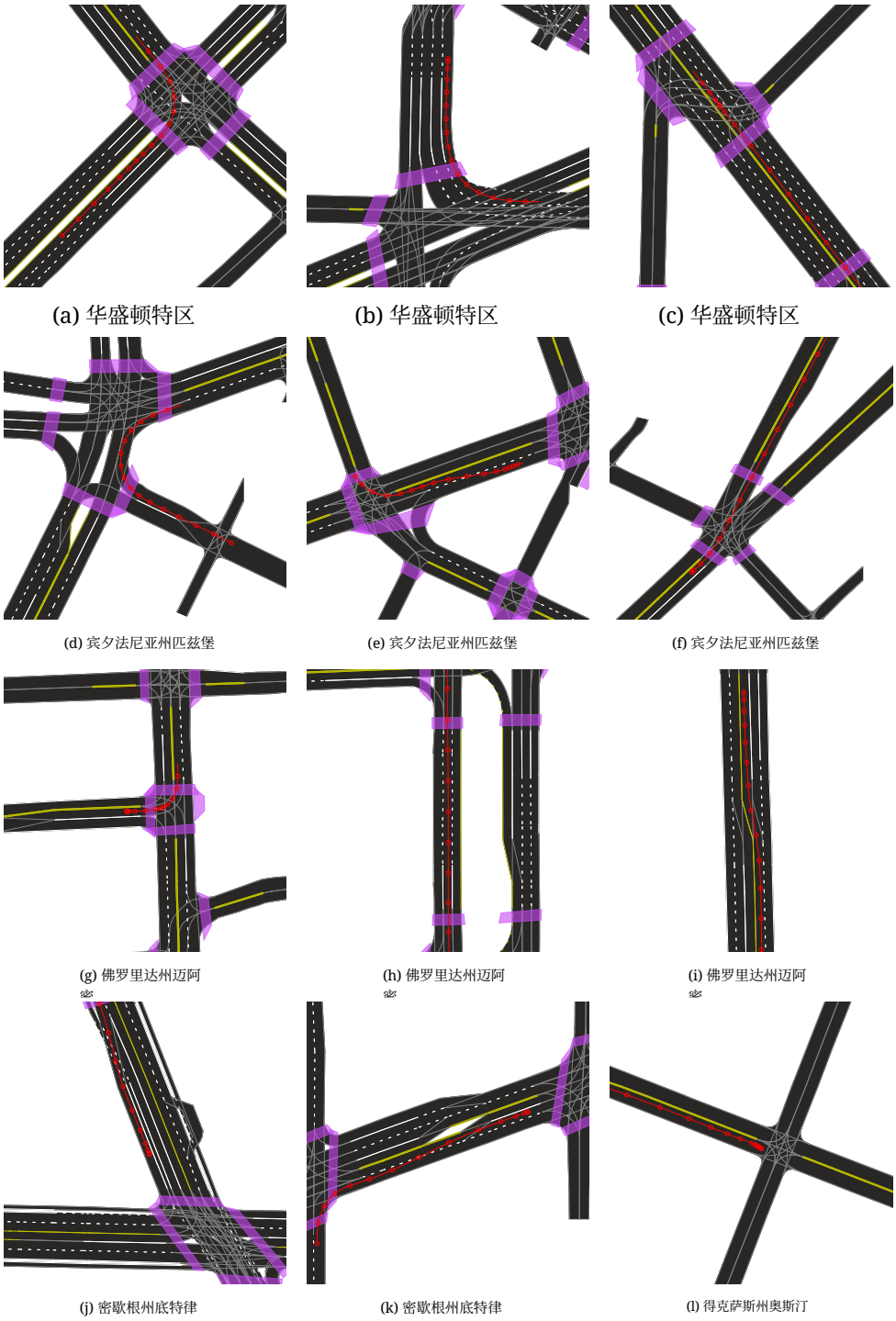

图12：传感器数据集中的自车（自动驾驶车辆）在多个不同城市的局部矢量地图上的轨迹示例。图中展示了一个 100m × 100m 的局部地图区域。人行横道以紫色标示。红色圆圈表示为便于说明而以 1 赫兹离散采样的自动驾驶车辆位姿。数据集中以 >20 赫兹的频率提供位姿，如红线所示的轨迹路径所示。城市布局差异显著，例如，迈阿密的道路通常与南北向、东西向网格平行，而匹兹堡的道路通常并非如此。

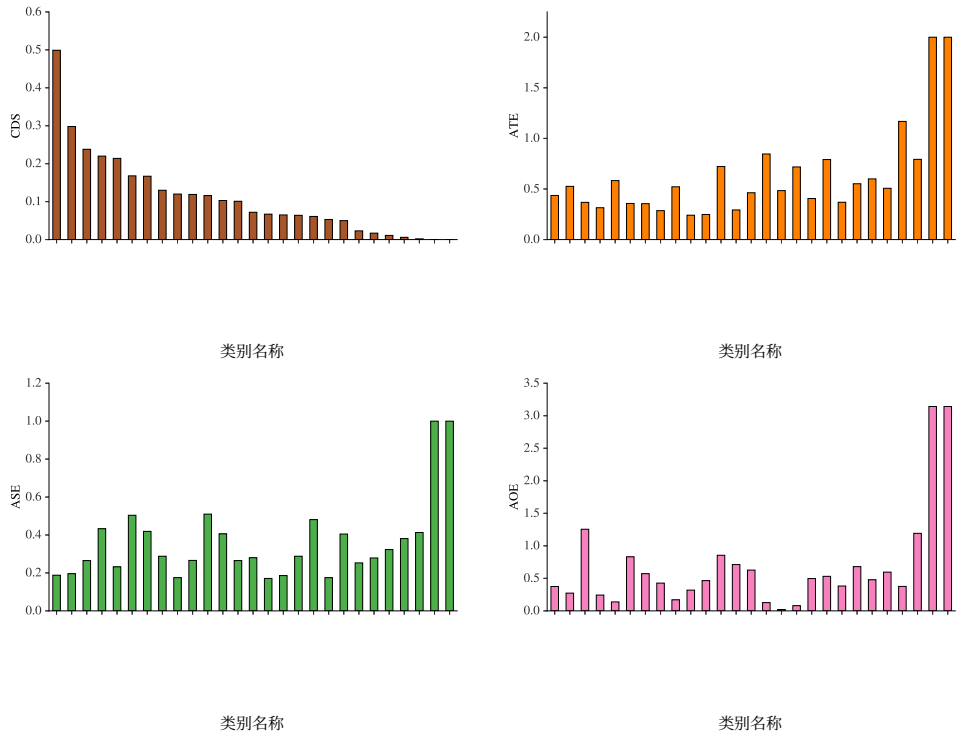

图13：在验证集 划分的*Sensor Dataset*（*Beta*）上进行三维目标检测的性能。**上排：** 综合检测得分（左）。平均平移误差（右）**下排：** 平均缩放误差（左）和平均方向误差（右）。结果展示了在验证集 传感器数据集上的表现。

6.5 SPF2baseline 的训练细节

我们每隔0.5秒采样一次2秒的训练片段（包含1秒的过去数据和1秒的未来数据）。因此，对于一段时长为30秒的训练日志，会采样出59个训练片段。我们使用 Adam优化器训练模型16个训练轮次，学习率为$4e-3$，betas为0.9和0.999，每个 GPU 的批量大小为16。

6.6 其他运动预测实验

6.6.1 迁移学习

迁移学习实验的结果汇总于表8。我们在不同设置下使用 Argoverse 1.1 和 Argoverse 2 对 WIMP 进行训练和测试。不出所料，当模型在同一分发的数据上进行训练和测试时（即训练集和测试集都来自 Argoverse 1.1，或都来自 Argoverse 2），效果最佳。例如，在 Argoverse 2（6秒）上测试 WIMP 时，使用 Argoverse 2（6秒）训练的模型最小最终位移误差 为2.91，而使用 Argoverse 1.1（3秒）训练的模型最小最终位移误差 为6.82（即约差2.3倍）。同样，在反向设置中，当在 Argoverse 1.1（3秒）上测试 WIMP 时，使用 Argoverse 1.1（3秒）训练的模型最小最终位移误差为1.14，而使用 Argoverse 2（6秒）训练的模型最小最终位移误差 为2.05（即约差1.8倍）。这表明，尽管场景数量较少，从*Argoverse 2*（*Beta*）迁移学习到*Argoverse 1.1* 比反向迁移更有用。不过，公开发布的*Argoverse 2*运动预测版本（非 Beta 的2.0版本）与 Argoverse 1.1 的规模相当。

我们注意到，使用不同序列长度训练和测试序列模型是一种常见做法（例如机器翻译）。因此，合理地预期使用 3 秒数据训练的模型在 6 秒预测时域上也能表现良好。造成分布偏移的因素可能有多种，包括预测时域、城市、挖掘流程和物体类型的差异。不过值得注意的是，这些结果表明，Argoverse 2 比其前代数据集更具挑战性，且多样性显著更高。

6.6.2 尝试不同的物体类型

表 9 展示了最近邻基线（不使用地图先验）在不同物体类型上的结果。正如预期，行人的位移误差显著低于其他物体类型。这是因为行人的移动速度显著较慢。不过，这并不意味着行人运动预测问题已经解决，也不意味着我们应该转而关注其他物体类型。相反，这意味着我们需要提出更好的指标，以反映这样一个事实：行人的位移误差较低，往往比车辆的位移误差较高更值得重视。我们将这项工作留待未来开展。

表8：在不同版本的 Argoverse 运动预测数据集上训练和测试时，WIMP 的性能。训练与评估仅限于类车辆物体类型（车辆、公交车、摩托车骑手），因为 Argoverse 1.1 中仅包含车辆。所有结果均针对 K=6，预测时域在括号中注明。值得注意的是，在 3 秒时域上训练的模型在更长的 6 秒时域上表现不佳。下文的“Argoverse 2”指的是*Argoverse 2*（测试版）运动预测数据集。

训练划分（预测时域）	测试划分（预测时域）	最小平均位移误差 ↓	最小最终位移误差 ↓	MR ↓
Argoverse 1.1（3秒）	Argoverse 1.1（3秒）	**0.75**	**1.14**	**0.12**
Argoverse 2（6秒）	Argoverse 1.1（3秒）	1.68	2.05	0.27
Argoverse 1.1（3秒）	Argoverse 2（3秒）	0.94	1.88	0.26
Argoverse 1.1（3秒）	Argoverse 2（6秒）	4.93	6.82	0.77
Argoverse 2（6秒）	Argoverse 2（6秒）	**1.48**	**2.91**	**0.43**

表 9：最近邻基线在不同物体类型上的表现（K=6）。每项评估指标中预测最准确的物体类型以粗体突出显示。

物体类型	#样本	最小平均位移误差 ↓	最小最终位移误差 ↓	MR ↓
All	9955	2.48	5.52	0.66
车辆	8713	2.62	5.87	0.70
Bus	439	2.69	5.59	0.73
行人	677	**0.69**	**1.31**	**0.17**
摩托车骑手	39	2.33	5.07	0.61
骑自行车的人	87	1.48	2.80	0.42

7 Argoverse 2 数据说明书

该数据集是出于什么目的创建的？是否针对某项特定任务？是否旨在填补某个特定空白？请提供说明。

创建 Argoverse 旨在支持全球研究社区，推动自动驾驶关键机器学习任务的发展。本文介绍的 Argoverse 2数据集在最初的 Argoverse 数据集基础上进行了改进。这些数据集支持多种任务，包括三维感知、运动预测和高清地图自动化。

本文提出的三个数据集分别针对这一领域中的不同空白。有关更详细的分类，请参阅正文中的对比图表。

Argoverse 2 传感器数据集的分类体系比类似数据集更丰富。它是同等规模的数据集中唯一包含立体图像的数据集。该数据集中的 1,000 条日志经过筛选，涵盖了多种物体类型及其多样化的交互。

Argoverse 2运动预测数据集的分类体系也比现有数据集更加丰富。该数据集中的场景经过筛选，着重关注难以预测的异常行为。

Argoverse2 激光雷达数据集是规模最大的激光雷达数据集。同期的 ONCE 数据集规模相近，能够支持在激光雷达空间中进行自监督学习。与 ONCE 不同，我们的数据集包含高精地图和高帧率激光雷达数据。

是谁创建了这个数据集（例如，哪个团队或研究小组），以及代表哪个实体（例如，公司、机构或组织）？

Argoverse 2数据集由 Argo AI 的研究人员创建。

制作此数据集需要哪些支持?（例如，谁资助了该数据集的制作？如果有相关资助，请提供资助方名称以及资助项目名称和编号；如果该项目由公司或政府机构支持，请提供相关详情。）该数据集的制作由 Argo AI 资助。

还有其他意见吗? 不适用

组成

构成该数据集的实例代表什么（例如，文档、照片、人物、国家）？是否包含多种类型的实例（例如，电影、用户和评分；人物及其相互作用；节点和边）？请提供描述。

Argoverse 2 的三个组成数据集具有不同的属性，但每个数据集的核心实例都是时长为 11、15 或 30 秒的简短“场景”或“日志”，代表对自动驾驶车辆周围场景的连续观测。

这三个数据集中的每个场景都配有高精地图，其中包括车道边界、人行横道、可行驶区域等。传感器数据集中的场景还包含一张地面高度栅格地图，分辨率为 0.3 米。

总共有多少个实例（如适用，分别统计每种类型）？该传感器数据集包含 1,000 个时长为 15 秒的场景。

该激光雷达数据集包含 20,000 个时长为 30 秒的场景。

该运动预测数据集包含 250,000 个时长为 11 秒的场景。

该数据集是否包含所有可能的实例，还是从更大的集合中抽取的实例样本（不一定是随机样本）？如果该数据集是样本，那么更大的集合是什么？该样本是否能代表更大的集合（例如，地理覆盖范围）？如果是，请说明如何验证/核实其代表性。如果不能代表更大的集合，请说明原因（例如，为了涵盖更多样的实例，或因为某些实例被排除或无法获取）。

该数据集中的场景是自动驾驶车辆车队所采集观测数据的一个样本。这些数据并非均匀抽样。所选样本旨在实现地理多样性（涵盖 6 座城市——匹兹堡、底特律、奥斯汀、帕洛阿尔托、迈阿密和华盛顿特区），包含有趣的行为（例如，汽车进行意外机动）、有趣的天气（例如，雨雪天气），以及包含多种类型物体运动的场景（例如，人群步行、骑手驾驶电动滑板车在多辆车之间穿行、挖掘机在建筑工地作业等）。

每个实例由哪些数据组成? “原始”数据（例如未经处理的文本或图像）还是特征？无论是哪一种，请提供描述。

每个传感器数据集场景持续 15 秒。每个场景包含来自 7 个环视摄像头的 20 帧/秒视频、来自两个前向双目摄像头的 20 帧/秒视频，以及来自两个异相 32 线束激光雷达的 10 赫兹激光雷达回波。环视摄像头经过同步，会在任一激光雷达扫描经过其视场时触发。每个场景都包含随时间变化的车辆位姿，以及用于关联各个传感器的标定数据。

每个激光雷达数据集场景持续 30 秒。这些场景与传感器数据集中的场景类似，但不包含图像。

每个运动预测场景持续 11 秒。这些场景不包含传感器数据，而是包含车辆、行人和自行车等物体的轨迹。轨迹标明每个物体的类别（例如公交车或自行车），以及其位置和航向，采样间隔为 10 赫兹。

与三种类型场景相关联的高精地图包含描述车道、人行横道和可行驶区域的折线。车道构成一个图，包含前驱车道和后继车道，例如，一条分叉车道可以有两个后继车道。车道边界经过精确定位，并包含标线类型（例如双黄实线）。可行驶区域也由多边形描述，是可以行驶但不一定合法的区域，其中包括路肩等区域。

每个实例是否都关联有标签或目标？如果有，请提供描述。

每个传感器数据集场景都包含动态物体（如车辆、行人、婴儿车、狗等）的三维轨迹标注。这些轨迹适用于三维目标检测和三维跟踪等任务的真实值。测试集有意不提供三维轨迹标签。高精地图也可以视为每个实例的标签，可用作车道线检测或地图自动化的真实值。车辆位姿数据可视为视觉里程计的真实值标签。激光雷达深度估计可作为单目或立体深度估计的真实值。

激光雷达数据集除了高精地图外，没有人工标注。不过，不断变化的点云本身可以视为点云预测的真实值。

每个运动预测数据集场景都提供标签，指出哪些轨迹对应“评分对象”。这些轨迹呈现出有趣的行为，并且保证在每个场景的整个时段内都能被观测到；算法需要预测这些轨迹的未来运动。测试集中有意不提供每个场景中参与者的未来运动。

单个实例是否缺少任何信息？如果是，请提供说明，并解释缺少这些信息的原因（例如，因为当时无法获取）。这不包括有意删除的信息，但可能包括删节文本等内容。

在 传感器数据集中，物体仅在可行驶区域周围 5 米范围内进行标注。例如，坐在自家门廊上的人不会被标注。

在 传感器数据集和运动预测数据集中，如果物体移出观测范围或被遮挡，则实例不一定会在整个场景持续期间都被标注。

Z 是否明确记录了各个实例之间的关系（例如，用户的电影评分、社交网络链接）？如果是，请描述如何明确记录这些关系。

三个数据集中的实例彼此互不相同。每个实例都带有对应场景所在区域的高精地图。不过，这些高精地图在空间上可能存在重叠。例如，许多预测场景可能发生在同一个交叉路口。如果数据集用户希望恢复场景之间的空间关系，可以通过 Argoverse API 实现。

是否有推荐的数据划分（例如训练集、开发集/验证集和测试集）？如果有，请描述这些划分，并解释其背后的原因。

我们为每个数据集定义了数据划分。传感器数据集按 700 / 150 / 150 划分为训练集、验证集和测试集。激光雷达数据集按 16,000 / 2,000 / 2,000 划分，运动预测数据集按 200,000 / 25,000 / 25,000 划分。在所有情况下，这些划分都旨在尽可能扩大训练数据集，同时确保验证集和测试集足够大且多样，能够准确地对基于训练集学习的模型进行基准评估。

数据集中是否存在错误、噪声来源或冗余？如果有，请进行描述。

数据集中使用的每种传感器——环视摄像头、立体摄像头和激光雷达——都存在相关噪声。像素强度、激光雷达强度和激光雷达点三维位置都存在噪声。激光雷达点还会被量化为半精度浮点数（float16），因此会产生大约一厘米的量化误差。车辆六自由度位姿也存在噪声。用于说明传感器之间关系的标定可能并不完美。

每个场景的高精地图都可能存在噪声，包括车道边界位置和精确地面高度方面的噪声。

传感器数据集中的 3D 目标标注并不总是与现实世界中物体的空间范围和运动相符。例如，我们假设物体在一个场景中不会改变大小，但汽车开门就可能打破这一假设。对于像素和激光雷达回波相对较少的远处物体，其三维标注的准确性较低。

运动预测数据集中的目标轨迹并不完美，也包含实时三维跟踪方法常见的误差。我们认为，运动预测算法应能在存在此类噪声的情况下良好运行。

该数据集是自包含的，还是链接到或以其他方式依赖于外部资源（例如网站、推文、其他数据集）？如果它链接到或依赖于外部资源：a）是否能保证这些资源会一直存在且内容保持不变；b）是否有完整数据集的官方存档版本（即包含数据集创建时的外部资源）；c）这些外部资源是否存在可能适用于未来用户的任何限制（例如许可、费用）？请描述所有外部资源及其相关限制，并酌情提供链接或其他访问途径。

数据由我们自行托管，与 Argoverse 1 [类似，详见 https://www.argoverse.org/], ；如有更新，我们会保留指向所有先前版本数据集的公开链接。该数据独立于任何先前的数据集，包括 Argoverse 1。

该数据集是否包含可能被视为机密的数据（例如受法律特权或医患保密义务保护的数据，或包含个人非公开通信内容的数据）？如果包含，请予以描述。否。

数据集中是否包含直接查看时可能令人反感、受辱、受到威胁或以其他方式引发焦虑的数据？如果有，请说明原因。否。

数据集是否与人有关？如果没有，您可以跳过本节的其余问题。 on.
是，该数据集包含公共街道上数千人的图像及其行为数据。

数据集是否识别出任何亚群体（例如按年龄、性别划分）？如果有，请说明如何识别这些亚群体，并描述它们在数据集中的分布情况。否。

是否有可能直接或间接（即结合其他数据）从数据集中识别出个人（即一个或多个自然人）？如果可以，请说明如何识别。
我们认为不太可能。所有图像数据均已匿名化。人脸和车牌通过遮挡处理：将其替换为 5×5 网格，每个网格单元的颜色取自该单元内原始像素的平均颜色。此匿名化处理由人工完成，不受我们的三维标注政策限制。例如，坐在自家门廊上、距离道路 10 米的人不会被标注为三维长方体，但其人脸仍会被遮挡。

该数据集是否包含任何可能被视为敏感的数据（例如，揭示种族或族裔出身、性取向、宗教信仰、政治观点或工会成员身份，或位置信息的数据；财务或健康数据；生物识别或基因数据；政府身份证明（例如社会安全号码）；犯罪记录）？如果有，请提供描述。没有。

还有其他意见吗？不
适用

收集

与每个实例相关联的数据是如何获取的？这些数据是直接观测到的（例如，原始文本、电影评分），由受试者报告的（例如，调查问卷回复），还是根据其他数据间接推断或得出的（例如，词性标签、基于模型对年龄或语言的猜测）？如果数据由受试者报告，或是根据其他数据间接推断或得出的，那么这些数据是否

经过验证/核实？如果是，请说明验证方式。传感器数据由一支自动驾驶
车辆车队直接采集。

数据是在什么时间段内收集的？这一时间段是否与实例相关数据的创建时间段一致（例如，近期抓取旧新闻文章）？如果不一致，请说明实例相关数据的创建时间段。最后，请列出该数据集首次发布的时间。
数据收集于2020年和2021年。该数据集在NeurIPS 2021之后，即2022年3月公开。

收集数据时使用了哪些机制或流程（例如硬件设备或传感器、人工手动整理、软件程序、软件 API）？这些机制或流程是如何验证的？

Argoverse 2 数据来自 Argo“Z1”车队车辆。这些车辆使用 Velodyne 激光雷达和传统 RGB 摄像机。所有传感器均由 Argo 进行校准。高精地图和三维目标标注通过计算工具与人工标注相结合的方式创建并验证。运动预测数据集 中的目标轨迹由三维跟踪算法创建。

收集数据的资源成本是多少？（例如，所需的计算资源、相关财务成本和能源消耗是多少？请估算碳足迹。有关此领域的方法，请参阅 Strubell等人。）

数据是在车队正常运营期间采集的，因此记录特定事件的额外开销极小。对数 TB 数据进行转换和后处理，估计耗费了 1,000 个机器小时。根据以中央处理器为中心的工作负载，我们估计碳足迹约为 1,000 磅。

如果该数据集是从更大的数据集中抽取的样本，采用了什么抽样策略（例如，确定性抽样，或具有特定抽样概率的概率抽样）？
传感器数据集中的场景是通过人工审核从更大的集合中选出的。激光雷达数据集和运动预测数据集中的场景则是通过启发式方法选出的，这些方法会在车队运营期间寻找有趣的目标行为。

哪些人参与了数据收集过程（例如，学生、众包工作者、承包商）？他们是如何获得报酬的（例如，众包工作者获得了多少报酬）？Argo 员工和 Argo 实习生整理了数据。数据收集和数据标注由 Argo 员工完成。未使用众包工作者。

是否进行了任何伦理审查流程（例如由机构审查委员会进行的审查）？如果进行了，请描述这些审查流程及其结果，并提供相关支持文件的链接或其他获取途径。没有。

该数据集是否与个人有关？如果无关，您可以跳过本节其余问题。是。

您是直接从相关个人处收集数据，还是通过第三方或其他来源（例如网站）获取数据？

数据采集自公共道路上的车辆，而非来自第三方。

是否已告知相关个人正在进行数据收集？如果是，请描述（或通过屏幕截图或其他信息展示）通知是如何提供的，并提供通知原文的链接或其他访问途径，或者直接复述通知的确切内容。没有，但数据收集并未被隐瞒。Argo 车队车辆上有醒目的标识，并配有明显的

摄像头和激光雷达传感器。这些车辆仅采集来自公共道路的数据。

相关个人是否同意收集和使用其数据？如果是，请说明（或通过屏幕截图或其他信息展示）如何征求并取得同意，并提供相关链接或其他访问方式，或者以其他方式复现这些个人同意的确切措辞。
没有。数据集中的人们身处公共场所，且其外貌已匿名化。驾驶员、行人和弱势道路使用者是公共道路驾驶活动中不可或缺的一部分，因此数据集纳入人员十分重要，这样社区才能开发出更准确的感知系统。

如果已取得同意，是否为同意的个人提供了撤回同意的机制，以便其今后撤回同意或针对特定用途撤回同意？如果是，请提供相关说明，并提供该机制的链接或其他访问方式（如适用）不适用

是否已分析该数据集及其用途对数据主体的潜在影响（例如，是否开展了数据保护影响分析）？如果是，请说明该分析的情况及其结果，并提供任何支持性文档的链接或其他访问方式。没有。

还有其他意见吗？不
适用

预处理／清理／标注

是否对数据进行了预处理／清理／标注（例如离散化或分桶、分词、词性标注、SIFT 特征提取、移除实例、处理缺失值）？如果有，请提供描述。如果没有，可以跳过本节其余问题。
是。图像已从完整分辨率缩小。三维点位置已量化为半精度浮点数。地面高度图已从完整分辨率量化至 0.3 米精度。高精地图多边形顶点位置已量化至 0.01 米精度。三维标注已进行平滑处理。对于运动预测数据集，临时三维轨迹已被抑制，物体位置也已随时间进行平滑处理。

除了预处理／清理／标注后的数据，是否也保存了“原始”数据（例如，以支持无法预见的未来用途）？如果有，请提供“原始”数据的链接或其他访问方式。
是的，但此类数据不公开。

用于预处理／清理／标注实例的软件是否可用？如果可以，请提供链接或其他访问途径。否。

还有其他意见吗？不
适用

USES

该数据集是否已用于任何任务？如果是，请提供描述。
是的，这篇论文在传感器数据集上评测了一种当代三维目标检测方法，并在运动预测数据集上评测了一种当代运动预测方法。

是否有代码仓库链接到使用该数据集的部分或全部论文或系统？如果有，请提供链接或其他访问途径。

是的，t hArgoverse 2 API 可以 be found 网址为 https: //github.com /argoverse /av2-api。
对于 Argoverse 2数据集，我们维护了两个排行榜，分别用于三维检测 [https://eval.ai/web/challenges/challenge-page/1710] 和运动预测 [https://eval.ai/web/ challenges/challenge-page/1719]。
对于 Argoverse 1数据集，我们维护了四个排行榜，分别用于三维跟踪[https://eval.ai/web/challenges/challenge-page/453/overview]， 3D 检测[https://eval.ai/web/challenges/challenge-page/725/overview]， 运动预测[https://eval.ai/web/challenges/challenge-page/454/overview]， 和立体深度估计 [https://eval.ai/web/challenges/challenge-page/917/overview]。Argoverse 1 也曾被用作流式感知挑战赛 [https: //eval.ai/web/challenges/challenge-page/800/overview]的基础。

数据集还可用于哪些其他任务？ or?
这些数据集可用于视觉里程计、姿态估计、车道线检测、地图自动化、自监督学习、运动恢复结构、场景流、光流、接触时间估计、伪激光雷达和点云预测等研究。

数据集的组成，或其收集、预处理/清理/标注方式，是否存在可能影响未来用途的因素？例如，未来的使用者是否需要了解某些事项，以避免因使用数据集而导致对个人或群体的不公平对待（例如，刻板印象、服务质量问题）或其他不良危害（例如，经济损失、法律风险）？如果有，请予以说明。未来的使用者可以采取哪些措施来减轻这些不良危害？没有。

是否有不应使用该数据集执行的任务？如果有，请予以说明 n.
由于人脸和车牌的数据已被模糊处理，因此不应将该数据集用于依赖人脸或车牌真实外貌的任务。例如，运行人脸检测器来估算行人使用人行横道的频率，将无法获得有意义的数据。

还有其他意见吗？不
适用

分发

是否会将该数据集分发给代表其创建方的实体（例如，公司、机构或组织）之外的第三方？如果会，请予以说明。是，该数据集托管于`https://www.argoverse.org/` 与 Argoverse 1 和 1.1 一样。

数据集将如何分发（例如，网站上的 tar 压缩包、API、GitHub）？数据集是否有数字对象标识符（DOI）？
我们为两个 Argoverse 2 数据集（运动预测、传感器）同时提供 tar.gz 压缩档案和原始文件，但仅为激光雷达数据集提供原始文件，均可通过 AWS 传输获取。请参阅 https: //www.argoverse.org/av2.html#download-link。

Argoverse 1 和 Argoverse 1.1 以一系列 tar.gz 文件的形式发布（请参阅 https://www.argoverse.org/av1.html#download-link）。这些文件被拆分开来，以使过程更能抵御中断（例如，一个 2 TB 文件在 3 天后下载失败会令人十分沮丧），同时也便于文件操作（终端用户可能没有 2 TB 的可用空间存放在同一个驱动器上；即使有，也可能无法一次性解压整个文件）。

数据集何时会发布?
数据于 NeurIPS 2021 之后、2022年3月提供下载。

数据集是否会依据版权或其他知识产权（IP）许可，和/或适用的使用条款（ToU）发布? 如果是，请说明此许可和/或使用条款，并提供相关许可条款或使用条款的链接或其他访问途径，或直接重述相关内容，以及这些限制所涉及的任何费用。
是的，该数据集依据与 Argoverse 1 相同的知识共享许可协议发布——CC BY-NC-SA 4.0。详情请参见 https://www.argoverse.org/about.html#terms-of-use。

是否有任何第三方对与这些实例相关的数据施加基于知识产权或其他方面的限制? 如果有，请说明这些限制，并提供相关许可条款的链接或其他访问途径，或直接重述相关内容，以及这些限制所涉及的任何费用。没有。

数据集或其中的单个实例是否受到任何出口管制或其他监管限制? 如果有，请说明这些限制，并提供相关支持文件的链接或其他访问途径，或直接重述相关内容。没有。

还有其他意见吗? 不
适用

维护

谁在支持、托管或维护该数据集? Argo AI

如何联系该数据集的所有者/策展人/管理者（例如，电子邮件地址）?
Argoverse 团队通过 Argoverse 2 API 的 GitHub 页面 https://github. com/argoverse/av2-api/issues 回复。
Argoverse 团队通过 Argoverse1 API 的 GitHub 页面 https://github. com/argoverse/argoverse-api/issues 回复。目前该页面有 2 个未解决的问题和 126 个已关闭的问题。

如有隐私问题，可通过此处获取联系信息：https://www.argoverse.org/ about.html#privacy

是否有勘误? 如果有，请提供链接或其他访问途径。没有。

是否会更新数据集（例如，修正标注错误、添加新实例或删除实例）? 如果会，请说明更新频率、由谁进行更新，以及如何向用户告知更新信息（例如，通过邮件列表或 GitHub）?

Argoverse 2数据集中的各个数据集可能会进行更新以修正错误。Argoverse 1 就曾进行过这种更新，版本升级为 Argoverse 1.1。相关更新将通过 GitHub 和我们的邮件列表发布。

如果该数据集涉及人员，是否对与这些实例相关的数据保留期限设有限制（例如，是否告知相关个人，其数据只会保留一段固定时间，之后将被删除）? 如果有，请说明这些限制，并解释将如何执行。

No.

旧版本的数据集是否会继续获得支持、托管或维护？如果会，请描述具体方式。如果不会，请描述如何向用户告知其将停止维护。
是的。尽管我们已声明 Argoverse 1 已“弃用”，但我们仍在托管它。请参阅 https://www.argoverse.org/av1.html#download-link。如果我们将来弃用 Argoverse 2，也会发出相同的警告。注意：Argoverse 2 并未取代 Argoverse 1。它们是相互独立的数据集集合。

如果其他人想扩展、增强或基于该数据集进行开发或贡献，是否有相应机制？如果有，请予以说明。这些贡献会经过验证或核实吗？如果会，请描述具体方式。如果不会，为什么？是否有向其他用户传达或分发这些贡献的流程？如果有，请予以说明。
是的。例如，流式感知挑战赛由 CMU 研究人员创建，他们为 Argoverse 1.1 数据添加了新的二维物体标注。我们为 Argoverse 2 采用的知识共享许可协议确保社区无需征得 Argo 的许可，也能进行同样的工作。

我们没有机制将这些贡献或补充内容整合回 Argoverse 2「基础」数据集。通常，我们倾向于保持「基础」数据集原样，并通过链接的方式对有价值的补充内容予以认可，正如我们在流式感知挑战赛中所做的那样（请参阅 Argoverse 页面顶部的链接：https://www.argoverse.org/tasks.html）。

还有其他意见吗？不
适用

环境影响声明。计算资源使用量：我们估计在数据集收集和基线实验执行过程中使用了 2,000 个中央处理器小时和 500 个 GPU 小时。