一种用于推导自行车基础设施路面质量的群智感知方法

Ahmet-Serdar Karakaya柏林工业大学 & *ECDF*柏林，德国
ask@mcc.tu-berlin.de

Leonard Thomas柏林工业大学柏林，德国
leth@mcc.tu-berlin.de

Denis Koljada柏林工业大学柏林，德国
deko@mcc.tu-berlin.de

David Bermbach柏林工业大学 & *ECDF*柏林，德国
db@mcc.tu-berlin.de

arXiv:2305.01712v2 [cs.DC] 2 Aug 2023

摘要—全球各地的城市都在努力提高自行车交通的出行方式分担率，以应对交通和碳排放问题。除安全性外，骑行舒适度也是实现这一目标的关键因素，其中包括自行车道的路面质量。本文提出了一种新颖的基于边缘的群智感知方法，利用智能手机传感器数据分析自行车道的路面质量：骑行者记录骑行记录，并在手机上进行预处理后再上传到私有云后端。在那里，额外的分析模块汇总所有可用骑行记录中的数据，以获取路面质量信息，进而用于表面质量感知路由和基础设施维护规划。

索引词——**城市规划、运动传感器数据、传感器数据分析、路面质量、自行车骑行数据**

一、引言

世界各地的城市都致力于提高自行车交通的出行方式分担率，例如解决排放问题和频繁的交通拥堵，同时也希望通过增加日常活动来促进市民健康。实现这一目标的关键因素包括交通安全、地形，以及骑行者的舒适度 [1], [2], [3]。这些因素都受到自行车基础设施可用性和质量的影响，因此需要持续监测和维护。然而，仅监测路面质量就可能是一项艰巨的任务，因为现有基础设施规模庞大。例如，截至2023年3月，德国首都柏林的自行车道总长约为2,300公里，因此，依靠人工检查自行车基础设施以监测路面质量并不可行。

为解决这一问题，我们提出采用一种众包感知方法，让骑行者通过智能手机应用记录日常骑行。利用内置运动传感器，可以将振动或凸起作为路面质量（或更确切地说，是路面质量缺陷）的测量指标。随后，我们可以整合多次骑行记录的数据，估算路面质量。这样一来，只需较低成本便可在很大程度上实现路面质量监测自动化，所得数据还可用于维护规划或表面质量感知路由。尤其是在骑行频率较高的自行车道上，可以迅速发现路面质量问题。

尽管已有其他项目通过众包研究自行车基础设施的路面质量（例如 Luede- mann 等人 [4]），但我们的工作具有独特性和创新性，因为其他方法 (i) 要么侧重于对路面类型进行分类（例如鹅卵石），而我们则关注路面质量 （即，粗糙程度）或（ii）对手机定位及其他属性有严格要求，在这种情况下，我们依靠“群体智慧”策略来过滤噪声。后者也是一种优势，因为与使用专用应用测量道路表面质量的相关方法不同（这可能会迫使骑行者同时运行多个应用），我们的方法可以轻松集成到现有应用中，甚至可以用于分析已经存储的数据集。

为此，我们作出以下贡献：

- 我们提出了一套用于推导路面质量的数据处理流水线，将信号处理与地理聚类技术相结合，并通过手机端的基于边缘的预处理和云端分析来实现（§IV）。
- 我们介绍如何将这套数据处理流水线整合到现有的众包感知项目 SimRa [3]中。该项目迄今为止仅关注交通安全（§IV）。
- 我们介绍如何利用所得数据，通过考虑路面质量的路线规划提升骑行者的舒适度，以及如何向城市管理部门提供这些数据（§V）。
- 我们通过分析 SimRa 数据集 [5], [6], [7] 并将其与八条路面质量各异的街道的现场路况进行比较，评估了我们的方法（§VI）。
- 我们讨论了我们的方法在多大程度上能够实现路面质量监测自动化，以及增加传感器可能如何帮助提高数据质量（§VII）。

II. 场景与约束

利用智能手机和众包感知来追踪路面质量的想法并不新颖。因此，我们专门用本节简要讨论本研究的具体目标，以及目标场景所面临的独特约束。这样做是为了说明为什么我们将在下一节讨论的现有方法无法解决我们所针对的问题。

首先，我们的研究关注骑行者在“颠簸度”方面的体验。因此，我们尝试量化路面的粗糙度，而非路面类型。例如，骑行沥青路面通常比骑行鹅卵石路更平顺。但在实际情况下，沥青路面可能有很多（小）坑洼和修补路面，而鹅卵石路上的石块可能相对平坦（而非圆润），且石块之间的缝隙大多已填平。示例见图 1。在这种情况下，骑行鹅卵石路



(a) 平整的鹅卵石路面



(b) 粗糙沥青

图 1：平整的鹅卵石路面可以提供良好的路面质量，而劣质沥青可能导致颠簸的骑行体验（来源：Flickr.com）。

可能会平顺得多。因此，量化路面的质量而非路面类型会得出不同结果——因此，大量相关研究并不适用于我们所针对的问题。

其次，由于自行车轮胎的宽度，在某个街道路段的一次骑行记录只能覆盖路面中非常小的比例。此外，骑行者很可能会绕开最严重的坑洼和树根。将这两个方面结合起来可以看出，必须基于大量骑行记录来进行粗糙度测量，从而采用“群体智慧”的方法。因此，我们必须吸引庞大的用户群体——这会带来一系列影响。

1) 该方法需要尽量少占用手机的资源（电池、数据传输、计算能力、应用大小），否则用户可能会卸载相应的应用，或不再使用它。因此，该方法必须在手机上对数据进行预过滤，但不太可能完全在手机上运行；基于机器学习的方法可能会存在问题，而且该方法不能指望获得高分辨率的原始数据。

2）该方法不应有任何物理设备安装要求。一些骑行者会将手机安装在车把上（这通常能使加速度计传感器获得最佳效果），而另一些骑行者则将手机放在口袋、背包等处。任何要求骑行者 采用特定设备配置的方法，都会使大量潜在用户望而却步。因此，许多在实验室条件下有效的方法，在实际道路环境中可能并不适用。

3) 该方法需要能够轻松集成到现有的骑行应用中。正如我们在 SimRa [3], 中看到的那样，许多骑行者不会同时使用多个骑行应用。为了最大限度地扩大潜在用户群体，该方法需要以能够轻松集成到现有骑行应用中的方式进行设计。总体而言，这意味着该方法需要作为独立的后台任务运行，并且不应涉及手动标注等类似活动。它们还可能受到与隐私相关的限制。

我们在本文中提出的方法正是围绕这些约束进行设计的。我们已将其改造并集成到 SimRa 应用 [3] 中，但它同样可以轻松作为插件集成到 Strava 或 BikeCitizens 等应用中。作为一项次要约束，我们尝试以能够处理现有数据集 [5], [6], [7]的方式设计该方法，并且成功实现了这一点。

III. 背景与相关工作

如上所述，我们提出了一种基于智能手机的众包感知方法，利用加速度计传感器量化路面质量。本节概述相关工作中侧重于路面质量测量的方法（§III-A），以及其他领域中的类似方法（§III-B）。

A. 路面质量测量

路面质量通常根据国际粗糙度指数（IRI）进行量化，例如 [8]。传统上，这项工作使用所谓的剖面仪（外形大致类似带轮子的大型梯子）完成，这些设备由汽车牵引或人工推动。20世纪80年代的一项研究 [9] 发现，这些设备经常发生故障，而且在狭窄的街道上难以操控，因此不适合用于测量自行车道的质量。此外，这类测量非常耗费人力，因此不现实，无法大规模开展。

Taniguchi等人 [10] 检测碎片、坑洼或凸起等道路危险。为此，他们将超声波距离传感器安装在自行车车把上，扫描自行车前方的地面。虽然这种方法可以实时提醒骑行者前方道路上即将出现的凸起，但由于所需传感器数量实在太多，这种方法无法扩展到全市范围分析。

Peng等人 [11]遵循相同的目标，使用运动传感器而非距离测量。数据通过分类器进行离线分析，该分类器可以识别沥青、卵石和非常颠簸的路面。与我们的工作相比，他们更注重识别特定的路面类型 而不是

路面 质量。此外，他们的方法再次需要专用硬件。

Zhou 等人 [12] 尝试检测井盖。为此，他们使用卷积神经网络分析由安装在自行车车把上的智能手机摄像头拍摄的视频流。虽然他们的方法在硬件方面与我们相似，但持续录制视频意味着功耗较高。此外，手机需要以一个不便的角度安装，因此不适合日常使用。

Luedemann 等人 [4]使用智能手机应用记录加速度计数据，作为路面质量的指标。然而，他们的方法依赖单次骑行记录，没有汇总多次骑行记录数据的概念。

与我们类似，Yamaguchi 等人 [13] 希望测量路面质量。为了获得高精度的测量结果，他们将智能手机的运动传感器与自行车里程计相结合。这种方法始终能够提供比我们的方法更精确的结果，但同样需要专用硬件，因此无法在整个城市范围内实际应用。

除此之外，还有几种以汽车为中心的方法，它们要么使用专用硬件，例如 [14],；要么对手机的放置位置有非常具体的要求，例如 [15], ；要么专注于检测瞬态特征，即单个坑洞，例如 [16],[17], [18]。

B. 边缘端辅助的数据分析

在之前的工作中 [3], [19], [20], ，我们介绍了 SimRa 项目，该项目利用市售智能手机的内置运动传感器，追踪自行车交通中的险些相撞事件。骑行期间，SimRa 以 1/3Hz 记录 GPS轨迹，并以 50Hz 记录运动传感器数据。为节省带宽和存储空间，手机会对运动传感器数据进行降采样：计算窗口大小为 30 的移动平均值，然后仅保留每第五个值。骑行结束后，用户可以在上传前裁剪并标注记录的路线。本文基于该项目，使用相同的数据测量方式，将我们的数据处理流水线集成到现有的 SimRa 应用中。

其他边缘端辅助的数据分析工作包括：Mei 等人 [21], 基于智能手机摄像头和众包感知测量紫外线辐射；Cao 等人 [22] 使用运动传感器检测患者跌倒在地时的中风；以及 Pham 等人 [23] 通过在停车位配备 RFID 芯片来追踪其占用状态，从而实现智能停车系统。此外，还有多篇出版物探讨了将边缘到云端数据处理流水线的不同组件部署在

各种用例，例如 [24], [25], [26]。

IV. 数据处理流水线

本节首先概述我们用于推导路面质量的数据处理流水线（§IV-A），然后介绍其中的各个处理步骤（§IV-B 至 §IV-E）。

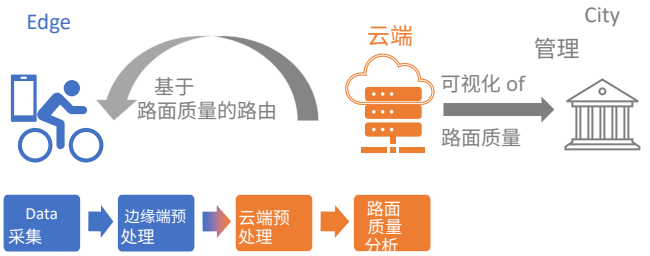

图 2：我们的路面质量分析流水线概览。

A. 概览

我们的数据处理流水线中的任务分布在边缘端（即智能手机上）和云端，另请参见图 2。使用 SimRa 应用的内置运动传感器采集数据后，会先在本地对数据进行预处理，然后上传至我们的后端服务器。在后端服务器上，还会执行其他预处理步骤，之后才进行实际的路面质量分析。

预处理过程中用于减少数据量的部分需要在边缘端执行，以保护用户隐私并减少带宽消耗。数据清洗部分的计算开销过大，因此在云端执行，以降低手机的功耗。

B. 数据采集

这些数据基于众包，由公民科学家自愿使用 SimRa 应用 [3], [20] 收集。虽然 SimRa 最初是为了检测自行车交通中的险些碰撞热点 [19], 而开发的，但它提供的骑行数据和险些碰撞事件数据也可用于其他目的，例如在交通仿真中创建自行车模型 [27], [28] ，或者像我们在本文中所做的那样，用于分析自行车道的路面质量。由于我们早期就规划了路面质量分析这一二次用途，因此 SimRa 还会收集详细的加速度计数据，如第 III-B 节所述。虽然按照 SimRa 的方式采用众包感知可以轻松扩展到大量用户，从而产生海量数据，但这种方式往往会产生噪声数据，因此在实际数据分析之前需要进行预处理。

C. 边缘端预处理

为了减少记录的数据量并保护用户隐私，我们在手机上执行两项预处理任务。首先，在骑行过程中，我们对加速度计数据进行降采样。原始数据的采样频率为 50 Hz，我们先计算长度为 30 的移动平均值，然后每隔五个值提取一个值。通过这种方式，我们旨在尽可能捕捉更多峰值，同时将数据量控制在合理范围内。其次，用户可以在上传骑行记录之前对其进行裁剪，以隐藏起点或终点。

D. 云端预处理

云端预处理有两个主要目标：清理噪声数据，并计算一个代表骑行平顺程度的数值。为了实现第一个目标，我们会去掉骑行开始和结束时各十秒的数据，以消除用户上下自行车或将手机放入

口袋时产生的噪声。当然，理想情况下，这一步应当在手机上完成，但由于 SimRa 的主要用例需要这些数据，因此无法在此处完成；否则我们就必须上传两次数据。我们还会去除停驻片段，即只考虑最低速度达到 5 km/h 且持续时间至少为 1 分钟的骑行路段。这种低速通常出现在骑行者推着自行车而非骑行，或在交通信号灯前停下时。即使骑行者能够以如此缓慢的速度骑行，行程也容易出现不稳定的动作，从而产生噪声数据。这样也能去除用户骑行结束后忘记停止记录的片段。

对于第二个目标，我们通过计算 X、Y 和 Z 轴的三个移动方差（窗口大小为 10）的平均值，对加速度计数据进行归一化。这样可以保留振幅信息，但忽略方向，因为方向会受到手机放置位置的显著影响。因此，我们获得一个由单个数值组成的时间序列，用于表示每段运动中的骑行路段的路面颠簸度。

E. 路面质量分析

每位骑行者在同一条道路上都会产生不同的数据轨迹，若不进行校准，就无法准确地单独分析每条轨迹。理想情况下，我们至少会使用针对骑行者的个人档案（即按骑行者汇总的数据），但出于隐私原因，这些数据无法获取（骑行记录均已单独进行假名化 [3]）。我们的方法旨在利用数据集的规模，并运用大数定律获得可靠的结果，无需校准数据，同时尽量避免噪声对结果造成过大影响。

为此，我们首先逐条考虑骑行记录：取预处理后的骑行数据（即一条由汇总后的运动传感器读数组成且不含停顿的时间序列，以及 GPS轨迹），并计算该时间序列的百分位数（0.2、0.4、0.6、0.8 和 1）。随后，为进行归一化，我们根据各运动传感器读数所属的区间，将其替换为 1 到 5 之间的值；也就是说，若某个运动传感器读数落在区间 $(0.2; 0.4]$，则将其替换为数值 2。其基本思路是，较长的骑行记录很可能会经过路面质量截然不同的路段，因此，我们计算某一区域相对于整段骑行其余部分颠簸度的相对颠簸度。

在第二步中，我们使用所有这类归一化骑行记录的数据，并根据 GPS轨迹将它们映射到由 10m2 单元组成的网格中。因此，每个单元都有一个介于 1 到 5 之间的数值分布。我们使用所有数值的平均值作为该单元颠簸度的指标，例如用于地图着色；同时也提供其他统计指标（例如，参见评估部分中的分布函数图表）。

V. 道路表面质量信息的使用

本节将介绍如何使用此类路面质量结果。在 §V-A 中，我们将介绍如何在 SimRa 应用中实现导航功能，并将道路表面质量作为路径规划的附加参数。随后

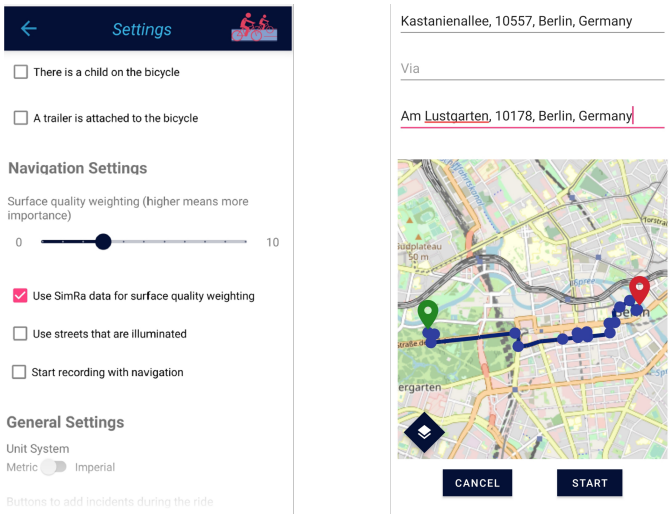

图 3：可通过设置菜单中的滑块设定路面质量在路径规划中的权重。

我们将在 §V-B 中展示如何向城市管理者提供管线的输出数据。

A. 综合路面质量的路径规划

计算出路面质量评分后，我们便可提供路径规划功能，规划时不仅考虑距离和时间，也考虑路面质量。这里的主要问题是，在计算从 A 到 B 的最佳路线时，路面质量应占多大权重，或者说可接受的绕行距离是多少。我们决定让用户通过滑块调节这一因素：滑块范围为 0 到 10，0 表示路径规划不考虑路面质量，10 表示路面质量具有最高优先级（图 3）。我们使用经过修改的 GraphHopper[1]进行路径规划。GraphHopper 使用街道边，并通过节点将这些边连接起来。每条边都有一个用于路径规划的权重，并且可以根据自定义数据调整权重。我们正是通过这种方式利用路面质量：根据路面质量和用户指定的影响因子增加权重。

B. 输出数据与可视化

我们的自行车路面质量处理流程会生成一个 GeoJSON 文件作为输出。该文件包含网格中的单元格，每个单元格的表面积为 10m2，并以 要素形式表示，几何类型为多边形，其属性键中包含路面质量评分信息，例如平均值、中位数、标准差、经过该单元格的骑行记录数量以及着色信息。有了这样的输出文件，就能轻松创建简单的可视化效果[2],例如，使用 Leaflet[3]，如图4所示。借助这类可视化，即使是不熟悉技术的用户，也能轻松监测大范围区域的自行车路面质量，并在需要时采取行动。

[1]https://github.com/graphhopper/ [2] https://simra-project.github.io/surfaceQuality/Berlin.html [3] https://leafletjs.com/

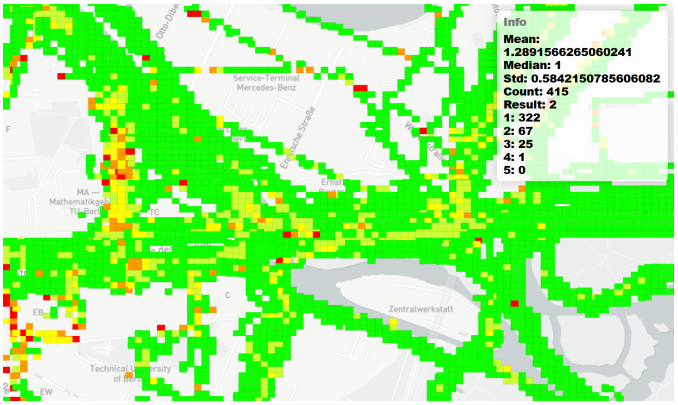

图4：输出文件的可视化效果。将鼠标悬停在方框上时，显示其路面质量。颜色编码基于平均值。

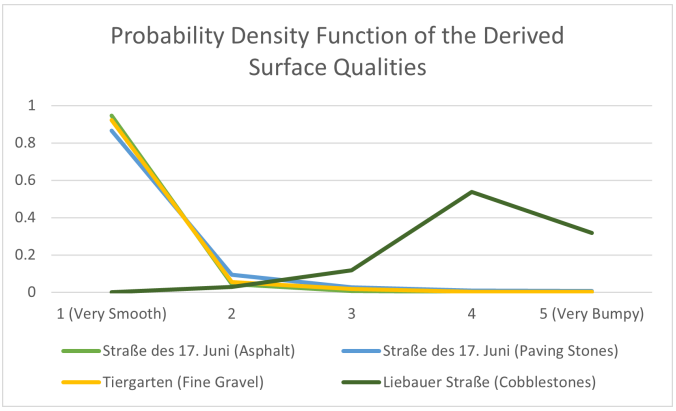

图5：推导出的路面质量的概率密度函数显示，这些路段的道路表面质量值非常明确，因为它们要么非常好，要么非常差。

VI. 评估

在本节中，我们通过比较柏林各地选定街道路段的计算路面质量与其在现实世界中的路面类型，来评估路面质量分析方法。其中，现实世界中的路面类型数据来自 OpenStreetMap（OSM）。作为数据输入，我们使用现有的 SimRa 数据集 [5], [6], [7]（包含近 90,000 条骑行记录，总里程超过 650,000 公里）。基于不同路面类型与路面质量之间也存在一定相关性的直觉，我们随机选取了四处具有不同路面类型的地点。为了展示我们方法的局限性，随后我们进一步探索了该数据集，并手动选取了另外四处地点，在这些地点中，我们的方法似乎得出了错误的结果。每个评估路段的路面面积均为 10 m2，为了便于相互比较，我们分析了它们的平均值、中位数和标准差。

我们首先描述第一组中的明确结果（§VI-A）。随后，我们将另外四个“问题点”归类为结果混杂（§VI-B）或看似错误的结果（§VI-C）。

A. 结果明确的路段

我们针对以下每种路面类型至少评估一个示例，这些类型按照预期路面质量 [29] 从高到低排序：沥青、平整铺路石、细碎石、鹅卵石路面。我们选择这些路段时遵循两个原则：(i) 确保这些路段拥有充足的数据；(ii) 涵盖所有不同的路面类型。根据这些标准筛选后，我们随机选取了四个路段。

表I和图5展示了所选路段的路面质量分析结果，结果非常清晰直观。

可以观察到，上述按照路面质量从高到低排列的表面类型列表，其排序是正确的。位于六月十七日大街 的一段新近养护的沥青路面得分接近满分，这意味着在几乎所有经过该路段的骑行记录中，其路面质量均位于前20%。其次是两段路面，其表面类型分别为平整铺路石和细碎石，】【。

表I：路面质量分析评估结果，显示具有明确结果的路段的平均值、中位数和标准差

街道名称	路面	GPS位置	Mean	中位数	标准差
六月十七日大街	沥青	52.515369,13.630855	1.03	1	0.17
六月十七日大街	铺路石	52.513501,13.335127	1.25	1	0.5
蒂尔加滕	细碎石	52.514745,13.34622	1.32	1	1.06
Liebauer Straße	鹅卵石路面	52.509132,13.453806	4.15	4	1.07

结果非常相似。这意味着，从骑行者的角度来看，平整铺路石和细碎石路面的颠簸度相近。不过需要注意的是，在降水量较大的地区，细碎石可能不太适宜（详见§VII）。*Liebauer Straße* 的路面质量得分非常差，这并不令人意外，因为该路段铺有鹅卵石路面，人行道上餐馆和咖啡馆林立且十分繁忙，使骑行者无法在此（非法）骑行，只能骑行在机动车道上。

B. 结果混杂的路段

虽然大多数结果都符合直觉，例如，采用沥青铺设并与机动车交通隔离的新建自行车道，其路面状况非常好；但另一些路段（见图6）乍看之下似乎令人费解。

根据*Invalidenstraße* 和*Kaiserin-Augusta-Allee*各街道路段的概率密度函数（PDF），每个路段的路面质量似乎都截然不同。进一步观察具体路段，可以发现其中的原因：

表II：路面质量分析评估结果，显示结果混杂的路段的平均值、中位数和标准差。

街道名称	路面	GPS位置	Mean	中位数	标准差
Invalidenstraße	沥青	52.526236,13.369196	2.24	3	0.96
Kaiserin-Augusta-Allee	沥青	52.524429,13.327253	2.17	2	1.05

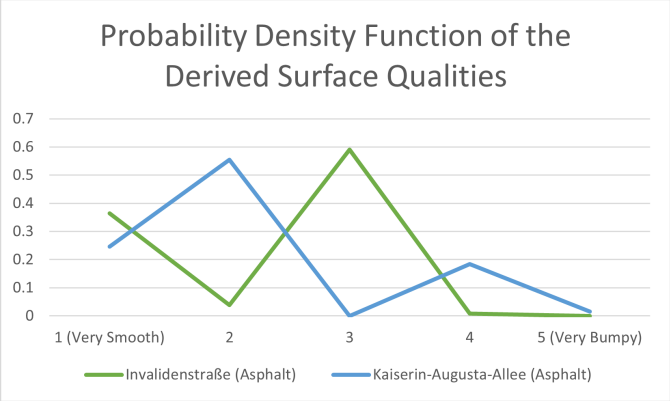

图6：推导出的路面质量的概率密度函数显示，这些路段的路面质量值令人困惑：视骑行情况而定，其路面质量看起来可能非常好，也可能非常差（呈现多个峰值）。



图7：在一小片区域内，人行道上的公交专用道、公交车站和自行车道可能会产生不同的结果。（来源：苹果地图）

在*Invalidenstraße*，设有一条公交专用道、一个公交站和一条路缘式自行车道（见图 7）。柏林的大多数公交专用道在法律上允许骑行者使用，但有些骑行者可能仍更喜欢相对颠簸的自行车道。此外，公交车停靠在公交站时，使用公交专用道的骑行者可能会从左侧非常平坦的汽车车道超越公交车。这两个因素很可能正是该处记录呈现双峰分布的原因。

进一步仔细查看<风格 id='1'>凯瑟琳-奥古斯塔大街</风格>的险些碰撞事件数据表明，人们在一个方向使用自行车道，而在另一个方向使用街道 [5], [6], [7] ，因为一侧似乎处于无法使用的状态，而另一侧则尚可接受。由于自行车道铺设的是铺路石，而街道铺设的是沥青，这会导致道路的路面质量因骑行方向不同而有所差异。不过，两个方向最终都会定期落入同一个 10x10m 方格中，因为街道并不算太宽，尤其是在考虑 GPS 精度的情况下。

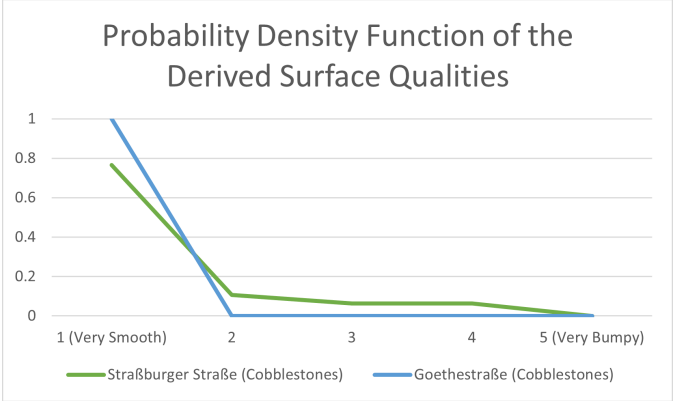

图 8：推导出的路面质量的概率密度函数显示，尽管这些路段铺有鹅卵石路面，但路面非常平整。

表 III：路面质量分析评估结果，展示（看似）结果令人困惑的路段的平均值、中位数和标准差

街道名称	路面	GPS位置	Mean	中位数	标准差
施特拉斯堡大街	鹅卵石	52.532273,13.416521	1.43	1	0.87
歌德街	鹅卵石	52.508889,13.308333	1	1	0

C. 结果（看似）令人困惑的章节

此外，与 OSM 的表面类型标签相比，也有一些路段的路面质量评估结果似乎完全错误。表 III 展示了两个路段，根据 OSM 的数据，这两个路段均铺设了鹅卵石路面，且没有独立的自行车道，但评估结果却非常好，而图 8 中展示的 PDF 文件也证实了这一点。

人们可能会想到，这一部分的标签有误，实际上并非铺有鹅卵石路面。然而，正如图 9 所示，这两条街道确实铺有鹅卵石路面。

图 9 还展示了这些路段的自行车骑行记录为何非常平稳的可能原因。基础设施状况促使骑行者更愿意选择人行道，而非实际的街道。首先，街道上的鹅卵石让骑行者骑行时感到非常不适。其次，人行道宽敞、安静，铺有表面平坦的大块铺路石。第三，街道上停放着车辆，或有汽车四处寻找停车位，还有服务车辆、施工车辆等，这些车辆往往会挡住骑行者，也会造成安全隐患。考虑到这些情况，看到这两条街道及其他类似街道的路面质量数值较好也就不足为奇了，因为骑行者很可能会（违规）改走人行道。事实上，苹果地图上的照片（本文未收录）确实显示骑行者在使用人行道。

七、讨论

总体而言，本文所呈现的结果表明，我们的方法能够利用数据推导出路面质量信息



(a) 施特拉斯堡街



(b) 歌德街

图 9：柏林的施特拉斯堡街和歌德街铺设着大型鹅卵石，路上还有停车或寻找车位的汽车，彼此争抢道路空间。相比之下，人行道铺设着表面平整的铺路石，由于附近没有商店、餐馆或咖啡馆，因此十分安静。（来源：苹果地图）

这些数据记录在 SimRa 应用中。不过，该方法仍存在一些局限性，我们将在本节中讨论。

A. 方法论挑战

由于 SimRa 数据集由智能手机生成的众包数据构成，记录的数据具有很强的异质性。这是因为不同型号的智能手机配备了不同的 GPS 模块和运动传感器，导致同一条道路在一部手机看来可能非常平坦，在另一部手机看来却可能十分颠簸。自行车类型多种多样（例如公路赛车或山地车），使这一问题更加突出；不同车型也会显著影响智能手机能够感知到的振动。为解决这些问题，我们依靠大数定律，并比较某一路段的路面质量相对于骑行过程中的整体情况如何。

B. SimRa 中的预处理

作为路面质量分析流程的输入，我们使用了 SimRa 数据集，因为据我们所知，它是唯一包含大量匿名化个人骑行记录且未经过聚合处理的公开数据集。不过需要注意的是，SimRa 的开发目的是识别自行车交通中的近失事件热点。为此，系统采用相对较低的 50 Hz 加速度计采样率，并通过窗口大小为 30 的移动平均进一步降低细节层级，随后每隔五个值取一个。虽然这样可以自动检测近失事件，但更高的数据点分辨率将使我们能够开发出更复杂的测量方法。不过，这会进一步增加 SimRa 应用的磁盘空间、内存、电池和带宽占用；而且如第 VI 节所示，我们得出的路面质量对于预期用例而言已足够精确。

C. 骑行者的行为

由于 GPS 存在误差，无法确定骑行者是在实际道路上行驶，还是在人行道上行驶。例如，正如上一节所讨论的，图 9 显示施特拉斯堡大街 位于柏林，按预期其结果应当较差，但实际上数值却出人意料地好。我们认为，这是因为骑行者非法使用平坦的人行道，而没有使用颠簸的道路。

此外，骑行者在实际骑行时也会尽可能避开最严重的坑洼及类似的凸起，因此这些路况通常不会出现在我们的分析中。

D. 传感器精度不足

GPS传感器 [30] 的精度不足，加上智能手机运动传感器产生的噪声，构成了数据集的另一项局限。我们尝试通过预处理步骤部分解决这些局限，但当然无法完全消除。唯一的替代方案是使用专用硬件，但由于存在采用门槛，记录的骑行记录将显著减少。

E. 时间影响

SimRa 数据集中的骑行记录最早可追溯至 2019 年，路面质量可能随时间发生变化。其原因之一可能是表面类型发生变化，例如从鹅卵石路面改为沥青路面，或者坑洼和裂缝得到修补。这很可能会使变化后的结果更好，但在我们的数据集中会呈现为双峰分布。因此，在实际使用这种方法时，我们建议只考虑最新的骑行记录。

另一种时间因素可能是天气。在下雨、刮风或下雪的日子里，道路的路面质量可能会显著下降，尤其是潮湿的碎石路面，而我们的路面质量分析方法并未考虑这一点。不过，将这一因素纳入分析会降低结果的有效性，因为这会减少数据集中纳入考量的骑行记录数量。

VIII. 结论

世界各地的城市都希望提高自行车交通的出行方式分担率，例如解决排放问题和频繁发生的交通拥堵，同时通过增加日常活动改善市民健康。除安全性外，一个关键影响因素

是舒适度，尤其体现在自行车基础设施的路面质量上。然而，由于这类基础设施规模庞大，人工监测其路面质量并不可行。

本文提出了一种众包感知方法，骑行者通过智能手机应用和手机的内置运动传感器记录日常骑行记录。我们提出了一套数据处理流水线，从边缘端（即手机）开始，最终在云端后端结束。此外，我们展示了该众包感知方法确实能够推导出路面质量，并实现了两个使用此类数据的用例。

参考文献

[1] P. Nyberg、U. Bj¨ornstig 和 L.-O. Bygren，“道路特征与自行车事故”，斯堪的纳维亚社会医学杂志，第 24 卷，第 4 期，第 293–301 页，1996 年。[2] A. Gadsby、J. Tsai 和 K. Watkins，“通过问卷调查和眼动追踪了解路面状况对骑行者安全感与舒适度的影响”，交通研究记录，第 2676 卷，第 12 期，第 112–126 页，2022 年。[3] A.-S. Karakaya、J. Hasenburg 和 D. Bermbach，“SimRa：通过众包识别自行车交通中的险些碰撞热点”，*Elsevier* 普适与移动计算，第 67 卷，文章 101197，2020 年 9 月。[4] K. Luedemann 和 M. A. Nascimento，“Bikevibes：一款从骑行者视角众包开放道路质量数据的应用”，载于第 *15* 届 *ACM SIGSPATIAL* 计算交通科学国际研讨会论文集，IWCTS ’22 系列。美国纽约：计算机协会，2022 年。 [在线]。网址：https://doi.org/10 1145 3557991 3567779 。/。D. [5] 06 19 12 Bermbach，“SimRa 柏林骑行记录 / - /20”。柏林工业大学，2021 年。[6] D. Bermbach 和 A.-S. Karakaya，“SimRa 柏林骑行记录 01/21 - 09/21”。柏林工业大学，2021 年。[7] A.-S. Karakaya 和 D. Bermbach，“SimRa 柏林骑行记录 10/21 - 09/22”。柏林工业大学，2022 年。[8] M. W. Sayers，“国际道路粗糙度试验（IRRE）：建立测量结果的相关性与校准标准”，1986 年。[9] S. L. Cumbaa，“道路剖面研究：最终报告”，技术报告，1986 年。[10] Y. Taniguchi、K. Nishii 和 H. Hisamatsu，“用于监测路面状况的自行车安装式超声波距离传感器评估”，载于*2015* 年第 *7* 届计算智能、通信系统与网络国际会议，2015 年，第 31–34 页。[11] D. Peng、Z. Strout、S. Jiang 和 P. Shull，“一种通过锁内嵌 IMU 对无桩共享单车进行路况分类的方法”，载于工业控制网络与系统工程研究国际会议论文集。美国纽约：计算机协会，2019 年，第 32–36 页。[12] B. Zhou、W. Zhao、W. Guo、L. Li、D. Zhang、Q. Mao 和 Q. Li，“基于智能手机的道路井盖检测与分类”，建筑自动化，第 140 卷，文章 104344，2022 年。[13] T. Yamaguchi、T. Nagayama 和 D. Su，“利用自行车的动态响应对自行车道状况进行简易估算”，载于第 *6* 届实验结构工程进展国际会议暨第 *11* 届先进智能材料与智能结构技术国际研讨会（*6AESE/11ANCRiSST*），2015 年 8 月。[14] K. Chen、M. Lu、G. Tan 和 J. Wu，“CRSM：基于众包的路面监测”，载于*2013* 年 *IEEE* 第 *10* 届高性能计算与通信国际会议暨 *2013* 年 *IEEE* 嵌入式与泛在计算国际会议。IEEE，2013 年，第 2151–2158 页。[15] Y.-A. Daraghmi、T.-H. Wu 和 T.-U. ˙Ik，“基于众包的路面评估与索引”，*IEEE* 智能交通系统汇刊，第 23 卷，第 5 期，第 4164–4175 页，2020 年。[16] M. Y. Alam、A. Nandi、A. Kumar、S. Saha、M. Saha、S. Nandi 和 S. Chakraborty，“真正来自大众的众包：利用智能手机传感器实现不依赖设备、车辆、路面及驾驶行为的道路剖面分析”，普适与移动计算，第 61 卷，文章 101103，2020 年。

[17] X. Li、R. Chen 和 T. Chu，“利用智能手机检测路面粗糙度的众包解决方案”，载于第 *27* 届导航研究所卫星部门国际技术会议论文集（*ION GNSS+ 2014*），2014 年，第 498–502 页。[18] L. C. Lima、V. J. P. Amorim、I. M. Pereira、F. N. Ribeiro 和 R. A. R. Oliveira，“运用众包技术和移动设备识别沥青路面质量”，载于*2016* 年第六届巴西计算机系统工程研讨会（*SBESC*）。IEEE，2016 年，第 144–149 页。[19] A.-S. Karakaya、T. Ritter、F. Biessmann 和 D. Bermbach，“CycleSense：利用移动运动传感器检测自行车交通中的近失事件”，*Elsevier* 普适与移动计算，2023 年。[20] M. Temmen、J. Eckart、J. Merk、A.-S. Karakaya 和 D. Bermbach，“自行车交通研究中的众包”，国际交通，第 74 卷，第 1 期，第 66–69 页，2022 年 1 月。[21] B. Mei、R. Li、W. Cheng、J. Yu 和 X. Cheng，“利用智能设备测量紫外线辐射”，*IEEE* 物联网期刊，第 4 卷，第 4 期，第 934–944 页，2017 年。[22] Y. Cao、S. Chen、P. Hou 和 D. Brown，“FAST：借助雾计算的分布式分析系统，用于监测跌倒以缓解卒中”，*2015* 年 *IEEE* 网络、架构与存储国际会议（*NAS*），第 2–11 页，2015 年。[23] T. N. Pham、M.-F. Tsai、D. B. Nguyen、C.-R. Dow 和 D.-J. Deng，“基于物联网技术的云端智能停车系统”，*IEEE Access*，第 3 卷，第 1581–1591 页，2015 年。[24] I. Lujic、V. D. Maio、K. Pollhammer、I. Bodrozic、J. Lasic 和 I. Brandic，“通过实时边缘分析和 5G 提升交通安全”，载于第 *4* 届边缘系统、分析与联网国际研讨会论文集。ACM，2021 年。[25] T. Pfandzelter、J. Hasenburg 和 D. Bermbach，“从零到雾：高效构建基于雾计算的物联网应用”，软件：实践与经验，第 51 卷，第 8 期，第 1798–1821 页，2021 年 8 月。[26] G. Hattab、S. Ucar、T. Higuchi、O. Altintas、F. Dressler 和 D. Cabric，“车载微云中计算任务的优化分配”，载于第 *2* 届边缘系统、分析与联网国际研讨会论文集。ACM，2019 年。[27] A.-S. Karakaya、K. K¨ohler、J. Heinovski、F. Dressler 和 D. Bermbach，“基于 SimRa 数据集的逼真骑行者模型（用于 SUMO）”，载于第 *20* 届 *IEEE* 地中海通信与计算机网络会议（*MedComNet 2022*）。塞浦路斯帕福斯：IEEE，2022 年 6 月，第 166–173 页。[28] A.-S. Karakaya、I.-A. Stef、K. K¨ohler、J. Heinovski、F. Dressler 和 D. Bermbach，“利用 SimRa 数据集在 SUMO 中实现逼真的骑行者行为”，*Elsevier* 计算机通信，2023 年。[29] W. Titov 和 T. Schlegel，自行车路面状况监测——利用众包获取的移动设备传感器数据，2019 年 6 月，第 340–356 页。[30] K. Merry 和 P. Bettinger，“城市环境中的智能手机 GPS 精度研究”，*PLOS ONE*，第 14 卷，第 7 期，第 1–19 页，2019 年 7 月。