利用街景视觉数据大规模检测社区绅士化

Tianyuan Huang*、Timothy Dai‡、ZhechengWang*、Hesu Yoon†、Hao Sheng‡、吴恩达‡、Ram Rajagopal* 和 Jackelyn Hwang†
*土木与环境工程系，
†社会学系，
‡计算机科学系，斯坦福大学{tianyuah, timdai, zhecheng, hyoon28, haosheng, ramr, jihwang}@stanford.edu, ang@cs.stanford.edu

arXiv:2301.01842v1 [cs.CV] 4 Jan 2023

摘要——社区绅士化在塑造个人和整个社区的社会经济福祉方面发挥着重要作用。虽然已有一些研究尝试检测城市中的绅士化，但现有方法主要依赖从调查数据中估算的指标，需要投入大量人力进行标注，而且难以全面刻画整个街区。我们提出一种新方法，通过结合历史街景视觉数据，根据街区的外观大规模检测社区绅士化。我们将该方法的结果与先前文献中的绅士化指标及案例研究进行比较，展示了所提方法的有效性。我们的方法有望补充现有的绅士化指标，并成为城市研究者和政策制定者的有效资源。

索引词——**城市计算、计算机视觉、变化检测、多实例学习**

I. 引言

绅士化是指对先前缺乏投资、处于衰退状态的城市社区进行再投资和更新，并吸引中产阶级及中上阶层居民迁入的过程 [1]。检测正在绅士化的社区，对于研究城市空间的动态变化至关重要，并关系到经济和社会不平等问题。传统的绅士化检测方法依赖于收集人口统计信息，例如美国人口普查局开展的十年一次的人口普查。然而，通过此类基于调查的方法获得的数据往往受到空间和时间粒度的限制：美国覆盖范围最全面的全国人口普查每 10 年开展一次，而更频繁的测量方式（如美国社区调查（ACS）提供的 1 年期人口统计估算）仅覆盖人口数为 65，000+ 的地区。由于年度样本量较小，较小地理区域的美国社区调查数据仅以 5 年期估算的形式提供。简言之，传统方法不可避免地受到空间和时间粒度的限制。因此，有必要采用其他数据来源来衡量社区绅士化，以了解城市变化的动态。

为克服上述局限，并更全面地测量城市环境，近期已有研究尝试采用街景视觉



图 1. 包含延时街景图像对的社区容器。对于每个街区，我们提出的框架会从采样的街景视觉数据中检测绅士化视觉线索（以高亮图像对标示），随后学习聚合后的街区表征，以判定其绅士化状态。

城市中的数据。绅士化的可见方面体现了由多种行动者复杂协同促成的街区社会转型 [2] [3]，而街景图像为捕捉街区建成环境的视觉特征提供了宝贵视角。此外，与人口普查调查缓慢的十年一次频率相比，街景图像的更新频率更高；例如，谷歌街景已收集了可追溯至2007年的历史影像，并且在美国大多数城市地区每1—3年更新一次，从而创建了城市的数字时间胶囊。 [2]通过调查谷歌街景图像来测量绅士化。研究者已利用街景视觉数据开展图像分割与分类，以测量人们对城市外观质量及其变化的感知 [4] [5]。 [6]通过分析渥太华（加拿大）谷歌街景图像的时间序列，检测建成环境属性的视觉改善，既确认了已知的正在绅士化的城市区域，也揭示了此前未知的正在绅士化区域。然而，大多数以往项目也具有特定城市性，并依赖于对大量训练数据进行人工标注，从而限制了其模型和表征向多座城市

迁移的能力。更重要的是，现有方法只能检测与绅士化相关特征的原子单元，却无法刻画整个街区。鉴于绅士化是在社区层级定义的，将特征的原子单元聚合为街区表征至关重要。如图1所示，聚合步骤面临信息稀释的挑战，因为街区中所有采样的街景图像里，只有很小一部分包含与绅士化相关的视觉线索。

在本项目中，我们提出了一个两步框架，利用街景视觉数据大规模检测绅士化。最后，我们评估了三个正经历广泛绅士化的不同城市中的街区：美国加利福尼亚州奥克兰、美国华盛顿州西雅图和美国科罗拉多州丹佛。我们的主要贡献有三方面：1）我们提出了一种框架，通过有效利用其他模态的辅助数据集，以可扩展的方式从大量且嘈杂的街景视觉数据中识别绅士化信号。2）我们将绅士化信号的原子单元汇总至社区层级；各边界均可任意定义，从而能够灵活地与其他绅士化指标进行比较。3）我们通过定性考察这三个城市中可能正在绅士化的街区的具体情况，验证了所提出的方法。

II. 相关工作

A. 社区绅士化指标

尽管大多数绅士化研究学者依赖公开的人口统计与住房数据（通常来自美国人口普查和美国社区调查），但仍有少数研究系统考察了绅士化的视觉特征。为了大规模衡量城市中的绅士化，研究者曾在城市街区层级开展实地调查，由评估人员步行穿过街区，记录建筑物翻新和再投资等可见迹象 [7]。近来，人工编码员通过考察谷歌街景图像，检测由理论推导出的绅士化指标 [2]。 [6] 利用街景图像和深度神经网络及孪生网络，绘制了加拿大渥太华房产改善情况的地图；作者还通过网站问卷收集了成对图像的众包标签，方法与 [8]类似。此外，商业活动数据也被用于量化城市变化和衡量绅士化[9] [10]。得益于多个城市中采样密度更高的谷歌街景图像，以及通过注意力机制学习到的街区表征，我们提出的框架提供了一种更便捷、全面的工具，可用于大规模检测绅士化。

B. 街景视觉数据

街景视觉数据已应用于城市中的各种场景，例如指示区域功能 [11]、研究环境与慢性健康结局之间的关联 [12] ，以及衡量人群的福祉[13]。它们也被用于时空表示学习 [14] [15] [16] ，以生成城市社区的区域嵌入。近期研究表明可以从街景数据中推断收入、种族、教育和投票模式等社会经济属性 [17]。此外， [8]旨在根据街道视觉数据预测人们感知的安全水平； [4]则开发了类似机制，分析分别于 2007 年和 2014 年拍摄的成对图像，以描绘物理环境的改善与衰退，并分析其与社会经济属性的关联。然而，现有方法通过均值算子将坐标级信号聚合为街区级表示，这可能会稀释分布稀疏的关键元素。

C. 多实例学习

多实例学习是一种监督学习，其中一个类别标签被赋予一个实例包。在简单情况下，如果一个包中至少有一个正样本实例，该包就会被标记为正样本。传统上，多实例学习方法在将实例聚合到包 [18] [19]时，依赖平均池化或最大池化。全连接神经网络已被采用，并证明有益 [20]，；近期关于引入门控注意力机制 [21] 的研究表明，在多实例学习中，这种机制的表现优于常用的池化算子。我们沿用了这一研究方向，因为它提供了一种可学习的方式来为实例分配权重。然而，我们的问题设定更为复杂：即便包中的所有实例并非负样本，该包仍可能被标记为负样本（例如，几乎没有改善迹象的未发生绅士化社区）。我们提出的方法同时利用包标签和实例标签，并通过门控注意力机制将二者联系起来。

III. 方法

A. 问题陈述

定义 *1*（延时街景图像对）：每个延时街景图像对由 2 张图像组成，用于捕捉同一街道层级场景 $t^{(i)} = (s_e^{(i)}, s_l^{(i)})$，其中 $s_e^{(i)}$的记录时间戳早于 $s_l^{(i)}$。

城市社区中的跨时段街道层级图像捕捉了建成环境外观的变化，这些变化能够揭示与绅士化相关的信号。

定义 *2*（社区容器）：一个大都市区由一组城市社区 $N = \{n_1, n_2, \ldots, n_N\}$ 组成，这些社区的地理几何形状彼此不重叠。每个社区单元 n_j 包含一组延时街景图像对 $n_j = \{t_j^{(1)}, t_j^{(2)}, \cdots, t_j^{(K)}\}$。

我们的目标是以准确、可扩展且可解释的方式识别城市中的绅士化街区。为此，我们提出了一种两步方法，如图 2 所示，每一步都有各自的训练数据和标签，具体如下：

- 步骤1：从每个街区的延时街景图像对中检测建筑物的新建和重大翻修。
- 步骤2：将街景表示聚合到社区层级，并预测该街区的绅士化状态。

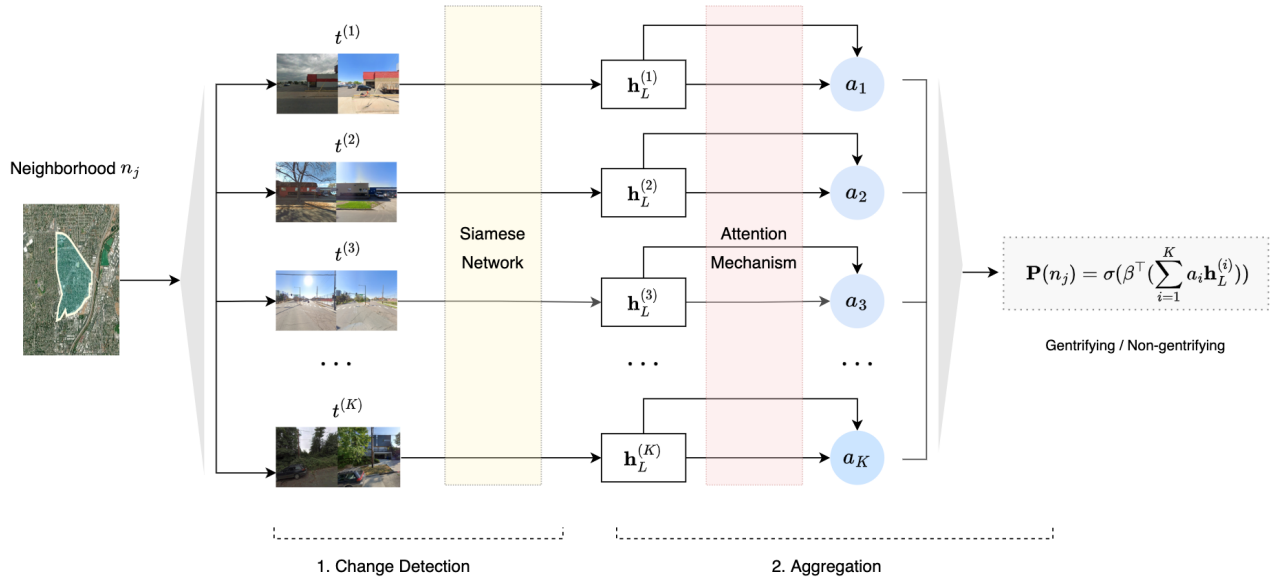


图2. 所提方法概览。步骤1通过孪生网络检测延时街景图像对中的建筑物新建和翻修。步骤2利用注意力机制将街景表示聚合至社区层级，以预测绅士化状态。

B. 步骤*1.* 街景变化检测

由于新建建筑和建筑物翻新通常被视为正在绅士化的街区的特征，因此我们的第一步是针对跨时段街景图像对构建变化检测任务。为定义这项任务，我们将每组跨时段街景图像对 t_i 标记为一个二元类别：包含有意义的变化（例如新建建筑）的图像对标记为正样本，不包含此类变化的图像对标记为负样本。这里，“有意义的变化”表示绅士化，而没有有意义的变化则包括完全没有变化，或道路上的汽车、天气状况等随机变化；这些因素可能会带来噪声信号。

针对这一任务，我们构建了一个孪生网络 [22]，包含 $L+1$层，其中包括一个双分支 ResNet [23] 模块，用于将输入域——跨时段街景图像对——非线性嵌入到某些欧几里得空间 $\mathbb{R}^d$中。令$\mathbf{h}_e^{(i)}{}_{,L}$ 表示跨时段街景图像对 $t^{(i)}$ 中较早时间戳 $s_e^{(i)}$对应街景图像的第(L)个隐藏向量，令$\mathbf{h}_l^{(i)}{}_{,L}$表示较晚街景图像 $s_l^{(i)}$ 对应的同一隐藏向量。我们注意到，仅对隐藏向量逐元素计算距离度量，并不足以从跨时段街景图像对中学习街道层面的变化。我们的考虑基于这样一个事实：大多数街道层面的变化都与一定层级的场景上下文（例如摄像机角度）相关。因此，我们将跨时段街景图像对 $t^{(i)}$ 的最终第(L)个隐藏向量定义为两幅图像的隐藏向量及其距离的拼接，其中距离以逐元素差值表示，如下所示： $h_L^{(i)} \in \mathbb{R}^M$

$$\mathbf{h}_L^{(i)} = \left[(\mathbf{h}_{l,L}^{(i)} - \mathbf{h}_{e,L}^{(i)})^\top, (\mathbf{h}_{l,L}^{(i)})^\top, (\mathbf{h}_{e,L}^{(i)})^\top\right]^\top, \quad (1)$$

其中 $M = 3d$。最终的向量表示后接一个全连接层，其结果作为单个输出单元，并通过 Sigmoid激活函数处理。更准确地说，预测标量表示为：

$$\mathbf{p}(t^{(i)}) = \sigma(\alpha^\top(\mathbf{h}_L^{(i)})). \quad (2)$$

令$\mathbf{y}(t^{(i)})$表示跨时段街景图像对 $t^{(i)}$的标签。我们假设，当 $t^{(i)} = (s_e^{(i)}, s_l^{(i)})$发生有意义的变化时，$\mathbf{y}(t^{(i)}) = 1$；否则，$\mathbf{y}(t^{(i)}) = 0$。我们的目标函数采用如下形式的二元交叉熵损失：

$$\mathcal{L}_S(t^{(i)}) = \mathbf{y}(t^{(i)}) \log \mathbf{p}(t^{(i)}) + (1 - \mathbf{y}(t^{(i)})) \log(1 - \mathbf{p}(t^{(i)})). \quad (3)$$

C. 步骤*2.* 街区层面聚合

一个街区 j 包含一组 K 跨时段街景图像对 n_j，并且与n_j 相关联的单个二元标签 Y 用于指示该街区是否正在绅士化。我们根据政府划定的人口普查区（采用2010年边界）获取所谓“街区”的物理边界，并从公开可用的美国社区调查数据 [24]中获取其二元标签 Y。我们假设包 n_j中的实例既无序也互不依赖，因此提出通过多实例池化聚合 n_j 中的街景嵌入。具体而言，我们从变化检测模型中提取 n_j 中每个 $t_j^{(i)}$ 对应的跨时段街景图像对$\mathbf{h}_L^{(i)}$ 的嵌入，然后通过对每个跨时段街景图像对的实例嵌入求加权平均，计算街区层面的特征向量$\mathbf{n}^j$：

$$\mathbf{n}_j = \sum_{i=1}^{K} a_i \mathbf{h}_L^{(i)}. \quad (4)$$

门控注意力机制 我们注意到，在预测包的标签时，每个包中有些实

例比其他实例更重要。例如，一对住宅改造幅度很大的跨时段街景图像对能提供强烈的绅士化信号，而一对几乎相同、呈现平淡无奇的高速公路场景的街景图像，无法像前者那样提供关于街区绅士化状态的丰富信息。如果在聚合时使用均值算子（即，$a_i = \frac{1}{K}$），重要实例所传递的信号可能会在街区级表示中被稀释。因此，我们提出采用门控注意力机制 [25]，具体做法是将方程 4 中的 a_i 参数化如下：

$$a_i = \frac{\exp\{\mathbf{w}^\top(\tanh(\mathbf{V}\mathbf{h}_L^{(i)}) \odot \mathrm{sigm}(\mathbf{U}\mathbf{h}_L^{(i)}))\}}{\sum_{j=1}^{K}\exp\{\mathbf{w}^\top(\tanh(\mathbf{V}\mathbf{h}_L^{(j)}) \odot \mathrm{sigm}(\mathbf{U}\mathbf{h}_L^{(j)}))\}}, \quad (5)$$

其中，$\mathbf{w} \in \mathbb{R}^{W\times 1}$，而$\mathbf{U}$，$\mathbf{V} \in \mathbb{R}^{W\times M}$是可学习参数。值得注意的是，门控注意力层在学习每个$\mathbf{h}_L^{(i)}$ 实例的权重 a_i 时引入了非线性。我们提出的模型旨在为每一对时间间隔的街景图像分配不同的权重，从而让最具代表性的街景级表征决定街区的绅士化状态。

最后，街区表征将输入分类器，该分类器由一个全连接层和 sigmoid 激活函数构成，如下所示：

$$\mathbf{P}(n_j) = \sigma(\beta^\top(\mathbf{n}_j)). \quad (6)$$

同样，我们使用交叉熵损失函数来训练此分类器，其中当 n_j 是正在绅士化的街区时，$\mathbf{Y}(n_j) = 1$；当 n_j 是未发生绅士化的街区时，$\mathbf{Y}(n_j) = 0$：

$$\mathcal{L}_{\mathcal{N}}(n_j) = \mathbf{Y}(n_j)\log\mathbf{P}(n_j) + (1-\mathbf{Y}(n_j))\log(1-\mathbf{P}(n_j)). \quad (7)$$

IV. 实验

为展示我们框架的有效性，我们针对奥克兰（加利福尼亚州）、西雅图（华盛顿州）和丹佛（科罗拉多州）开展了实验。我们采用美国人口普查局定义的人口普查区作为街区单元，因为所需的绅士化标签 [24] 可以直接从这些相同的街区划分中获取。尽管采用了这些预定义单元，我们的框架在应用于其他街区地理单元（例如人口普查街区组）甚至自定义边界时仍具有灵活性。

A. 数据集

历史谷歌街景图像 这些带时间戳的街景图像是通过 GoogleStatic Street viewAPI[1] 获取的，覆盖上述城市奥克兰、西雅图和丹佛，时间范围为 2007 年至 2022 年。为选取用于下载这些图像的地理空间位置，我们沿道路网络每隔 50 至 100 米采样一个坐标。表 I 显示了我们为每座城市采样的街景图像数量。对于每一对跨时段街景图像，我们确保较早拍摄的图像不晚于 2010 年，而较晚拍摄的图像不早于 2018 年，以尽可能增加该时间段内可观察到城市变化的情况。图 3 展示了这些街景图像的地理空间分布。

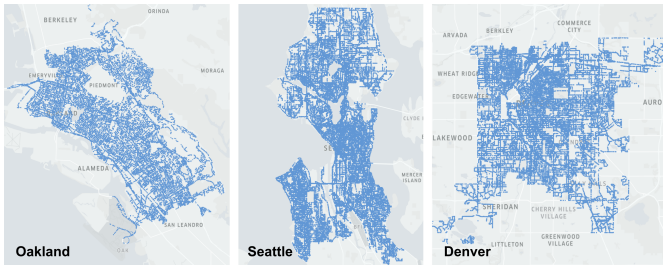
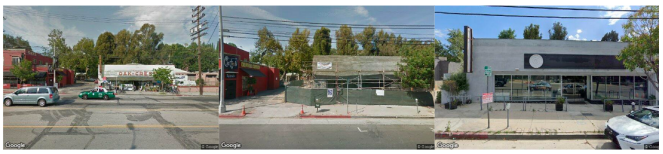


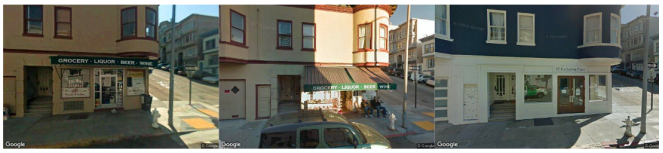
图 3。 三个研究对象中采样街景图像的地理空间分布

消亡的城市。

建筑许可证 建筑许可证数据来自相应城市政府的在线许可证中心。每条许可证数据都包含许可证的签发日期、类别、地理空间坐标和成本估算等信息，以及其他政府要求提供的各类信息。作为数据预处理步骤，我们根据通货膨胀调整工程金额，仅保留“新建”“改建”和“扩建”类别，并移除单年总成本低于 60,000 美元的建筑工程。这样一来，经过筛选的许可证记录所反映的城市变化足够显著，可以作为社区绅士化的潜在信号。



New construction of a full-service restaurant



Business conversion from a local grocery to a Yoga studio

图4. 通过建筑许可证和商业名录获取的绅士化相关视觉线索示例。

商业名录历史商业数据经许可后取自 Data Axle 的 ReferenceUSA[2] 产品，其中按年记录商家的名称、NAICS 分类、地址等信息。在此，我们提取了从必需品零售业（例如洗衣店、杂货店）转变为非必需品零售业（例如高档餐厅、咖啡馆、美术馆等）的商家，具体而言，是 2007 年至 2020 年间发生的转变。我们将此类商业变化视为绅士化信号的另一来源。

绅士化指标 为便于比较，我们使用基于社会和人口统计数据得出的绅士化指标

[1]可在 https://developers.google.com/maps/documentation/streetview 获取

[2]可访问：http://www.referenceusa.com/

这些数据通过2005—2009年和2015—2019年美国社区调查数据获得，并采用了 [24]中的方法。这些指标将人口普查区分为3类：正在绅士化、未发生绅士化和不可绅士化。不可绅士化街区在研究期初就是富裕街区，因此不具备发生绅士化的条件。需要明确的是，可绅士化街区包括正在绅士化和未发生绅士化的街区，且根据这些指标，只有正在绅士化的街区经历了绅士化。我们的工作聚焦于城市中的可绅士化街区；我们的目标是从可绅士化街区中检测出所有正在绅士化的街区。表 I 展示了我们所研究的每个城市中正在绅士化和未发生绅士化的街区数量（即人口普查区）。

表 I 数据集统计

City	# 街景图像	# 街区（人口普查区）	
		正在绅士化	未发生绅士化
西雅图	81,676	37	26
奥克兰	41,796	23	31
丹佛	72，620	47	22

B. 训练细节

变化检测模型 我们的实验采用 ResNet18[23] 作为骨干网络，将输入从图像空间转换到 $\mathbb{R}^d$ 的向量空间，其中 $d = 512$，从而将 $M = 512 \times 3$ 设为$\mathbf{h}_L^{(i)}$ 的维度。此外，二元预测器$\mathbf{p}(t^{(i)})$采用 0.5 作为阈值。根据建筑许可证和商业名录数据中的地理空间坐标，我们抽取了 2，476 张街景图像（即 1，238 对跨时段街景图像）来训练变化检测模型。对于每个有效坐标，我们下载 2 对跨时段街景图像：一对标记为正样本（即$\mathbf{y(t^{(i)})} = 1$），包含分别在变化日期之前和之后拍摄的 2 张图像；另一对标记为负样本（即$\mathbf{y(t^{(i)})} = 0$），其中两张图像均在变化日期之前拍摄。最后，我们将数据集划分为训练集（70%）和测试集（30%）。

门控注意力模型我们根据人口普查区的大小和道路密度，为每个人口普查区抽取 $K = 100$ 至 200 对延时街景图像。设置矩阵 $\mathbf{V}$、$\mathbf{U}$ 和 $\mathbf{w}$的形状时会用到 $W = 128$。与变化检测模型类似，我们将街区数据集划分为训练集（70%）和测试集（30%），并在二元预测器 $\mathbf{p}(t^{(i)})$ 中采用 0.5 作为阈值。此步骤中的标签表示街区 n_j 是否正在绅士化（即 $\mathbf{Y}(n_j) = 1$），或未发生绅士化（即 $\mathbf{Y}(n_j) = 0$）。

C. 基线模型

我们通过三项消融研究评估所提出的方法。1）**预训练且无注意力**同时跳过步骤1中的变化检测模型和步骤2中的注意力机制。相反，它通过在 ImageNet 上预训练的主干模型生成图像嵌入，并在步骤2中聚合实例嵌入时使用均值算子（即 $\mathbf{n}_j = \sum_{i=1}^{K} \frac{1}{K} \mathbf{h}_L^{(i)}$）。2）**无注意力** 保留步骤1中相同的训练设置，跳过步骤2中的注意力机制，并在聚合步骤中使用均值算子，与 [4] 类似。3）**端到端** 移除步骤1中的变化检测标签，以端到端方式进行训练，从成对图像输入到街区预测输出，仅以绅士化属性作为标签。这三个模型作为我们的基线模型。

V. 结果与讨论

A. 绅士化属性预测

首先，我们按照步骤1中的描述训练并评估变化检测模型。基于孪生网络的模型在预测通过许可证和商业名录获得的正样本对时，准确率达到 93%。接下来，我们提取延时街景图像对的嵌入，并将其输入完整模型的步骤2以及**无注意力**模型，以基准评估各模型根据 [2] 中的测量结果预测街区绅士化属性的能力。为抵消正在绅士化与未发生绅士化标签之间的类别不平衡，我们报告模型在测试集上的平衡准确率。如表 II 所示，我们的完整模型显著优于所有基线模型。具体而言，**预训练且无注意力**模型在训练过程中出现损失振荡，并将三个研究城市中的测试街区全部预测为未发生绅士化或全部预测为正在绅士化。**无注意力**模型将奥克兰测试集中的部分非绅士化街区误分类，并将西雅图和丹佛的所有街区都预测为正在绅士化。**端到端**模型出现过拟合，在三个城市的测试集上泛化能力较差；这可能是因为街景中的某些噪声信号被识别为绅士化相关线索，因为**端到端**模型省略了步骤1中的变化标签。在这些消融研究中，我们证明步骤1和步骤2对于从延时街景图像对中提取足够的城市变化信息以预测绅士化状态都至关重要。我们的完整模型在三个城市的平衡准确率达到 74%，召回率达到 89%，并且预测为正在绅士化的街区数量多于测量标签所示的数量。不过，我们注意到，绅士化测量 [2] 依赖于美国社区调查（ACS）中的人口统计和住房数据，而我们的方法利用了不同的数据来源：街景影像所捕捉的城市外观。因此，为了进一步验证我们提出的模型，我们进行了定性分析和案例研究。

B. 解读学习到的权重

为了揭示所提出的方法如何区分正在绅士化与未发生绅士化的街区，我们通过可视化所有街区容器的权重分布，仔细考察注意力机制。在图5中，我们为每个街区抽取 $K = 100$ 组延时街景图像对 $t^{(i)}$，并根据其数值按降序排列

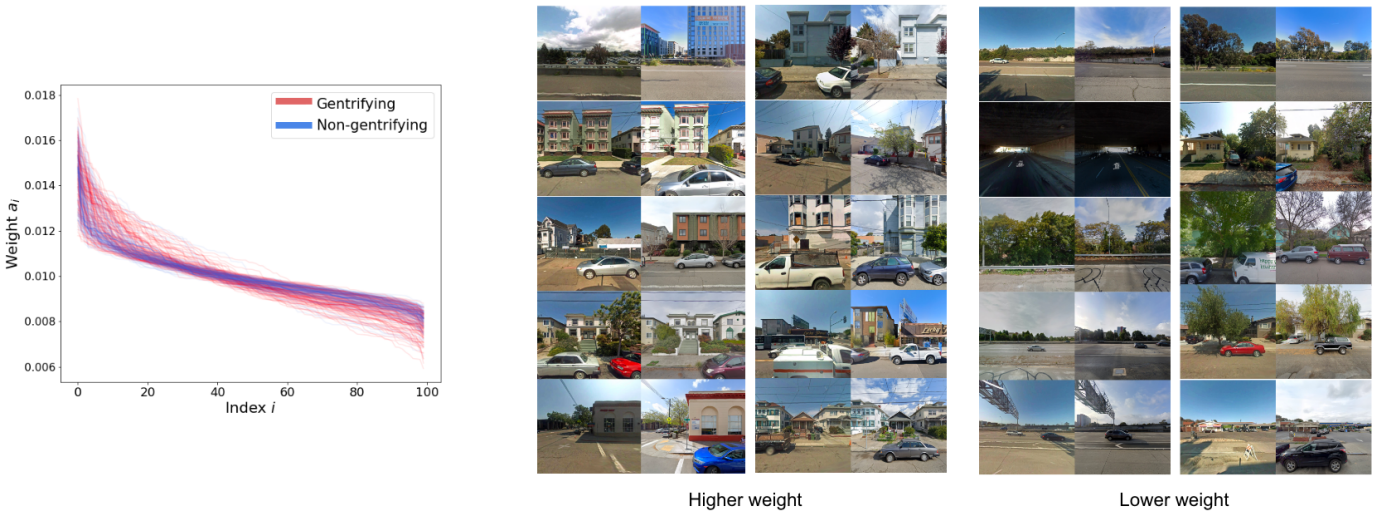


图 5. **左图**：正在绅士化街区（红色）与未发生绅士化街区（蓝色）的权重分布对比，其中每条线代表一个包含 100 对延时街景图像的街区。横轴为每对延时街景图像的索引 $t^{(i)}$，按权重值 a_i 降序排列。纵轴为权重值 a_i。**右图**：权重较高和较低的延时街景图像对。

表 II 街区绅士化属性预测结果

评估城市	模型	Acc.	平衡准确率
奥克兰	预训练且无注意力	0.56	0.50
奥克兰	无注意力	0.63	0.57
奥克兰	E2E	0.56	0.50
奥克兰	完整模型	**0.81**	**0.83**
西雅图	预训练且无注意力	0.58	0.50
西雅图	无注意力	0.58	0.50
西雅图	E2E	0.41	0.35
西雅图	完整模型	**0.71**	**0.66**
丹佛	预训练且无注意力	0.68	0.50
丹佛	无注意力	0.68	0.50
丹佛	E2E	0.63	0.51
丹佛	完整模型	**0.74**	**0.72**
全部平均值	预训练且无注意力	0.61	0.50
全部平均值	无注意力	0.63	0.52
所有项的平均值	E2E	0.54	0.45
所有项的平均值	完整模型	**0.75**	**0.74**

a_i 。我们观察到，非绅士化街区的曲线下降速度远快于绅士化街区的曲线。这意味着，非绅士化街区中的大多数延时图像对所分配的a_i 数值都接近平均值 0。01，只有少数图像对的权重较高。相比之下，绅士化街区中权重较高的图像对多得多，因此其 a_i分布更为两极化。我们基于这样一个合乎逻辑的假设来解释这一观察：绅士化街区会包含更多发生显著变化的延时街景图像对。凭借绅士化街区与未发生绅士化街区之间学习到的权重差异，我们表明，所提出的模型能够从随机噪声中提取出绅士化的潜在信号，方法是通过赋予显著权重来对正在绅士化的街区进行分类。

为了进一步说明学习到的权重所具有的语义含义，我们从一个随机抽取的街区中选取权重最高和最低的跨时段街景图像对各 20 组，并在图5中展示这些街景图像。我们发现，描绘建筑物的图像对权重较高，而描绘高速公路和植被的图像对权重较低。这验证了注意力模型会优先考虑信号较强的图像对，并降低可能引入噪声或提供无信息信号的图像对的优先级，最终通过降低不太相关实例的权重将其过滤掉，从而避免信息稀释。

C. 案例研究：检测潜在的正在绅士化街区

鉴于我们提出的绅士化测量方法利用了不同的数据源，并且与测量标签相比，预测出更多正在绅士化的街区，我们希望探索补充现有标签的方法，以便构建更全面、更灵活的绅士化测量。具体而言，我们探讨以下问题：哪些街区被分类为正在绅士化，却被标记为未发生绅士化？我们能否在这些街区中找到绅士化信号？它们是否可能处于绅士化早期阶段？为探讨这些问题，我们将所有可绅士化街区的预测结果进行可视化，并叠加由 [2] 得出的绅士化标签，如图6所示。乍一看，我们观察到，差异类别中的许多街区都毗邻被标记为正在绅士化的街区，这使我们推测，这些街区可能正处于绅士化早期阶段。其依据是，已在绅士化的街区所产生的社会经济影响可能会溢出到周边街区 [26]。

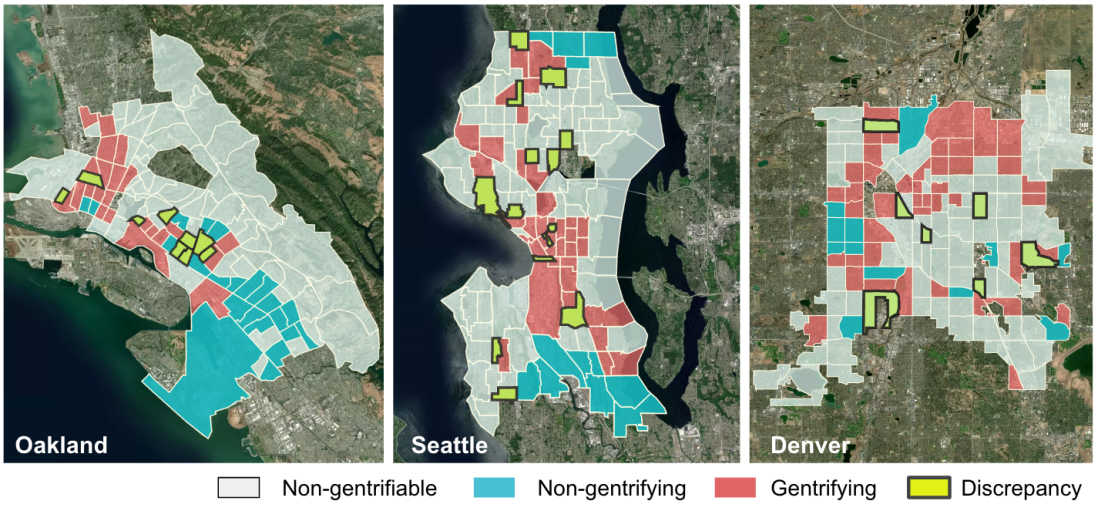


图 6. 研究城市中的社区绅士化标签与预测差异。预测差异标出了这些“假阳性”街区：它们被标记为未发生绅士化，但我们的模型却预测其正在绅士化。



图 7. 差异案例研究。通过为选定的延时街景图像对赋予较高权重，我们提出的模型能够检测到这些被标记为未发生绅士化的街区中的潜在绅士化信号。

为进一步评估预测差异，我们从这3个城市中选取了4个人口普查区，作为进一步的案例研究，逐一检查图像。这些人口普查区位于以下街区：奥克兰的Prescott；奥克兰的McClymonds；西雅图的信标山；以及丹佛的林肯公园。图7显示，我们提出的方法通过为具有相关信号的街景图像对分配更高权重，识别出一系列绅士化信号。具体而言，我们观察到住宅建筑发生了一些重大变化，包括修缮以及房屋重新粉刷，还有新建独栋住宅和公寓楼。此外，我们还检测到植被和绿化的恢复，以及道路和人行道等基础设施的翻新。由于基于美国社区调查数据开发的绅士化测量采用人口普查区社会经济指标的五年估算值，建成环境中的局部变化很容易被忽略。因此，我们提出的方法可以通过识别

对街区进行实体改善的具体信号来补充现有测量，而这些改善对绅士化进程至关重要；该方法还可以识别人口普查区内发生绅士化的具体地点。

VI. 结论与讨论

在本研究中，我们提出了一种框架，通过对街景视觉数据应用计算机视觉和统计分析，大规模检测正在绅士化的街区。我们首次通过一种可学习机制检测绅士化信号的原子单元，并将其汇总至社区层级；同时，我们在多个城市中预测绅士化属性，对这一机制进行了验证，涉及的街区数量远超以往研究。通过考察与绅士化相关的视觉线索的实地案例，我们观察到可能正在绅士化的街区，并通过若干案例研究对其进行评估。鉴于绅士化研究学者对于如何以定量方式衡量绅士化仍未达成共识， [27] [28] [29], 我们提出的方法展现出成为有效资源的潜力，可用于补充和完善现有方法。在我们分析的三个城市中，我们发现，与已处于绅士化进程中的街区相邻的街区，在视觉线索方面呈现出绅士化迹象，这表明这些街区可能正处于绅士化的早期阶段，或者绅士化仅发生在街区的部分区域。

尽管我们的数据驱动方法为衡量街区绅士化提供了一种新颖的方法，但仍存在以下局限性：1）我们对绅士化视觉线索的标注依赖于辅助数据集（即许可证和商业数据），而这些数据集的规模小于最终部署时使用的图像数据集，因此在测试集上面临分布外数据带来的挑战。具体而言，延时街景图像对中的随机噪声信号（例如视角差异、摄像机角度造成的遮挡）可能会影响模型在不同场景中正确泛化的能力。2）由于建筑施工和翻新等街景图像信息是绅士化最明显的视觉信号之一，我们的模型可能会对建筑物较多的地点投入更多注意力，而忽略人口较少的地点。3）最后，要全面衡量绅士化，仍有必要考虑人口统计和经济属性，以避免潜在的视觉偏差。尽管存在这些局限性，我们相信，所提出的方法和分析为学者和政策制定者提供了一个灵活的框架，有助于他们制定绅士化的量化指标。更广泛地说，定位和预测类似绅士化的城市变化，可以帮助地方政府和规划机构识别面临迁离风险的地点，并协助其优先安排某些基础设施投资和制定有针对性的政策干预。未来，我们可以将处理流程扩展到更多下游任务，以检测城市变化其他方面的相关要素和信号。

七、致谢

本项目获得了斯坦福大学以人为本人工智能研究所的 Google Cloud Grant 资助。作者谨向 Sarthak Kanodia、Herman Donner 和 Jeremy Irvin 致谢，感谢他们给予的大力指导。

参考文献

[1] R. J. Lawrence, “Willem van vliet（编），住房百科全书”，《住房、理论与社会》，第17卷，第1期，第43–44页，2000年。 [在线]。可访问：https://doi.org/ 10 1080 140360900750044773 。/ J. [2] Hwang和R. J. Sampson，“绅士化的不同路径：芝加哥街区的种族不平等与更新的社会秩序”，《美国社会学评论》，第79卷，第4期，第726–751页，2014年。[在线]。可访问：https://doi.org/ 10 1177 0003122414535774 。/ N. Smith 和 P. Williams，《城市绅士化》，第 1 版。伦敦：Routledge, 1986年。 [在线]。可访问：https://doi.org/10.4324/ 9781315889092[3]N. Naik、S. D. Kominers、R. Raskar、E. L. Glaeser和C. A. Hidalgo，“计算机视觉揭示城市实体变化的预测因素”，《美国国家科学院院刊》，第114卷，第29期，第7571–7576页，2017年。 [在线]。可访问：https: // www.pnas.org/content/ 114 29 7571 / / N. Naik、J. [4] Philipoom、R. Raskar和C. A. Hidalgo，“Streetscore——预测一百万处街景的感知安全性”，2014年IEEE计算机视觉与模式识别会议研讨会，第793–799页，2014年。[5] L. Ilic、M. Sawada和A. Zarzelli，“利用深度学习和谷歌街景，在加拿大一座大城市中对绅士化进行深度制图”，《PLOS ONE》，第14卷，e0212814，2019年3月。[6] D. J. Hammel和E. K. Wyly，“一种利用人口普查数据识别绅士化地区的模型”，《城市地理学》，第 17, 3, 248 268, 卷, 1996 [ ] 10 2747 0272 3638 17 3 17, 3, 248 268, 期，第 1996 [ ] 10 2747 0272 3638 17 3 17, 3, 248 268, – 1996 [ ] 10 2747 0272 3638 17 3 17, 3, 248 268, 页。在线。 1996 [ ] 10 2747 0272 3638 17 3 可访问：https://doi.org/。/ - . . . 248 ,[7] N. Naik、J. Philipoom、R. Raskar和C. Hidalgo，“Streetscore——预测一百万处街景的感知安全性”，载于《IEEE计算机视觉与模式识别会议研讨会论文集》，2014年，第779–785页。[8] R. Meltzer，“绅士化与小型企业：威胁还是机遇？”，《城市景观》，第18卷，第3期，第57–86页，2016年。 [在线]。可访问：http://www.jstor.org/stable/ 26328273 [9] E. L. Glaeser、H. Kim和M. Luca，“即时预测绅士化：利用Yelp数据量化邻里变化”，《经济网络电子期刊》，2018年。[10] Z. Gong、Q. Ma、C. Kan和Q. Qi，“利用街景图像对街道空间进行分类，以构建城市功能的空间指标”，《可持续性》，第11卷，6424，2019年。[11] T. T. Nguyen、Q. C. Nguyen、A. D. Rubinsky、T. Tasdizen、A. H. N. Deli- gani、P. Dwivedi、R. T. Whitaker、J. D. Fields、M. C. DeRouen、H. Mane、C. R. Lyles、K. D. Brunisholz和K. Bibbins-Domingo，“与冠心病、高血压和糖尿病相关的、基于谷歌街景得出的加利福尼亚州街区特征”，《国际环境研究与公共卫生杂志》，第18卷，2021年。[12] J. Lee、D. Grosz、B. Uzkent、S. Zeng、M. Burke、D. Lobell和S. Ermon，“从社区生成的街道级图像中预测生计指标”，载于AAAI，2021年。[13] Z. Wang、H. Li和R. Rajagopal，“Urban2vec：融合街景图像与兴趣点的多模态城市邻里嵌入”，2020年。[14] P. Jenkins、A. Farag、S. Wang和Z. Li，“通过多模态嵌入进行空间数据的无监督表示学习”，载于第28届ACM信息与知识管理国际会议论文集，2019年，第1993–2002页。[15] T. Huang、Z. Wang、H. Sheng、A. Y. Ng和R. Rajagopal，“从多模态多图中学习街区表征：图像、文本、移动性图及其他”，2021年。 [在线]。可访问：https://arxiv.org/abs/ 2105 02489 。T. Gebru、J. Krause、Y. Wang、D. [16] Chen、J. Deng、E. L. Aiden和L. Fei-Fei，“利用深度学习和谷歌街景估算美国各地街区的人口构成”，《美国国家科学院院刊》，第114卷，第50期，第13 108–13 113页，2017年。[17] J. Feng和Z.-H. Zhou ，“深度多实例多标签网络，”发表于*AAAI* ，2017。

[19] P. H. O. Pinheiro 和 R. Collobert，“从图像级到像素级的卷积网络标注”，*2015 IEEE Conference on Com- puter Vision and Pattern Recognition (CVPR)*，第 1713–1721 页，2015 年。[20] X. Wang、Y. Yan、P. Tang、X. Bai 和 W. Liu，“重新审视多实例神经网络”，*ArXiv*，第 abs/1610.02501 卷，2018 年。[21] M. Ilse、J. M. Tomczak 和 M. Welling，“基于注意力的深度多实例学习”，*ArXiv*，第 abs/1802.04712 卷，2018 年。[22] G. Koch、R. Zemel、R. Salakhutdinov 等，“用于单样本图像识别的孪生神经网络”，载于 *ICML deep learning workshop*，第 2 卷。里尔，2015 年，第 0 页。[23] K. He、X. Zhang、S. Ren 和 J. Sun，“用于图像识别的深度残差学习”，2015 [ ] 1512 03385 。在线。[24] 可访问：https://arxiv.org/abs/ 。J. Hwang，“没有隔离的绅士化？一个日益多元化城市中的种族、移民与更新”，《城市与社区》，第 19 卷，第 538–572 页，2020 年。[25] M. Ilse、J. M. Tomczak 和 M. Welling，“基于注意力的深度多实例学习”，2018 年。 [在线]。可访问：https://arxiv.org/ abs/1802.04712[26] V. Guerrieri、D. A. Hartley 和 E. Hurst，“内生性绅士化与住房价格动态”，*ERN*：行为经济学（主题），2010 年。[27] J. Brown–Saracino，“阐释关于绅士化与日益加剧的收入不平等的分歧路径”，《社会学评论》，第 43 卷，第 515–539 页，2017 年。[28] M. Zuk、A. H. Bierbaum、K. Chapple、K. G´ orska 和 A. Loukaitou-Sideris，“绅士化、迁离与公共投资的作用”，《规划文献期刊》，第 33 卷，第 31–44 页，2015 年。[29] A. S. Schnake-Mahl、J. L. Jahn、S. V. Subramanian、M. C. Waters 和 M. C. Arcaya，“绅士化、街区变化与人口健康：系统综述”，《城市健康杂志》，第 97 卷，第 1–25 页，2020 年。