城市语义计算机视觉：理解城市空间中人们的情境理解框架

Anthony Vanky，博士，通讯作者
哥伦比亚大学建筑、规划与保护研究生院，纽约州纽约市，
邮编 10025 电子邮箱：a.p.vanky@columbia.edu

Ri Le
哥伦比亚大学建筑、规划与保护研究生院，纽约州纽
约市 10025 邮箱：r.le@columbia.edu

关键词

人工智能、计算机视觉、城市空间、城市语境、城市主义、语义意义、深描、评估

摘要

计算能力的提升和深度学习方法的改进，使计算机视觉技术在城市环境中得到广泛应用。它们在警务、交通管理和记录公共空间等方面的应用也日益普遍。尽管算法训练中的偏差及其收益分配不均等问题常被讨论，但几乎所有应用都以同样的方式，将城市体验简化为简单、片面且机械的衡量指标。这些实践缺乏语境、深度和特殊性，因而无法在城市语境中，尤其是在使用和占据城市空间的语境下，形成语义知识或开展分析。本文将批判人工智能和计算机视觉在城市实践中的既有应用，并提出一个用于理解人、行动与公共空间的新框架。

本文重新审视格尔茨如何运用深描来构建关于文化与活动的阐释理论，并以此为视角，建立一个评估计算机视觉技术各种应用的框架，重点考量这些应用所承载的意义。我们将讨论该框架的定位在不同技术使用者之间可能存在的差异与冲突。本文还将探讨当前深度学习算法的使用与训练，以及这一过程如何限制语义学习；同时提出三种可能的方法，以获得对城市空间更具语境特殊性、具有城市语义的描述，从而满足城市主义者的需求。

本文通过突出人工智能技术在城市环境中的现有应用在这一语境下的失效，同时提出对这些算法进行演进的设想，使其最终能够适应并服务于这一空间与文化环境，从而为围绕人工智能日益普及所展开的批判性讨论作出贡献。

1. 引言

计算能力日益强大，通信基础设施成本不断下降，传感器的尺寸也日趋小巧，这些因素使技术得以广泛嵌入城市空间的肌理之中，催生了“智慧城市”运动。对于所谓智慧城市运动中的许多人而言，这一趋势是将城市工具化，以实现更高效率，并将城市生活的诸多方面问题化（Hollands, 2008）。在这一领域工作的技术专家，正把枚举应用于从出行和基础设施到公共安全和民主参与的所有领域（Eagle & Pentland, 2009; Goldsmith & Crawford, 2014; Jiang et al., 2013）。然而，城市空间的定义性社会特征无法被简化为单纯的优化问题。

这些数字技术及其产生的模型和数据成果，能够塑造我们对建成环境的认知。与其他算法过程中普遍存在的偏差所引发、广为人知的挑战并无二致（Crawford，2018；Kirchner 等，2016），黑箱方法及其不透明的结果同样可能不当地影响我们对所居住地方的解读（Schwarzer，2017）。尽管还原主义与复杂性之间存在这种冲突，但迫切需要通过新模型和工具来理解相关问题；这或许能为研究公共空间和城市形态开辟新路径，以应对快速城市化和城市空间私有化加剧的局面（E. Talen 与 Ellis，2002）。

深度学习和计算机视觉应用日益普及，仿佛是现代主义将摄影用于城市规划的延续，为通过图像理解城市创造了新的机遇（Lecun 等，2015）。尽管人们乐观地认为，这些新兴技术能够提供更精确（或许也更广泛）的方法来理解城市，但仍有一些问题尚待解答。

在这一计算语境中，本文特别关注这些算法工具在城市规划与管理中的应用——这一作用虽处于萌芽阶段，却正在不断扩大。一方面，它们能够量化城市空间中的人类行为，使决策者得以据此制定更加有据可依的政策。另一方面，将这种数字还原主义应用于城市空间，却忽视了城市的特殊性及其本质特征，而正是这些特征使城市成为独特的栖居之地。因此，本文认为，除了情境化技术所具有的还原主义倾向之外，还需要采取一种独特的方法，以理解人们如何栖居并使用这些城市公共空间。此外，随着计算机视觉和基于图像的方法日益广泛地应用于城市的工具化，有必要在城市技术的训练、实施与采用方面建立一套城市专属词汇。

尽管计算机视觉技术已应用于城市管理的诸多方面，例如基础设施利用和公共安全，但本文明确探讨如何运用这些技术来理解人们如何栖居并使用公共空间。本文认为，基于图像的人工智能是现代主义运用摄影图像采集城市数据这一做法的自然延伸。本文还简要回顾了人工智能如何应用于基于图像的数据，以说明在构建关于空间使用者及其所栖居空间的、更为丰富且专门化的本体时可能存在的弱点。此外，本文探讨了公共部门和私营部门主体如何推销薄数据，以及这可能如何损害以人为本的空间规划。不过，本文也探讨了如何重新构想当前所用方法，以克服这些局限，运用基于算法的方法对公共空间进行城市语义描述。

2.“城市主义”与语境问题

一个根本性的挑战在于，如何界定某种空间或实践具备哪些特征才算得上城市的，尤其是在当代城市建

具有城市的特质，而哈德逊城市广场却没有，尽管它们在曼哈顿相距不过几公里？又是什么让首尔的

从本质上说，城市空间是个人经验与陌生人相遇交汇之处，尽管他们很少与我们共享相同的价值观、历史

城市规划中，Whitaker Bill（1985）是什么是？居民与我们共享公共空间并创造集体体验。以及他们在哪些地方共享这些空间？居民能够影响看这些社会动态，城市空间的形态与布局构建了其中人们交互的框架。归根结底，公共生活唯有在这些空间公地——街道及城市公共空间——中才能获得独特体验，因为这些空间使居民深刻意识到彼此相互依存，以及这种依存关系所引发的社会义务（Chidster，1989）。从某种意义上说，城市是个人主义交融并创造集体体验的地方。

界定这些社会空间关系是一项艰巨挑战，而对于何为良好的公共领域，人们也有不同看法。关于这些空间的特征，存在着彼此分歧却并行不悖的理论；这些空间占据了实体城市的很大一部分，包括街道、人行道、广场、拱廊、非机动交通、绿道等。简·雅各布斯（1961）阐述了街道作为公共空间的重要性：

> “街道几乎总是属于公众：由公众所有；而当我们谈论公共领域时，很大程度上说的就是街道。如果我们能够将街道规划和设计成令人愉悦、充实的场所，成为营造社区的空间，成为城市和街区中吸引所有人的公共场所，那么我们就成功地直接设计了约三分之一的城市，也将对其余部分产生巨大影响。”

正如她所说，人行道是“城市的主要公共场所”，也是“城市最重要的器官”。同样，扬·盖尔（1987）也强调了建筑之间的生活。 正是在建筑之间的空间里，居民的社会联系得以建立和巩固，流动空间也与城市的社会生活并存。这些方面构成相互关联的流动与活动模式，而建筑则对此作出回应。对凯文·林奇（1960）而言，这种生活也赋予这些空间意义，使它们成为“具有意象性的”空间，其空间特质能够强化居民对城市空间的关注，并赋予其意义。由于彼此邻近，或因空间本身的几何构型，这些空间在建筑、路径、特色元素和便利设施之间建立起令人难忘的联系。

归根结底，这些概念旨在影响场所营造实践，或塑造能够支持前述公众活动的城市空间。设计师往往没有意识到自身的偏差及其根源会影响决策（A. Jacobs & Appleyard, 1987）。重新探求“良好城市形态”的构成要素，将有助于城市规划这一专业的发展；在这一探求中，实体城市的规划应以人和场所 为中心（Emily Talen & Ellis, 2015）。对于设计师和规划师而言，时间、预算等规范性限制常常决定着改善城市空间的机会。然而，由于缺乏对使用者和非使用者如何感知这些空间，以及使用者特征的了解，这些限制往往会进一步加剧（Chidster, 1989）。要为此类工作奠定坚实基础，就必须应对美学、伦理学和政治理论的复杂问题，因此也需要建立特定领域的本体，以适应城市主义的社会与实体环境。

3. 作为城市数据的影像

图像在证明公共空间中的人类活动方面发挥了巨大作用。当代社会中，摄像头已无处不在，并被用作记录社会弊病与不公正的工具。突尼斯蔬菜摊贩穆罕默德·布瓦吉吉抗议时自焚的视频（Noueihed, 2011），以及乔治·弗洛伊德遭警察强力压制时喘息着说“我无法呼吸”的视频（J. Collins, 2020），都成为抗议者要求司法公正与变革的证据。然而，将摄影影像用作数据来源，以学习和探索城市空间，其历史几乎与这项技术本身一样久远。

路易·达盖尔的《坦普尔大道景观》创作于1938年，是最早拍摄到繁忙街道上行人的照片。尽管画面中有两个男人站在街角，但他们之所以入镜，只是因为其中一人正在为另一人擦鞋，因此两人都必须保持静止。长时间曝光导致移动的车辆和其他行人都未能留下影像。19世纪60年代，初露头角的城市摄影艺术成为推动社会与城市改革的政治工具。在巴黎，Charles Meville受委托拍摄乔治-欧仁·奥斯曼男爵的现代化建设，记录了城市肮脏、拥挤的“改造前”景象，以及宏伟、凯歌式的“改造后”片段。Melville的照片一直是相关争论的素材：它们是否不仅作为城市留存昔日巴黎的档案，还充当了为夷平许多街区辩护的宣传工具（Dahlberg，2015）。历史学家 Shelley Rice（1997）批评Meville反复拍摄“临床式”的街景，指出摄影师对城市的宏大变革抱持“被动”态度。在大西洋彼岸，雅各布·里斯担任《纽约论坛报》警务记者期间，记录了曼哈顿下东区的贫民窟状况。里斯通过幻灯片放映、配图文章和印刷书籍，利用摄影影像作为住房与社会改革必要性的证据，促使有政治影响力的外邦人和决策者采取行动并推动改革。

摄像头曾一度鲜为人见；到了二十世纪，它们日益普及，并在现代主义时期成为记录人类行为的工具，能够以不眨眼般的实证主义进行观察。随着从飞机和卫星上拍摄照片变得更加容易，规划师和建筑师开始青睐飞行所带来的全新平面视角，将其用作记录城市形态发展的方式。对勒·柯布西耶（1935）而言，飞机的视角提供了一份“无情”、客观的现实记录；而在战后时期，美国和欧洲的许多规划师则将航拍照片作为城市衰败的证据。这种记录让人们有机会以全新的方式看待世界，同时也印证了重新规划的必要性（Hinchcliffe，2010）。在其开创性著作《向拉斯维加斯学习》中，文丘里与斯科特·布朗（1972）尝试运用多种视觉媒介，探索如何表征20世纪60年代末的美国城市形态。作为对拉斯维加斯汽车导向特征及其标志性表现力的批判性探讨，他们的耶鲁研究团队通过摄影和电影来探索这座城市，并将二者同时用作数据收集和表征的手段。值得注意的是，他们的后现代主义方法采用了在行驶中的汽车内拍摄的“驾车摄影”——这种技法借鉴自艺术家埃德·鲁沙——以实现语境化表征。

总而言之，作为数据的图像也会产生人们所理解的意义。在研究城市主义时，我们对城市及其居民的了解，既受到媒介特性与局限性的中介，也受到生产方式的塑造（Azar 等，2021）。历史先例依赖于对框定与表征的策划，而新兴技术对人类的依赖日益减弱——图像越来越多地由机器为机器生成（Paglan, 2016）。这一计算过程改变了城市知识的生成与理解方式，无论这些知识是通过枚举、数据还是模型产生的；与此同时，这一过程甚至可能完全不为人类所察觉。

4. 基于图像的分析与“智慧城市”

从马匹到汽车，再到电话，作为一种社会产物，实体城市长期以来一直随着当时盛行的技术而演变。如今，所谓的“智慧城市”兴起，引入了两种并行的技术干预：一是拥有移动电话等联网设备、实现数字互联的市民群体；二是通过“物联网”对城市这一实体构造进行无处不在的设备化。两者所促成的泛在感知，使人们能够更迅速地动态感知、分析和理解个体的移动轨迹，并随着时间推移积累详细知识，以发现模式与趋势。

这种“大数据”方法——通过获取大容量数据集来研究现象及其动态变化——拓展了城市空间的设计、开发和评估流程，并为建成环境的数据驱动分析与设计提供了机会。麦克卢汉预见到，技术将充当城市生活的恒温器，“以各种方式塑造生活，从而优化人类的意识”（Norden，1969）。他说：“如今，在技术上已经可以利用计算机以有益的方式对社会进行编程。”他强调：“社会的编程实际上完全可以以建设性且人道的方式进行。”Greenfield（2013）指出，“这一切……的最终目的是让城市中每一个展开的过程都清晰可见[……], 使此前晦暗不明或难以确定的事物不仅变得可知，而且可以付诸行动。”

尽管城市的复杂性难以厘清，但理解充满活力的城市之复杂本质，既是欧几里得现代主义者的梦想，也是那些采用有机方法进行城市规划者的追求。例如，对于城市规划师简·雅各布斯而言，“变量很多，但并非杂乱无章；它们相互关联，构成一个有机整体。”雅各布斯谈论的是20世纪60年代曼哈顿的哈德逊街（J. Jacobs，1970），但如今任何城市中的任何一条街道也都可以这样描述。关键区别在于，如今的数字数据让我们能够比雅各布斯所能想象的更好地描述和分析她所说的“相互关联的变量”。大数据或许能让我们发现引人入胜的依存关系，并获得深入细致的认识。（Jemielniak，2020）

随着城市官员、城市主义者和城市规划师分析大量可用数据，图像也被用于理解更广泛的城市动态。例如，Girardin 等人（2008）利用在线照片分享平台 Flickr 上用户上传的照片，评估了西班牙游客和来自英国的国际游客在巴塞罗那留下的不同游览足迹。为此，研究人员分析了可交换图像文件（EXIF）中的底层信息，其中存储着由摄像机生成的元数据，以及用户自愿提供的信息，例如地点、用户档案人口统计信息（包括摄影者来自的国家和城市）、标签和描述。综合这些信息，有助于了解不同群体可能以不同方式使用城市（Offenhuber 等人，2013）：英国人往往只逗留在兰布拉大道等主要的知名旅游景点，而西班牙人则愿意前往更远的地方。这也再次印证了“可意象化的城市”这一概念：[1] 那些吸引摄影者注意的地点（例如纪念碑、教堂和公共空间）由此显现，而缺少图像之处则可能揭示出被视为更内敛的地点。

尽管志愿者提供的图像仍被广泛使用，街景图像却提供了一种数据集：其覆盖范围极为全面（这得益于创作公司谷歌雄厚的财力），格式也保持一致。这项工作源于联合创始人拉里·佩奇的信念：这类街道层级影像包含大量信息，可以加以整理、开放获取并挖掘利用（Anguelov 等人，2010）。到2017年，该公司已[illegible]

[1] “意象”在此指一种认知的、基于记忆的意象，如前文提及 Lynch 的研究中所用的意象。

然而，一些项目尝试利用这一数据集，评估这些空间中更为短暂或具有特定文化属性的特征。Salesses

尽管用户生成图像在速度、容量和真实性方面都令人惊叹，但对于拍摄内容[2] 和拍摄时间，普遍缺乏标准化，这限制了对特定地点进行深入探究的可能性。例如，Flickr 上有超过 200 万张加利福尼亚州“迪士尼乐园”的照片，而位于主题公园入口对面街道上的“Captain Kidd's Family Dining”却只有一张照片。此外，尽管街景图像在大规模范围内解决了标准化问题，但拍摄图像的巨大成本限制了实时了解城市整体动态特性的可能性。然而，通过部署传感器发展起来的智慧城市技术，为深入探究城市提供了机遇，既能解决标准化和容量方面的顾虑，也能让人们实时并在更长时间内了解各个地方。

4.1. 街道上的自动化之眼

随着智慧城市运动兴起，传感技术领域的大量投资集中于计算机视觉、先进机器学习（或深度学习），更具体地说，是利用卷积神经网络（CNN）对单张图像或一系列图像进行分析。这

这股热潮伴随着建成环境中监控摄像头数量的增长而来。出于各种原因，众多公共和私人机构都在其中安装监控摄像头（Lin & Purnell，2019）。据估计，到2021年年底，全球监控摄像头数量将超过10亿台，其中仅美国就有8500多万台，其人均摄像机普及率已可与中国相媲美（IHS Markit，2019）。尽管摄像头本身并不具备实现上述计算能力的整体能力，但据估计，到2025年，这些已安装的摄像头中将有3.5亿台配备支持人工智能的内置技术；当年出货的摄像头中，超过65%将至少搭载一款人工智能芯片组（ABI Research，2021年）。

2这些共同构成了大数据中常说的“四个V”：速度、真实性、容量和多样性。

为了枚举并描述居民的行为，这些智能设备采用了一类称为“动作识别”的基于视觉的机器学习技术，旨在提取可识别的人体特征、行为举止和动作。通常，所有计算机视觉方法都会根据捕获的图像和视频生成数据抽象。从而将图像[illegible]作。

除非专门针对这一任务（例如使用人脸识别技术），否则这些处理过程通常不会明确识别某个具体的人。一般而言，识别人类动作的第一步，是在图像中识别出在统计特征上类似于人物形象的特征。在这里，可以将这个抽象化的人物形象视为斑块、边界框或骨架（图1）。这些表示方式在计算资源消耗和统计强度方面各有取舍，但呈现人类的方式各不相同。

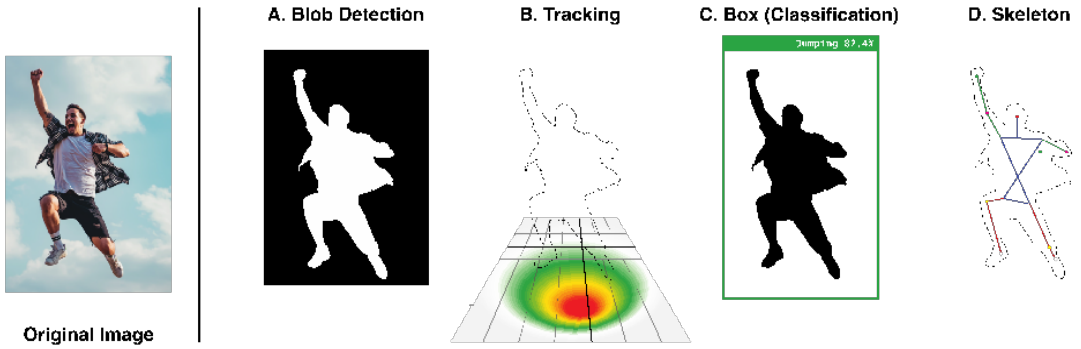


图1。 计算机如何解读图像中的对象与动作。（原照片来源：*Chetan，2019*。）

计算上最简单的方法是斑块检测：算法通过比较视频帧之间发生变化的属性，可以判断视频前景中的人物是否与背景不同。这种斑块检测方法常用于描述运动流。在图1中，我们只需识别人物当前或曾经所在位置的热点，以此衡量人物相对于摄像机视野的位置，并建立背景帧。不过，这种前景与背景之间的二分比较方式，可以通过将更复杂的边界框或骨架分析集中在图像中较小的区域内，提高动作分类的计算效率。在这一子集中，算法可以进行以下几种操作：在帧内比较并跟踪这个边界框选定的子集，将 RGB 像素值与训练模型进行比较（分类），或者进一步将人物简化为某些部分之间的关系，并将这些部分视为可能的肢体或关键身体特征，从而构建数字化的拓扑骨架。最终，算法正是通过比较这两种简化结果——边界框内的像素值或骨架——来评估人类动作。这些动作可根据活动的复杂程度分为几类：基本动作、单人动作、交互和群体动作（Al-Faris 等，2020），并构成计算机评估活动的基础。结合这些方法和类别，我们便能识别个体在城市空间中的活动。

3 尽管本文不会全面回顾这项技术及其应用深度，但其他论文曾尝试对各种方法进行分类。例如，参见 *Ibrahim* 等人，*2020*年。

4.2. 参与者与建设者

“智慧城市”一词通常指将计算工具纳入市政当局及各地区的运营和管理之中。然而，随着私营部门在城

规划或政策制定的核心，是将复杂问题重新整理为一组离散的标准化测量指标，使塑造环境的人能够理

尽管关注重点主要是街道的服务水平或通行量，但道路基础设施管理也受到了广泛关注。随着现有且日渐

类似技术也被用于分析室内空间，尤其是零售或商业环境。Indoor Atlas 和 Brickstream 等公司采用了

公众对网上购物的广泛采用，使得对街道这些区域的管理变得更为重要；这片地产如此宝贵，以至于城

在以人为本的公共空间，例如人行道和广场，Numina 和 Placemeter 等公司一直尝试对步行或骑自行车

如何准确理解行人行为的复杂性，相关研究仍然很少。虽然理解非通勤行为的研究正在兴起（Seer 等，

随着 Intersection（Sidewalk Labs 的子公司，而 Sidewalk Labs 又是谷歌的姊妹公司、Alphabet 的子

2018）。并

术被

尽管人们并非从根本上反对借此了解公共行为，但这些摄像机技术也可以利用同一套算法，大规模用于警

l传感平

术；

然而，在围绕使用这些技术展开的激烈公众辩论中，人们的担忧涉及公共和私人所有的数字基础设施。

）。

这些技术的应用招致了诸多批评，批评者甚至包括产业内部人士（Shepardson，2020）。批评主要针对

都

以

用于识别人脸的同类技术，也被用来识别面部特征，以服务于广告和零售。新兴技术能够读取面部特征，

4.3. 底层数据与根本问题

身份、轨迹或输出的生成，有赖于创建能够解读实时现实世界数据流的模型。因此，需要数据集来训练和

计算机视觉识别中最常用的训练数据集包括 ImageNet（Deng 等，2009），其中包含约 1,400 万张人工标注的图像和 20,000 个类别；以及 Open Images，其中有 900 万张图像和近 20,000 个经人工核验的类别（Krasin 等，2017年）。与此同时，许多数据集专为训练人类动作模型而创建（Ofli 等，2013），其中大多数是从电影、电视或在线视频分享平台等现有来源中 opportunistically 获取的（Idrees 等，2017年；Patron-Perez 等，2012；Smaira 等，2020年）。例如，Kinetics-700数据集利用 65 万段 YouTube 视频片段，构建了一个涵盖 700 种人类动作的数据集。商业利益也在推动专门数据集的创建，以提升计算机视觉在市场中的能力。Affectiva-MIT面部表情检测数据集（McDuff、El Kaliouby、Senechal 等，2013）是一组引人注目的图像，呈现了世界各地人们的面部情绪，使研究人员和企业能够测量情感，并应用于零售、媒介和广告领域（McDuff、El Kaliouby、Demirdjian 等，2013）。人们对自动驾驶汽车兴趣的日益增长，促成了 Cityscapes 数据集的创建（Cordts 等，2016）；该数据集对 25,000 张街道拍摄的城市景观图像进行了标注。

尽管后者有望塑造人们对城市空间的数字化理解，但其对车辆的关注，使图像集的空间视角局限于驾驶员，以及汽车如何通过城市的这一单一方面进行导航。《城市景观》确实识别了人物和30个类别，范围从“天空”到不同类型的街道物体和车辆，但这些信息仅用于说明车辆在街道上导航的能力，而非真正理解城市空间如何运作。此外，由于图像仅取自德国城市，该数据集可能缺乏潜在的多样性、基础设施历史、标准以及出行选择。

这一点在城市主义领域提出了一个关键问题：它们能否体现城市真实的体验与社会多样性？当然，不能假设汇集不同细节程度和深度的数据总是毫无问题（Brannen，2005年）。创建这些数据集的挑战在于，组织和标注图像的人所秉持的价值观和假设，往往嵌入其中且未被明言。

首先，这些数据集追求普遍适用性，因而构建出一种更重视普遍性而非特殊性的本体，而在某些领域，可能并不存在公认的标准。对建筑风格的社会认知，或

它在广泛社会中所发挥的作用，例如，在不同文化和语境之间差异很大（Berlyn，1971；Kubo等，2010）。

其次，由于更严重的社会不平等，这些通过机会性收集图像而构建的数据集（如 ImageNet 和 Kinetics-700）可能会遗漏特定群体的经验。电视和媒介中的性别与种族不平等（R. L. Collins，2011；Desmond与Danilewicz，2010），以及地理多样性的缺乏（Shankar等，2017），可能会因这些数据集被衍生用于训练新算法而进一步加剧，除非采取措施加以改善。例如，YouTube 上的使用者和内容中，占比最大的一部分来自美国。同样，美国在电视和媒介制作方面拥有巨大的全球软实力。任何基于该数据集训练的算法，在非美国语境中的表现都会更差；一种文化与这一标准越不同，误差就越大。要使数字图像发挥作用，就必须依据某种知识体系对其进行组织（Pasquinelli，2015）。当这些数据隐入算法的黑箱，并成为政策制定和城市实体改造的基石时，我们就有必要质疑，现有数据集和范式是否充分描述了城市主义。

5. 空间中的意义

计算领域的创新为城市研究开辟了新的途径，但在城市主义话语中运用这些技术，仍有若干关键问题值得考量。正如 Duarte 和 DeSouza（2020年）所指出的，城市技术必须更深入地考量大规模数据方法论背后的认识论，以及这些方法如何塑造关于城市的本体和启发式方法；毕竟，由于城市规划在塑造社区方面肩负着特殊责任，社会与空间最终都将发生持久的转变。

计算机视觉技术所遵循的认识论标准之一，是倾向于还原主义或简化的数值表示（Dreyfus，1992）。这种方法本身就会深刻影响居民和决策者对建成环境的理解（Winner，2017年）。虽然这种还原主义能够形成关于某个地点的通用智能，并可能用于预测，但它往往忽视了赋予城市独特性的短暂特质。开发这些技术的公司致力于提升其产品的可扩展性，或使产品具有普遍适用性。这种追求效率的策略与关注某个地[illegible]J13；[illegible]排

这一界定上的冲突与 Clifford Geertz（1973）关于采用解释性方法理解文化的论述相呼应。在格尔茨看来，人类学研究的既有范式是通过薄描述展开的，这类描述源自对行为表层的观察。与许多城市技术所[illegible]眼来[illegible]上，

用于记录城市的指标的制定，也可能使描述偏向某一种类型。此外，人们还担心，通过数据收集和分析[illegible]务[illegible]

基础设施效率指标之上。此类数据收集相对容易，这会影响城市的描述方式，并带来一种风险：人们

这种冲突是大数据研究这一新兴实践的根本特征，更不用说将其置于城市等复杂社会环境之中。格尔茨提出深描，旨在推动民族志研究探寻更深刻的意义。然而，他的做法以既有实践模式为基础，并据此展开批评。但在城市科学研究中，正如其所立足的计算社会科学一样，实践准则仍在形成之中（Jemielniak，2020年；Rossman 和 Rallis，2017年）。因此，对城市现象的批判性观察陷入一种尴尬的未定状态，介于数据科学话语与解释性人类学研究之间：前者的主流公理承认模型不可知，并认为在类似黑箱中进行分析所带来的准确性提升，足以弥补这一过程的代价。

从计算机出现之前的研究中，或许可以找到在还原主义与阐释之间这一过渡地带进行观察的经验。在开展街道生活项目、记录公共空间的过程中，威廉·H·怀特（1980年）曾运用超级8毫米胶片来了解广场和公共空间的使用情况。怀特及其同事录制并分析了数百小时的延时摄影胶片，在为数不多的案例中评估匿名行人行为的差异与规律。他的研究旨在提供数据，以衡量城市空间中的活动模式，而当时这类数据尚不存在。怀特的调查是最早尝试利用新技术大规模监视并了解人类活动的研究之一。尽管这一过程仍极其依赖人工，研究人员需要手动梳理数小时的视频，但它尝试将整个公共广场作为一个整体加以记录，也展示了一种当时公众尚无法轻易获得的技术。通过这一项目，研究者既揭示了普遍行为，例如公园使用情况与“可供坐憩的空间”数量直接相关，而非如先前所认为的那样取决于形状或面积；也揭示了个人的特异性行为，例如在座位与位置选择上的性别差异。

Whyte 的研究兼顾两种层面：一是探寻精简、概括性的行为模式，二是考量从具体语境中得出的丰富认识。这种双重视角使他得以发现能够为分区法规提供依据的行为实践，并提出干预措施。例如，研究发现，成功成为热门聚集场所的广场，往往有一个关键特征：供人坐下的地方种类繁多、数量充足，包括长椅、可移动椅子、矮墙座沿和台阶。研究还发现，树冠、水景设施、公共艺术和食品摊贩等物理特征，都有助于吸引人们前往城市广场和公园。这些特征有助于形成正反馈循环，促使人们聚集在一起；而那些设有无窗墙面，且没有商店、窗户或门的街道，则鲜少有人活动。用 Whyte 的话说：“最吸引人的，看来就是其他人。”如今，这些发现已被视为理所当然，但在当时，它们打破了现代主义城市设计的主流范式。这些发现尤其促使纽约市通过分区法规，要求广场和公有公共空间的地面高度与街道标高相差不得超过三英尺，以便人们看得见、容易进入；同时也推动了布莱恩特公园的全面改造。20世纪80年代，该公园鲜有人活动；改造旨在提高其可见度，并增设更多座位，其中包括如今人们熟悉的可移动家具。如今，公园全年都很受欢迎，设有1,000多把可移动椅子、数个咖啡亭，并举办许多定期活动。

5.1. 意义冲突

当时，人们对数字媒介促成的、更加开放的公民治理带来的新机遇抱有很大乐观态度（Goldsmith & Crawford，2014）。然而，新自由主义发展与产业以规模为导向的关注点之间目标不一致，而在许多情况下，城市管理中较为缓慢、以讨论为基础、面向社区的民主进程，则导致各方采用的指标彼此不兼容，产生了相互冲突的结果。尽管人们并不否认，了解并枚举居民如何使用城市空间的能力，能够促成更恰当、响应性更强的干预措施，但这些

数据集对城市主义者而言究竟有多大效用，仍受到质疑。这种冲突的根源在于各方如何看待枚举这一挑战，以及如何界定公共空间的“使用”。

同样，计算强度方面的权衡限制了技术人员处理图像时所能达到的保真度。由于图像处理需要强大的计算能力，而许多公司为了保护隐私而依赖性能较弱的边缘计算平台，因此可枚举内容的规模、细节和精度直接取决于可用的计算资源数量。这些限制贯穿技术的各个环节：从采集图像的像素分辨率，到图像的存储与传输，再到将图像转换为可用测量结果所需的处理能力。因此，特定技术所宣称的许多优势，都会受到此类计算成本及其对物理资源要求的削弱。

归根结底，这是一系列权衡取舍；容量限制、社会影响与其意义厚度之间的相互关系，为评估市民如何判断计算机视觉或任何智慧城市技术是否适用提供了一种方法。在图2中，我们评估了前文提到的公司，以梳理这三项指标各自涉及的权衡与局限。综合来看，这些指标既能帮助市民评估各种技术是否适合其社区，也能找出在不同领域中具有相似机遇与限制的技术本体论集群。

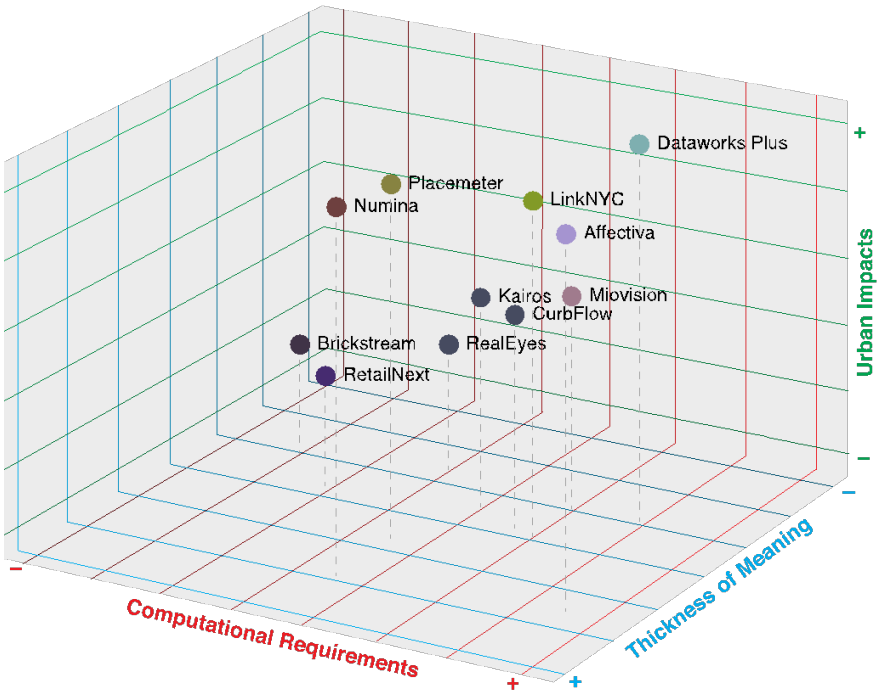


图2. 评估各种技术的含义、影响和计算成本。

这一权衡分类体系也为市民提供了一个视角，用以评估计算机视觉技术可能产生的社会影响。对于维护人行道这类对整个社会影响不大的活动而言，高保真度和细节可能并非必要，处理高分辨率实时图像所产生的成本也可能没有必要。然而，这些技术在警务中的应用会对市民产生重大影响，因为其中涉及司法公正问题，因此需要极高的精度和准确性。例如，底特律人脸识别项目的误报率很高，警察局长指出其错误率为96%

这反映了这些技术和算法的不足之处（Lee，2020），也应促使市民审慎看待其输出结果所代表的含义。

6. 将方法重新框定为城市语义

许多旨在理解人们如何使用空间的基于视觉的智慧城市技术，并未在城市主义话语中具体说明人们如何使用城市空间。这为计算机视觉研究提供了契机，使其能够开发语义上更具体、适切且充分考虑城市研究细微之处的方法。然而，城市语义计算机视觉会是什么样子？借用格尔茨的论述来说，目前的计算机视觉已经擅长识别眨眼，但在识别这些离散对象在城市中可能意味着什么方面，表现却很差。

当前最先进的技术已使计算机能够单独识别物体和动作。然而，即便识别准确，若没有情境来提供背景，相关定义的描述仍然十分单薄。图3中，四张“人们坐在长椅上”的照片经 ImSitu物体识别算法（Yatskar等人，2016）处理，从算法的视角识别图像。该算法尝试进行情境识别，并通过概述主要活动、参与者及其角色，提供简明的总结。该算法基于 FrameNet 训练，FrameNet 包含超过125,000张图像和200,000种独特情境。城市设计与规划领域的从业者也对相同图像进行了识别；每张图像采用出现频率最高的描述，尽管不同的人类描述之间差异不大。

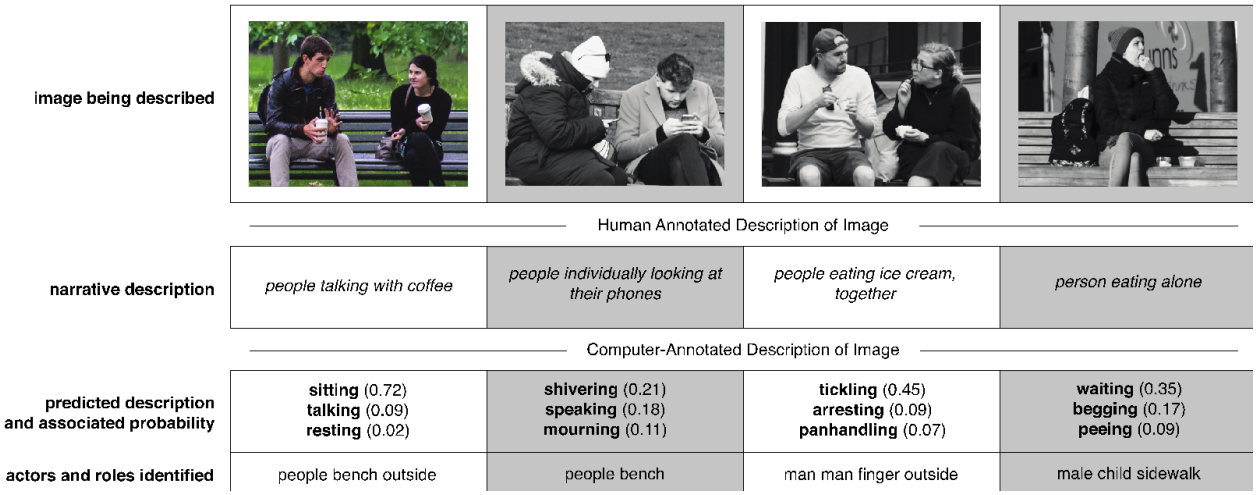


图3. *ImSitu*算法与人类对四张长椅上坐着的人的图像进行身份识别的结果。（照片来源：*hjl, 2012*；*byronv2，2019*、*2020a*、*2020b*）

研究结果表明，算法对四张图像的薄描述相当准确，但由于各图像的质量和颜色存在差异，置信度也有所不同。尽管算法能够识别出坐在长椅上的人这一点颇有意思，但它未能捕捉到各群体具体在做什么这一细微差别。这可能有多种原因，包括缺乏关于这些特定活动的词汇数据、图像质量或拍摄视角不佳，或数据集训练不当。无论如何，该算法未能达到本就简化的人类描述的水平，也凸显出当前评估城市活动厚描述所面临的挑战。

认识到计算机视觉最先进水平的局限性后，我们指出了技术开发者可以借以更深入理解城市空间中活动的潜在方法。使用算法黑箱所面临的挑战在于，计算结果所依据的过程并不透明。然而，理解算法运作的一般机制，或许能为计算机视觉应用于城市特定语境创造机会。

首先，我们可以考察模型的衍生方式，以及训练中常用的通用库。通过先验语境化，我们可以构建具有空间特异性的图像语料库，其中包含针对城市主义校准的描述和行为标注，从而实现城市特定识别。也就是说，必须以分类形式组织一个明确属于城市的城市数据集，以便算法判断[illegible]，哪些[illegible]，或者依赖一大批经过训练、能够依据特定城市标签对影像素材进行标注的人工标注员。

另一种选择是考虑后验方法。在这里，我们利用计算机视觉算法的优势，从图像中寻找常见模式，并由算法将相似的活动聚为簇。找到这些簇后，人类可以核验并为其添加标签，再用更丰富的描述加以标注。尽管在技术上可行，但基于影像素材进行无监督行为识别仍是新兴领域（O'Hara 等，2011；Soomro 与 Shah，2017），不过它可以让城市主义者重新审视 Whyte 等人的研究：计算机或许能发现人类无法发现的内容。

在城市社会语境中，这些技术与公众之间的关系也引发了一些问题。居民既不完全是这些服务的客户，也不是使用者，而且可能没有选择退出这些技术的选项。尽管这些黑箱过程可能会强化有害的现状，甚至进一步加剧既有偏见，但这些技术的创建方式因而如何推动规划、政策和设计决策，往往缺乏将受其影响者的意见纳入其中。波士顿 Beta Blocks 计划作为一种基于政策的机制而设立，旨在试点各类技术，并搭建一个促进社区参与和赋权的平台。除了开展技术方面的公民实验，该计划还与平台提供方组织教育工作坊和活动，无论这些提供方是否被纳入采购考虑（新城市机制市长办公室，2018）。

在公民实验的任务要求下，新技术会在预先选定的区域内公开试点，当地社区设有反馈机制，从而在选择、测试技术以及制定成功指标的过程中实现透明度和公民监督。居民受邀与市政府及企业共同确立与公民事务和隐私问题相关的价值观。他们还监督管理这些实验的流程和政策。因此，技术实施所处的社会环境，得以结合特定社区的人群、需求、地点和时间背景加以考量。

7. 结论

当前计算机视觉技术的发展方向，主要聚焦于自身的发展，以及提升其应用的广泛泛化能力。然而，俗话说："样样都行，样样稀松。"当这些技术被应用于城市的复杂环境中时，迄今为止的应用往往倾向于简化量化，却牺牲了公共空间的动态特征；正是这些特征吸引了数十亿人来到城市生活、工作和休闲。正因为这些空间具有动态特征，我们应重新审视这些工具的构想与开发，以充分认识其独特性。

借鉴人类学和社会科学领域以往的讨论，城市技术专家必须让算法枚举摆脱简单化的本体，转而采用能够更好地捕捉城市动态性的深描

。正如 Whyte 所言，枚举、描述与阐释之间的相互关系，对于得出有关城市空间的结论至关重要。随着这些技术不断普及，必须建立机制，审视这些数据集和算法在记录、训练、开发及使用过程中的偏差；这一点尤为重要，因为其带来的益处可能无法由居民公平享有，而且这些分析可能影响建成环境及其政策的形成。

尽管并不完美，这些提出的方法仍使城市主义者得以稍稍摆脱 David Hand（2020）所说的“暗数据”——即现有工具无法获取、也不符合现有方法，却仍可能影响最终决策和政策的数据。这些方法使城市主义者能够“付出理论研究的艰苦努力”，批判性地审视使用这些技术时所依托的本体论与认识论框架，并将实践重新导向与城市社会生活相关的指标，而非还原性的、以服务为导向的量化方法（Pickles，1997）。

8. 参考文献

ABI Research.（2021）。《智慧城市中的深度学习机器视觉》。
https://www.abiresearch.com/press/global-installed-base-smart-city-cameras-ai-chipset-reach-over-350-million-2025/ Ackerman, D.（2017年5月30日）。谷歌地图街景迎来十周年。CNet。
https://www.cnet.com/news/google-maps-street-view-celebrates-its-10th-birthday/ 美国公民自由联盟，N. Y.（2016）。纽约市的公共 Wi-Fi 引发隐私问题。Al-Faris, M., Chiverton, J., Ndzi, D., & Ahmed, A. I.（2020）。一篇关于基于计算机视觉的人体动作识别方法的综述。《成像学报》，6(6)。
https://doi.org/10.3390/jimaging6060046 Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J.（2010）。谷歌街景：在街道标高捕捉世界。《计算机》，43(6)，32–38。
https://doi.org/10.1109/MC.2010.170 执法记录仪分析的属性检测。（2020）。IBM 智能视频分析文档。
https://www.ibm.com/docs/en/iva/2.0.0?topic=video-attribute-detection-body-camera-analytics Azar, M., Cox, G., & Impett, L.（2021）。引言：机器视觉的方式。载于《人工智能与社会》（第1–12页）。施普林格科学与商业媒体德国有限公司。https://doi.org/10.1007/s00146-020-01124-6 Berlyn, D. E.（1971）。《美学与心理生物学》。Appleton-Century-Crofts。Brannen, J.（2005）。混合方法：定性与定量方法进入研究过程。《国际社会研究方法学杂志：理论与实践》，8(3)，173–184。https://doi.org/10.1080/13645570500154642 Brill, M.（1989）。探索当今城市公共生活的本体论。《地方》，6(1)，24–31。http://escholarship.org/uc/item/4kc602c7 byronv2。（2019）。互相发短信 *[照片]*。Flickr。https://flic.kr/p/23B3Jc4 byronv2。（2020a）。冰淇淋时光 *[照片]*。Flickr。https://flic.kr/p/2jjDBQv byronv2。（2020b）。户外午餐 *[照片]*。Flickr。https://flic.kr/p/2iEczU1 Chetan, V.（2019）。男子从岩石上跳下 *[照片]*。Pexels。https://www.pexels.com/photo/man-jumping-from-a-rock-2923157/ Chidster, M.（1989）。《公共场所，私人生活：广场与更广泛的公众》。《地方》，6(1)，32–37。http://escholarship.org/uc/item/9gr5n6hd Collins, J.（2020年7月15日）。警察执法记录仪视频展现乔治·弗洛伊德在致命逮捕过程中的痛苦。美国国家公共电台。
https://www.npr.org/2020/07/15/891516654/police-bodycam-video-provides-fuller-picture-of-george-floyds-fatal-arrest Collins, R. L.（2011）。《媒体性别角色的内容分析：我们现在身处何处，又该走向何方？》。《性别角色》，64(3)，290–298。https://doi.org/10.1007/s11199-010-9929-5 Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B.（2016）。用于城市场景语义理解的 Cityscapes 数据集。《IEEE计算机学会计算机视觉与模式识别会议论文集》，2016年12月，3213–3223。
https://doi.org/10.1109/CVPR.2016.350 Crawford, K.（2018年6月25日）。人工智能的白人男性问题。《纽约时报》。
https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html Czarniawska, B.（1992）。《探索复杂组织：文化视角：迈向人类学视角》。SAGE。Dahlberg, L.（2015）。《夏尔·马维尔，巴黎摄影师 / 刺穿时间：马维尔与阿杰之后的巴黎，1865–2012》。《摄影史》，39(2)，194–196。
https://doi.org/10.1080/03087298.2015.1035533 Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei。（2009）。ImageNet：大型层级图像数据库。《2009年IEEE计算机视觉与模式识别会议》，20(11)，248–255。
https://doi.org/10.1109/CVPR.2009.5206848 Desmond, R., & Danilewicz, A.（2010）。女性出现在新闻中，却未参与新闻：地方电视新闻中的性别角色。《性别角色》，62(11)，822–829。
https://doi.org/10.1007/s11199-009-9686-5 Dreyfus, H. L.（1992）。《计算机仍然做不到什么：对人工理性的批判》。麻省理工学院出版社。Duarte, F., & DeSouza, P.（2020）。《数据科学与城市：批判性方法》。《哈佛数据科学评论》。https://doi.org/10.1162/99608f92.b3fc5cc8 Eagle, N., & Pentland, A. S.（2009）。特征行为：识别日常活动中的结构。1057–1066。https://doi.org/10.1007/s00265-009-0739-0 Garvie, C., & Moy, L. M.（2019）。《监视之下的美国》。https://www.americaunderwatch.com/

Geertz, C.（1973）。深描：迈向文化阐释理论。载于《定性研究的转折点：在手帕上打结》（第143–168页）。Gehl, J.（1987）。《建筑之间的生活：使用公共空间》。Island Press。Gershenson, C.（2013）。交互对科学与哲学的启示。《科学基础》，18(4)，781–790。https://doi.org/10.1007/s10699-012-9305-8 Gill, K. S.（2020）。预测范式：工具主义的人类代价。载于《人工智能与社会》（第35卷，第3期，第509–517页）。施普林格。https://doi.org/10.1007/s00146-020-01035-6 Girardin, F., Calabrese, F., Fiore, F. D., Ratti, C., & Blat, J.（2008）。数字足迹：利用用户生成内容发掘游客。《IEEE普适计算》，7(4)，36–43。https://doi.org/10.1109/MPRV.2008.71 Goldsmith, S., & Crawford, S.（2014）。作为数字平台的城市。载于《响应型城市》。Jossey-Bass。Greenfield, A.（2013）。反对智慧城市。Do Projects。Hand, D. J.（2020）。暗数据：你不知道的事情为何至关重要。普林斯顿大学出版社。Harwell, D.（2019年8月28日）。Ring 这家门铃摄像头公司已与400个警察部门合作，扩大监控范围。《华盛顿邮报》。
h
t
t
p
s://
w
w
w.
w
a
s
h
i
ngtonpost.co
m/technology/2019/08/28/doorbell-camera-firm-ring-has-partnered-with-police-forces-extending-surveillance-reach/ Hernandez, J., Hoque, M. (Ehsan), Drevo, W., & Picard, R. W.（2012）。情绪计：在自然环境中统计微笑。《2012年ACM普适计算会议论文集——UbiComp '12》，301。https://doi.org/10.1145/2370216.2370264 Hill, K.（2020年8月3日）。遭算法错误指控。《纽约时报》。https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html Hinchcliffe, T.（2010）。伦敦的航空摄影与战后城市规划师。《伦敦学刊》，35(3)，277–288。https://doi.org/10.1179/174963210X12814015170232 hjl。（2012）。盲约——格林公园 *[照片]*。Flickr。https://flic.kr/p/cBGctS Hollands, R. G.（2008）。真正的智慧城市，请站出来？《城市》，12(3)，303–320。https://doi.org/10.1080/13604810802479126 Ibrahim, M. R., Haworth, J., & Cheng, T.（2020）。用机器之眼理解城市：城市分析中的深度计算机视觉综述。《城市》，96，102481。https://doi.org/10.1016/j.cities.2019.102481 Idrees, H., Zamir, A. R., Jiang, Y.-G., Gorban, A., Laptev, I., Sukthankar, R., & Shah, M.（2017）。面向"野外"视频动作识别的THUMOS挑战赛。《计算机视觉与图像理解》，155，1–23。https://doi.org/10.1016/j.cviu.2016.10.018 IHS Markit。（2019）。2019年安全技术十大趋势。载于《IHS Markit安全技术》。https://technology.informa.com/Research-by-Market/551540/security-technology Jacobs, A., & Appleyard, D.（1987）。迈向城市设计宣言。《美国规划协会学刊》，53(1)，112–120。https://doi.org/10.1080/01944368708976642 Jacobs, J.（1961）。《美国大城市的死与生》。Vintage Books。https://books.google.com/books?hl=en&lr=&id=P_bPTgOoBYkC&oi=fnd&pg=PA7&ots=JW1O38Fpf5&sig=X-9dkYK56vjYblU9O1I-kh0yYFQ#v=onepage&q&f=false Jacobs, J.（1970）。《城市经济》。兰登书屋。Jemielniak, D.（2020）。《厚大数据》。载于《厚大数据》。牛津大学出版社。https://doi.org/10.1093/oso/9780198839705.001.0001 Jiang, S., Fiore, G. A., Yang, Y., Ferreira, J., Frazzoli, E., & González, M. C.（2013）。手机轨迹数据的城市计算综述：当前方法、挑战与机遇。UrbComp。Kirchner, L., Mattu, S., Larson, J., & Angwin, J.（2016年5月23日）。机器偏见。ProPublica。https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing Kofman, A.（2018年9月8日）。纽约免费的LinkNYC互联网信息亭是否在追踪你的行踪？《拦截》。https://theintercept.com/2018/09/08/linknyc-free-wifi-kiosks/ Krasin, I., Duerig, T., Alldrin, N., Ferrari, V., Abu-El-Haija, S., Kuznetsova, A., Rom, H., Uijlings, J., Popov, S., Kamali, S., Malloci, M., & Pont-Tuset, Jordi and Veit, Andreas and Bel, K.（2017）。OpenImages：用于大规模多标签和多类别图像分类的公共数据集。https://storage.googleapis.com/openimages/web/index.html Kubo, M., Pasnik, M., & Grimley, C.（2010年4月6日）。严厉的爱：为粗野主义辩护。《建筑师》杂志。https://www.architectmagazine.com/design/tough-love-in-defense-of-brutalism_o Kwet, M.（2020年1月27日）。视频监控产业复合体的崛起。《拦截》。https://theintercept.com/2020/01/27/surveillance-cctv-smart-camera-networks/ 勒·柯布西耶。（1935）。《飞机》。《工作室》。Lecun, Y., Bengio, Y., & Hinton, G.（2015）。深度学习。载于《自然》（第521卷，第7553期，第436–444页）。《自然》

出版集团。https://doi.org/10.1038/nature14539 Lee, T. B.（2020年6月30日）。底特律警察局长承认人脸识别错误率高达96%。*Ars Technica*。
https://arstechnica.co
m/tech-policy/2020/06/detroit-police-chief-admits-facial-recognition-is-wrong-96-of-the-time/ Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., & Zhang, W.（2015）。利用谷歌街景和改进的绿视指数评估街道层面的城市绿化。《城市林业与城市绿化》，*14*(3)，675–685。https://doi.org/10.1016/j.ufug.2015.06.006 Lin, L., & Purnell, N.（2019）。一个由十亿台摄像机监视你的世界近在眼前。《华尔街日报》。
h
tt
ps://www.wsj.com/articles/a-billion-surveillance-cameras-forecast-to-be-watching-within-two-years-11575565402 Lynch, K.（1960）。《城市意象》。麻省理工学院出版社。Massaro, E., Ahn, C., Ratti, C., Santi, P., Stahlmann, R., Lamprecht, A., Roehder, M., & Huber, M.（2017）。汽车作为环境感知平台 [观点]。《*IEEE* 会议论文集》，*105*(1)，3–7。https://doi.org/10.1109/JPROC.2016.2634938 市长新城市机制办公室。（2018）。*BETA BLOCKS*。波士顿市。McDuff, D., El Kaliouby, R., Demirdjian, D., & Picard, R.（2013）。基于互联网收集的微笑反应预测在线媒体效果。*2013*年第*10*届*IEEE*国际自动人脸与手势识别会议及研讨会，*FG 2013*。https://doi.org/10.1109/FG.2013.6553750 McDuff, D., El Kaliouby, R., Senechal, T., Amr, M., Cohn, J. F., & Picard, R.（2013）。Affectiva-MIT 面部表情数据集（AM-FED）：在“野外”收集的自然、自发面部表情。*IEEE* 计算机学会计算机视觉与模式识别会议研讨会，881–888。
https://doi.org/10.1109/CVPRW.2013.130 Mozer, P.（2019年4月14日）。一个月、50万次人脸扫描：中国如何利用人工智能识别少数民族。《纽约时报》。
https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html Naik, N., & Philipoom, J.（2014）。街道评分——预测一百万处街景的感知安全。《IEEE 会议论文集……》。
https://doi.org/10.1109/CVPRW.2014.121 Norden, E.（1969）。马歇尔·麦克卢汉——与流行文化的高阶祭司、媒介形而上学家的坦诚对话。《麦克卢汉精要》，233–270。Noueihed, L.（2011年1月19日）。小贩的殉难引发突尼斯革命 | 路透社。路透社。https://www.reuters.com/article/tunisia-protests-bouazizi-idAFLDE70G18J20110119 O’Hara, S., Lui, Y. M., & Draper, B. A.（2011）。无监督学习人类表情、手势和动作。人脸与手势 *2011*，1–8。
https://doi.org/10.1109/FG.2011.5771473 Offenhuber, D., Nabian, N., Vanky, A., & Ratti, C.（2013）。数据维度：获取城市数据并使其易于获取。《*ICE*论文集——城市设计与规划》，*166*(1)，60–75。https://doi.org/10.1680/udap.12.00011 Ofli, F., Chaudhry, R., Kurillo, G., Vidal, R., & Bajcsy, R.（2013）。伯克利多模态人体动作数据库：一个全面的多模态人体动作数据库。*2013*年*IEEE*计算机视觉应用研讨会（*WACV*），53–60。https://doi.org/10.1109/WACV.2013.6474999 Paglan, T.（2016年12月8日）。隐形图像（你的照片正在看着你）——The New Inquiry。*The New Inquiry*。
https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/ Pasquinelli, M.（2015）。异常检测：元数据社会中异常现象的数学化。*Transmediale* 艺术节，1–10。Patron-Perez, A., Marszalek, M., Reid, I., & Zisser man, A.（2012）。电视节目中人类互动的结构化学习。《*IEEE* 模式分析与机器智能汇刊》，*34*(12)，2441–2453。
https://doi.org/10.1109/TPAMI.2012.24 Picard, R. W.（1995）。情感计算。载于感知计算部门技术报告（第221期）。Pickles, J.（1997）。工具还是科学？地理信息系统、科技科学与理论转向。《美国地理学家协会年刊》，*87*(2)，363–372。
https://doi.org/10.1111/0004-5608.00058 Rice, S.（1997）。《巴黎景观》。麻省理工学院出版社。Rossman, G. B., & Rallis, S. F.（2017）。定性研究导论：在田野中学习。载于《定性研究导论：在田野中学习》（第四版）。SAGE 出版社。
https://doi.org/10.4135/9781071802694 Salesses, P., Schechtner, K., & Hidalgo, C. a.（2013）。城市协作意象：绘制城市感知不平等图谱。《公共科学图书馆·综合》，*8*(7)。https://doi.org/10.1371/journal.pone.0068400 Schwarzer, M.（2017）。计算及新技术对建筑摄影的影响与

城市主义。《建筑》_*MPS*。https://doi.org/10.14324/111.444.amps.2017v11i4.001 Seer, S., Brändle, N., & Ratti, C. (2014). Kinect设备与人体运动学：研究行人行为的新方法。《交通研究C辑：新兴技术》，《48》，212–228。https://doi.org/10.1016/j.trc.2014.08.012 Selinger, E., & Fox Cahn, A. (2020, July 17). 最近参加过抗议活动吗？你的脸可能已经进入数据库。《卫报》。https://www.theguardian.com/commentisfree/2020/jul/17/protest-black-lives-matter-database Shankar, S., Halpern, Y., Breck, E., Atwood, J., Wilson, J., & Sculley, D. (2017). 没有表征就没有分类：评估发展中世界开放数据集中的地理多样性问题。arXiv。http://arxiv.org/abs/1711.08536 Shepardson, D. (2020, September 11). IBM表示，美国应对人脸识别系统实施新的出口管制。《路透社》。https://www.reuters.com/article/us-ibm-facial-recognition-exports/ibm-says-u-s-should-adopt-new-export-controls-on-facial-recognition-systems-idUSKBN2621PV Smaira, L., Carreira, J., Noland, E., Clancy, E., Wu, A., & Zisserman, A. (2020). 关于Kinetics-700-2020人类动作数据集的简短说明。arXiv。http://arxiv.org/abs/2010.10864 Soomro, K., & Shah, M. (2017). 视频中的无监督动作发现与定位。《2017年IEEE国际计算机视觉会议（ICCV）》，696–705。https://doi.org/10.1109/ICCV.2017.82 空间分析实验室。(2019). 《族裔语言景观数据》。https://slab.today/2019/09/ethnicity-lld/ Stanley, J. (2019). 机器人监控的黎明。载于美国公民自由联盟（6月刊）。https://www.aclu.org/report/dawn-robot- surveillance Sun, P., Hou, R., & Lynch, J. P. (2020). 利用深度学习衡量公共开放空间的使用情况：底特律河滨的一项基准研究。arXiv，《1》，2228–2237。Talen, E., & Ellis, C. (2002). 超越相对主义：重拾对良好城市形态的探索。《规划教育与研究杂志》，《22》(1), 36–49。https://doi.org/10.1177/0739456X0202200104 Talen, Emily, & Ellis, C. (2015).《超越相对主义：重拾对良好城市形态的探索》。36–49。Venturi, R., Brown, D. S., & Izenour, S. (1972).《向拉斯维加斯学习》。麻省理工学院出版社。Whyte, W. (1980).《小型城市空间的社会生活》。保护基金会。http://trid.trb.org/view.aspx?id=521122 Winner, L. (2017).《人造物有政治性吗？》劳特利奇出版社。世界经济论坛。(2020). 最后一公里生态系统的未来。载于《面向公共和私营部门参与者的转型路线图》（1月刊）。https://www.weforum.org/reports/the-future-of-the-last-mile- ecosystem Yang, S., Bailey, E., Yang, Z., Ostrometzky, J., Zussman, G., Seskar, I., & Kostic, Z. (2020). COSMOS智能交叉路口：用于鸟瞰视角目标跟踪的边缘计算与通信。《2020年IEEE国际普适计算与通信会议研讨会，PerCom研讨会2020》。https://doi.org/10.1109/PerComWorkshops48775.2020.9156225 Yatskar, M., Zettlemoyer, L., & Farhadi, A. (2016). 情境识别：用于图像理解的视觉语义角色标注。《IEEE计算机学会计算机视觉与模式识别会议论文集》，《2016》-《Decem》，5534–5542。https://doi.org/10.1109/CVPR.2016.597 Yin, L., Cheng, Q., Wang, Z., & Shao, Z. (2015). 行人流量的“大数据”：探索使用谷歌街景图像进行行人计数。《应用地理学》，《63》，337–345。https://doi.org/10.1016/j.apgeog.2015.07.010 Zukin, S. (2020). 像城市一样观察：技术如何成为城市的。《理论与社会》，《49》(5–6), 941–964。https://doi.org/10.1007/s11186-020-09410-4