arXiv:2303.02230v2 [cs.CV] 6 Jun 2023

中国建筑楼面面积：数据集与学习流程

PETER EGGER，应用经济学讲席教授，苏黎世联邦理工学院；CEPR；CESifo；瑞士诺丁汉全球化与经济政策研究中心（GEP）SUSIE XI RAO，应用经济学讲席教授；计算平台研究所，苏黎世联邦理工学院，瑞士SEBASTIANO PAPINI，应用经济学讲席教授，苏黎世联邦理工学院，瑞士

本文首次尝试测量40个中国主要城市的建筑物楼面面积（即建筑物轮廓和高度）。其目标是最大限度地扩大城市覆盖范围，并最终提供纵向数据。为此，需要基于具有中等至精细粒度的图像开展研究，因为更大范围的城市横截面数据以及更长的城市时间序列目前仅以这种格式提供。我们采用多任务目标分割器方法，在同一框架中并行学习建筑物轮廓和高度：（1）确定任何建筑物所覆盖的表面积（即占用土地的平方英尺）；（2）利用从不同角度获取的建筑物多图像表征来确定建筑物高度，进而计算楼面面积。我们使用Sentinel-1和Sentinel-2卫星图像作为主要数据来源。这些数据的优势在于横截面和纵向覆盖范围广泛，且可自由访问。我们详细介绍了数据、算法和评估方法。此外，我们还分析了参考数据的质量及其在以最小误差测量建筑物楼面面积中的作用。我们以深圳为案例，使用多任务学习器开展了广泛的定量和定性分析。最后，我们研究了研究结果（在像素层面和城市区域汇总层面）与夜间灯光数据之间的相关性，以评估该方法在研究城市发展方面的价值。我们的数据和代码库可通过https://gitlab.ethz.ch/raox/urban-satellite-public-v2公开访问。

CCS 概念：• **计算方法** → **目标检测**；• **应用计算** → **经济学**。

其他关键词和短语：建筑物轮廓、建筑高度、城市发展、多任务学习

1 引言

研究宏观和中观区域发展的社会科学面临的一项关键挑战，是完全缺乏（或至少极度缺乏）高质量数据的大规模横截面覆盖和时间覆盖范围。人们希望利用这类数据来确定此类发展的程度、地点，以及最终决定其发展的因素和后果。在理解发展这一议题中备受关注的一个领域，是人群在空间中的集聚所发挥的作用：在宏观层面，体现为城市、大都市区及其腹地之间的集聚；在中观和微观层面，则体现为城市和街区内部的集聚。在宏观、中观和微观空间尺度上准确衡量集聚，对于了解其在发展进程中的根本作用、资源利用与节约情况，以及评估基础设施发展总体上的效益，尤其是不同类型基础设施发展的效益等，都十分重要。

然而，覆盖众多甚至理想情况下覆盖所有不同尺度街区的大型数据集，以及用于比较的其他数据，并不总是可用。如果这些数据可用，也主要集中于发展进程已经成熟的地区，即美国 [3, 30] , 德国 [1], 或英格兰 [18] 等工业化国家。鉴于近年来人们对发展中国家 [5, 20, 35], 城市的关注，希望能够在较长时间跨度内获得此类数据，涵盖那些处于经济增长和追赶轨迹陡峭上升阶段的国家，因为在发生重大变化的背景下，因果关系可能更容易得到识别；同时，这些变化在以标准化方式收集的数据中也应具有足够的异质性。本文旨在通过利用计算机视觉的现代方法，结合来自各种来源的遥感卫星影像，对中国大量城市的住房存量实现这一目标。

中国是非常适合这一目的的经济体。首先，自1978年对外贸易开放以来，尤其是20世纪90年代初以来，中国的增长速度远高于数据丰富的工业化国家。因此，中国在遥感数据已经可用的时期实现了非常迅速的增长。其次，中国经历了从农业经济向全球领先制造业经济体的重大结构转型，近来某些服务业也大幅增长；其大学总体入学率和各领域入学率的变化便体现了这一点。伴随结构转型，中国见证了人口在农村地区与集聚区（城市和都市区）之间分布的转变。后者部分反映在户籍制度（称为户口）的改革中；该制度本质上赋予人们合法获得住房所有权、教育或卫生系统服务的权利，同时影响范围也更广（例如，在集聚区从事非法流动人口就业，只要有利于经济发展，就会在一定程度上得到当局的容忍）。因此，在过去四十年中，中国城市的楼面面积需求与供给大幅增长。

谈到集聚时，人口统计学家和其他社会科学家通常会想到空间中的人口密度。在大多数经济体和司法辖区，关于中观或微观空间居民的详细数据少得出奇，而且往往只能从家庭以某种方式披露其居住情况的来源中获取（例如官方工作记录、正式购房记录等）。然而，即使在美国，非法移民也可能在黑市工作并租房，而不留下居住记录；在一些集聚区，未登记人员的数量可能相当庞大。此外，也可以从生产的角度谈论集聚，此时人们关注的是生产所占据的空间。在这种情况下，官方数据也可能因诸多原因而存在问题。例如，公司登记数据通常列出母公司及其地址。这恰恰给那些据称占生产大部分份额的公司带来了问题，即多工厂和多地点单位。此外，此类记录可能只列出核心业务。然而，例如大型制造企业的总部往往位于市中心，而生产运营则位于城市的郊区或腹地。这表明，与生产相关的集聚方面可能测量不准，家庭与居住相关的集聚方面也是如此。总之，这使得人们需要另辟蹊径来测量集聚现象。

本文为这一领域提供了第一个里程碑，发布了一套关于建筑楼面面积（即建筑物轮廓和高度）的丰富开源数据集。当前数据集涵盖基础设施在2014年至2017年间发生重大变化的40座城市。我们的目标是扩大数据集的范围，包括横截面覆盖范围和时间覆盖范围。我们构建了一套处理流程，在多任务学习器中联合学习建筑物轮廓和高度。这一尝试面临以下挑战。

(1) 数据源与采集。 高分辨率卫星影像以及关于影像内容（例如建筑物、建筑物高度等）的可信参考数据，只在少数城市（甚至仅在这些城市的部分区域）存在。因此，研究人员通常需要对卫星影像和参考数据进行大量预处理。此类预处理的质量必然会影响检测结果和模型有效性。**(2) 建筑物轮廓与高度的联合学习方法**。 深度学习方法从根本上改变了人们使用遥感数据的方式。新的技术不断涌现，可通过多任务或两阶段学习流程结合不同目标。然而，对于研究城市变化所需的楼面占用测量与发展评估而言，此类联合方法的优势尚未得到评估。

(3) 预测质量评估。如何解读和验证有关城市变化以及建筑物位置与高度的模型预测，是这一领域的关键问题。仅依据单一指标或模型评估预测可能会产生误导。综合考虑不同数据源（例如人口、基础设施密度和夜间灯光数据）、指标与模型，对于全面了解城市动态至关重要。

本文提供了一个关于中国建筑楼面面积扩张的丰富数据集。我们首先收集了覆盖40座城市、分辨率为 10×10 米的哨兵卫星日间卫星数据，以及相应的三维参考数据。随后，我们采用多任务学习方法，在同一框架下估算建筑物轮廓和高度：(1) 确定建筑物覆盖的表面积（即占用土地的面积）；(2) 根据建筑物图像确定其高度。最后，我们评估了多种在计算机视觉领域广受欢迎的 U-Net 架构，以预测这两个指标。我们以深圳为案例，实证考察楼面面积预测的质量，并特别评估楼面空间密度与夜间灯光亮度数据之间的相关性。我们的数据和代码库可通过 https://gitlab.ethz.ch/raox/urban-satellite-public-v2 公开获取。

主要贡献可概括如下。我们提供了一个详细的处理流程，用于研究中国40座城市的建筑楼面面积。我们通过使用夜间灯光数据进行稳健性检验，说明如何利用所得的楼面空间密度预测结果分析城市内部的微区域集聚模式。例如，这有助于探讨城市是单中心还是多中心结构的问题，也可能帮助识别城市核心区随时间变化的位置。本研究为我们今后的工作奠定了基础，以便在更长的时间跨度内确定所有中国城市楼面面积的演变。

2 建筑占地轮廓识别与建筑高度估算

2.1 数据准备

我们使用哨兵卫星影像作为主要数据源，并受益于其开放访问、免费使用的政策。影像集包括 Sentinel-1（合成孔径雷达，合成孔径雷达）和 Sentinel-2（多光谱光学）产品，两者均覆盖近乎全球的范围，并具有较高的空间分辨率。我们通过 Google Earth Engine API 下载数据并处理影像，涵盖中国40座城市；这些城市在2017年至2022年间发生了显著的基础设施变化。[1] 这些城市是根据众包参考数据选出的，这些数据在学术遥感领域被广泛使用，参见 Cao 和 Huang [6]。

对于Sentinel-2光学图像，我们选用2A级产品，该产品提供经过正射校正的大气底层（BOA）地表反射率数据。我们使用的波段包括B2（蓝）、B3（绿）、B4（红）和B8（近红外），空间分辨率为 10×10 米。为减少云遮挡的影响，我们首先筛除云覆盖率超过60%的图像，并进一步利用额外的哨兵二号云概率影像集合应用云掩膜。[2] 我们根据上述条件查询2017年到2021年的数据，并对每年返回的多幅图像取平均值，从而为每座城市覆盖感兴趣区域的图像生成年度单幅影像。[3]

[1]我们样本中的40座城市包括保定、沧州、长沙、成都、佛山、福州、贵阳、香港、呼和浩特、惠州、嘉兴、廊坊、临沂、南昌、宁波、秦皇岛、泉州、三亚、绍兴、沈阳、深圳、苏州、台州（浙江）、太原、唐山、温州、潍坊、威海、芜湖、无锡、厦门、西安、徐州、烟台、扬州、银川、枣庄、中山、珠海和镇江。[2]参见 https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless（最后访问日期：2023年3月3日）。[3]请注意，我们在第3节报告的结果是使用2017年的卫星图像得出的，因为我们推测众包参考数据来自2017年。在未来的类似研究中，我们将使用2017—2021年的数据，尤其是在获得更高质量的地面实况数据后，具体讨论见第6节。请注意，现阶段我们不使用2022年的数据，因为2022年对校准（波段相关偏移）进行了一些重大调整，参见 https://developers.google.com/earth-engine/datasets/catalog/sentinel-2（最后访问日期：2023年4月23日）。

对于Sentinel-1雷达图像，我们选用由C波段合成孔径雷达传感器获取的VV和VH极化数据，并结合升轨和降轨方向，以收集尽可能多的信息。我们查询2015年至2021年的数据，并在每一年采用与Sentinel-2相同的影像集合归约方法。

2.2 方法：多任务学习

我们采用多任务学习方法，利用神经网络架构分割建筑物轮廓并预测建筑高度。我们采用 U-Net [31]，并搭配多种骨干网络。 [31]中所述的原始 U-Net 架构不依赖任何特定的预训练骨干网络，而是采用其自身的编码器-解码器卷积结构，并带有跳跃连接。该架构最初是为生物医学图像分割而开发的。后来，它在城市研究中逐渐流行起来，用于土地覆盖分类，或从卫星影像中提取道路和建筑物（参见 Li 等人 [25]）。我们测试了多种 U-Net 骨干网络变体，包括 EfficientNet-B4 [34],、ResNet-34 [17],和 SE-ResNet50 [21]。我们对这些架构进行了修改，使其能够接收多个输入并生成多个输出。

我们对众包数据集中的数据质量进行了人工评估。不同城市之间以及同一城市内部的数据质量都存在显著差异。我们重点关注参考数据质量相对较好的 40 座城市。这些城市对应 1035 个镶嵌瓦片，我们使用 GitLab 上的脚本从 Google Earth Engine 下载了这些瓦片（哨兵卫星 1 号和 2 号）。为准备训练数据，我们将这些图像划分为 256×256 个网格，每个网格对应相应的瓦片。随后，我们将参考数据中的逐栋建筑多边形栅格化，使其与 Sentinel 影像的 10×10 米网格分辨率一致。我们剔除了建筑物像素占比低于 10% 的瓦片。训练集/验证集划分比例为 90%/10%。测试集则使用深圳市的所有图像。选择深圳市的原因如下：（1）过去三十年间，深圳经历了大规模的基础设施和城市变化；该市还通过深圳市政府数据开放平台 [33]提供了许多开源数据，例如企业注册名录和三维建筑点。（2）与其他城市相比，先前研究已对深圳开展了深入研究，并对其从 1986年 到 2017 [39] 的三维建筑楼面面积进行了建模。后者可用于与我们的结果进行比较，也为模型验证提供了理想基础。这一点很重要，因为我们计划将处理流程扩展到其他缺乏参考数据的主要中国城市。因此，我们需要一个在样本外城市中测试时依然稳健的处理流程。

我们在一个多任务模型中学习建筑物轮廓和高度。对于建筑物轮廓预测，我们预测感兴趣区域是否被任意建筑物覆盖。对于建筑高度，我们采用基于回归的模型，预测并学习区域被建筑物覆盖时的高度。为使训练更稳定，我们不直接预测建筑物的绝对高度，而是预测归一化高度。归一化时，我们将高度值除以 400 米，因为数据集中所有建筑物的最大高度都低于 400 米。我们使用交叉熵进行建筑物占地预测，并使用平滑L1损失进行高度预测。计算高度预测的损失时，我们会屏蔽不属于建筑物的像素。最终损失是两项损失的线性组合，其中建筑物占地预测损失的权重为 0.1，建筑高度回归损失的权重为 1。请注意，我们尝试了第3节.2.2中讨论的数据增强技术，以及用于区域分类和高度回归的不同任务系数。对于建筑轮廓预测，我们报告精确率、召回率和Dice系数（精确率和召回率的调和平均数）；对于高度，我们报告平均绝对误差（MAE）和均方根误差（RMSE）。

相应神经网络架构的重要超参数包括多步学习率：初始值为 0.001，并在第 50 个训练轮次衰减为原来的 0.1 倍。我们在一台配备 64 个 AMD EPYC 7313 16 核处理器、504 GB 内存以及一块图形处理器（NVIDIA GeForce RTX 3090，24GB 内存）的机器上对算法进行基准测试，平均每个训练轮次的训练时间约为 15 秒。我们对所有架构训练 100 个训练轮次。

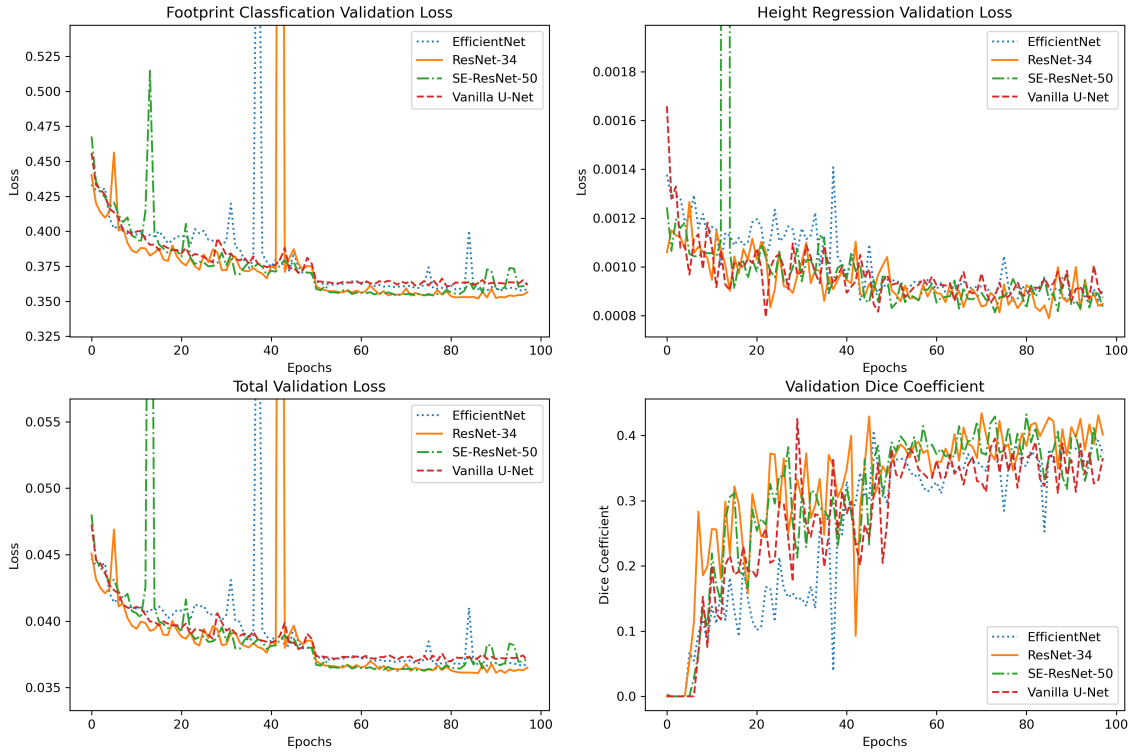


图1. 验证集上不同 U-Net 骨干网络的结果。

3 结果与评估

本节首先在验证集上评估采用不同骨干网络的 U-Net 在建筑轮廓分类和高度预测方面的性能。随后，我们反思模型的选择，并分享两阶段模型、对数归一化和数据增强方面的消融研究结果。最后，我们针对深圳市，提供 50 个基于所提出模型与参考数据对楼面面积进行估算的案例研究。

3.1 定量分析

骨干网络选择。 我们将 Sentinel-1 和 Sentinel-2 的叠加影像作为多任务学习器的输入（关于数据源的消融研究，参见附录 B）。我们首先在多任务学习设置下考察不同的 U-Net 骨干网络。图 1 展示了这些不同 U-Net 骨干网络在验证集（占数据的 10%）上的性能，包括 Dice系数和相应的损失。根据这一比较，我们选择 ResNet-34 作为 U-Net 骨干网络。其模型架构见附录 A 的图 9。

建筑物轮廓和高度的评估指标。对于建筑物轮廓，我们在像素级评估模型性能。所使用的评估指标包括精确率、召回率和 Dice系数。精确率和召回率仅根据真正例（建筑物）计算；Dice系数（亦称 F1）是精确率和召回率的调和平均数。在学习建筑高度时，我们计算平均绝对误差（MA）、平均相对误差（MRE）和均方根误差（RMSE）。MAE 是预测值与

	建筑占地面积			高度	
	精确率	**召回率**	**Dice（F1）**	**MAE**	**RMSE**
ResNet-34	0.5881	0.2959	0.3872	3.4566	9.8385

表 1。采用 ResNet-34 骨干网络的多任务学习器 U-Net 在可信测试集（深圳市）上的结果。

实际高度之间绝对差值的平均值，而 MRE 是绝对差值与实际高度之比的平均值。RMSE 衡量数据集中预测值与实际值之间的平均偏差。MAE 和 MRE 是衡量模型性能的绝对指标，表示误差的平均幅度。RMSE 对较大的误差赋予更高权重，因此对离群值（例如较高的建筑物）更敏感；MAE 则对所有误差一视同仁，受离群值的影响较小。

深圳的建筑高度。 在 Yu 等人 [39], 的研究中，作者报告了深圳市 2017年三维建筑楼面面积的参考数据。我们使用他们推荐的分类方案，分析了我们在深圳市的结果（测试数据）。他们拥有由深圳市规划和自然资源局提供的 567,000 栋建筑的高质量（非公开）参考数据。根据这一分类，建筑物分为四类：低层建筑（1–3 层）、多层建筑（4–6 层）、中层建筑（7–9 层）和高层建筑（≥10 层）。2017年，这四类建筑的占比分别为 51.9%、31.0%、12.5% 和 4.5%。请注意，这一分布反映在可信参考数据的图 2（左图）中。

深圳的可信参考数据。为了获得可靠的测试集，我们选取了深圳的可信参考数据，并将其与 Maxar 的 0.5 米高分辨率卫星影像进行人工比对 [29]。我们认为，建筑物标注正确率大致超过 90% 的区域相对可信，并将其用于评估。

我们在深圳的预测结果。表 1 展示了表现最佳的模型——采用 ResNet-34 骨干网络的 U-Net——在预测深圳可信参考数据时的结果。平均绝对误差为 3.4566，这表明模型预测值与实际高度值的绝对差平均约为 3.5 米（约 1 层）。均方根误差为 9.8385，表示预测高度值与实际高度值之间差值平方的平均值的平方根，可用于衡量模型误差的典型幅度。在本例中，这意味着模型预测的误差平均约为 9.8385 米。均方根误差对较大的误差（此处为较高建筑物的误差）较为敏感，因此相比平均绝对误差会对其施加更大的惩罚。平均绝对误差和均方根误差表明预测存在一定程度的误差；由于均方根误差对离群值较为敏感，其数值更高。

我们按照 Yu 等人 [39] 的分类方法对建筑物进行分类，并计算各类别的平均相对误差。低层建筑（1–3 层）的平均相对误差为 1.3159，多层建筑（4–6 层）的平均相对误差为 0.9405，中层建筑（7–9 层）的平均相对误差为 0.8842，而高层建筑（≥10 层）的平均相对误差为 0.8511。各类建筑物的平均相对误差表示预测高度与实际高度之间的平均相对差异。这些数值表明，模型倾向于高估 低层建筑的高度，预测高度平均比实际高度高出约 31.59%。模型对多层建筑、中层建筑和高层建筑的预测则更准确，预测高度与实际高度的平均差异分别约为 6.05%、11.58% 和 14.89%。

为展示预测高度与参考数据中建筑物高度分布之间的差异，我们在图 2 中呈现了深圳每个像素的预测高度和目标高度分布直方图。横轴为建筑高度，单位为米；纵轴为像素计数。两个直方图的上限均为 100 米。根据平均绝对误差和平均相对误差得分，我们看到

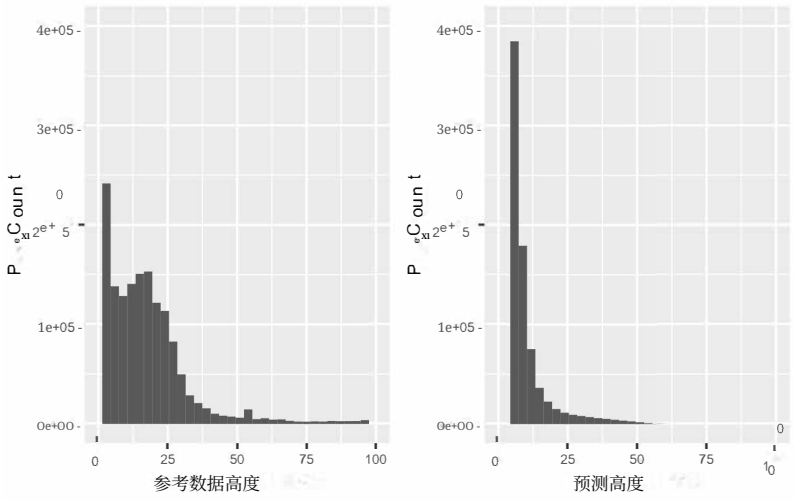


图 2. 深圳可信参考数据与预测结果中像素级高度分布的直方图（最高至 100 米）。

我们的模型低估了平均高度，同时高估了 1—3 层低层建筑的高度，这一点也反映在四类建筑类别的平均相对误差得分中。我们的预测遗漏了高度超过 60 米的高层建筑。不过，正如我们稍后将在案例研究（第3.3节）中看到的，参考数据也并非完美。因此，我们使用其他类型的地理空间数据开展稳健性检验，例如夜间灯光数据——这类数据经常用于城市经济学——以进一步验证所提出模型作出的预测（见第4节）。

3.2 模型架构讨论

3.2.1 两阶段模型。 Wen 等人 [36], 此前利用其获取的资源三号卫星多视角影像商业数据集，采用两阶段模型来建模三维建筑空间。Cao 和 Huang [6], 研究了 42 座中国城市，并获取了与 Wen 等人 [36] 相同的数据源，进而从中学习建筑物的精确高度。不过，需要注意的是，这类多视角影像在中国并非公开可用，且获取成本高昂。此外，这些数据的时间覆盖范围非常有限。

此处所讨论框架的目标是：(i) 覆盖中国所有城市地区；(ii) 覆盖远超十年的时间跨度；(iii) 基于开放且免费的数据源。本文提供的证据表明，只要重点关注边长为 10—20 米的网格单元的中观至宏观层面的区域汇总，就有望实现这一目标。这一分辨率足以满足大多数涉及人口、城市、区域或交通研究重点的学术研究需求。除目标分辨率不同之外，Wen 等人 [36] 以及 Cao 和 Huang [6] 的方法与我们的方法之间的另一个关键区别在于，我们在单阶段模型中结合了建筑物轮廓和建筑高度的学习。

在缺乏多视角影像的情况下，关于我们的两阶段模型，我们作出如下推测：与上述工作中的实现相比，我们的实现效率更高；两阶段模型的性能与单阶段模型不会有太大差异。作为稳健性检验，我们在

哨兵卫星 1+2 的最佳设置上开展了两阶段实验，结果证实了我们的主张。附录 C 中给出了实验设置和结果。

3.2.2 输出归一化与数据增强。我们尝试了多种方法对多任务学习器的输出进行归一化。对输出进行归一化会改变建筑物轮廓分类任务和建筑物高度回归任务的重要性。微调损失函数有助于确保训练过程中对这两项任务给予适当的重视。我们没有在损失函数中使用 400 米的缩放因子（即数据中的最大建筑高度），而是使用了对数尺度高度。此外，我们还尝试了多个系数（0.05、0.1、0.2、1、10），用于调整建筑物高度预测任务相对于轮廓预测任务的权重。

此外，我们还开展了数据增强实验，借鉴了 Liu 等人 [28]提出的技术。这些作者在实验中发现，像素变换的性能与基线相近或略差。他们还发现，随机旋转、X/Y 剪切以及随机擦除部分像素，能够通过人为扩大训练数据集来减少过拟合。我们据此设计了以下实验：实施范围为 [-10, 10], 的随机旋转、包含 X/Y 剪切、平移和旋转的随机仿射变换，以及随机掩蔽部分像素。在随机掩蔽实验中，我们只对 Sentinel-1 数据进行随机掩蔽，因为 Sentinel-2 数据中可能已经存在因云层遮挡而随机缺失的像素。根据这些实验，我们得出了以下见解。

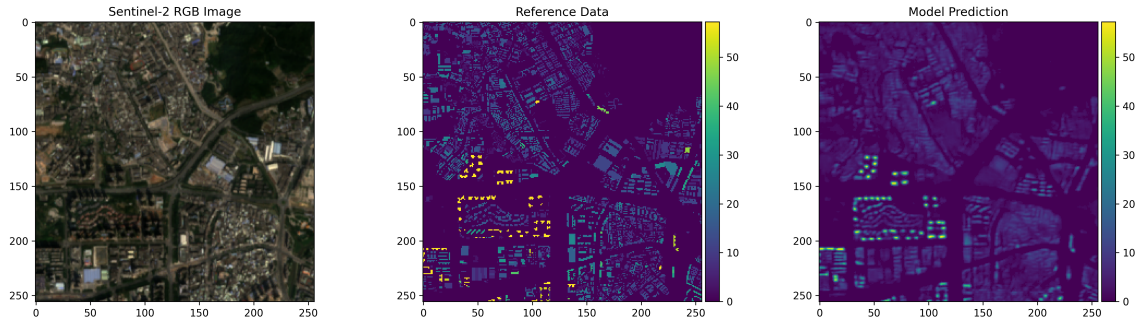
值得注意的是，将损失除以 400 进行缩放后，数据增强并未提升性能。然而，增强有助于采用对数尺度损失的模型。仅使用旋转时表现最佳。使用仿射变换也提升了性能，但不如仅使用旋转。采用对数尺度损失时表现最佳的是 SE-ResNet-50-log-reg-0.05-rotation（使用 SE-ResNet-50 作为骨干网络，并以旋转作为数据增强，为回归任务设置 0.05 的系数）。然而，正如我们在深圳市评估中所展示的（见附录 D），使用 Sentinel 数据时，缩放因子为 400 米的损失表现优于对数尺度损失加旋转。

3.3 案例研究

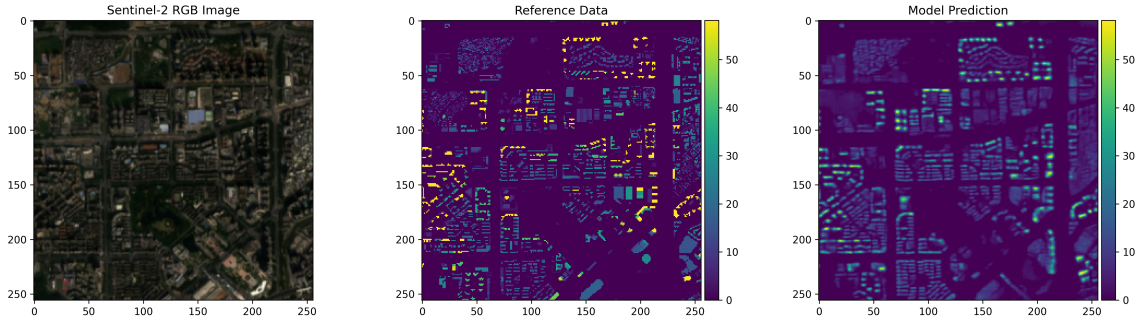
为了进一步评估与众包参考数据相比我们的预测质量，我们对随机选取的 50 个 256×256 瓦片进行了人工分析。在其中一个瓦片中，每个像素表示深圳地区预测结果中的一个 10米尺寸。

我们发现了两个主要挑战。首先，就建筑物轮廓而言，我们抽取的所有样本图块都成功识别了建筑物所覆盖的区域；然而，在38%的样本中，模型难以准确区分彼此距离较近的小型建筑物，导致高估建筑占地面积覆盖范围，如图3(a)所示。我们认为，这主要源于哨兵卫星的分辨率与城市拓扑之间存在固有的局限性。在深圳选定的参考区域内，约91%的建筑物与最近邻建筑物的距离小于10米，相当于单个像素的大小。此外，54.6%的建筑物与至少另外两栋建筑物相距不到10米。因此，我们建议在聚合级别（见第4节）而非像素级评估预测结果，以获得更有意义的见解。

其次，从高度估计的角度来看，尽管我们的预测有92%能够准确呈现建筑高度的层级结构，但如图3（b）所示，它们始终低估高层建筑的高度。此外，在分析过程中，我们发现模型确实能够正确识别出这些建筑物



(a) 高估建筑轮廓覆盖率。



(b) 高估高层建筑的建筑高度。

图 3. 深圳采样瓦片对比。坐标轴表示像素的尺寸，颜色表示建筑物的高度。(a)：高估建筑轮廓覆盖率；(b)：高估高层建筑的建筑高度。

正在施工的建筑物，这些建筑物出现在三张抽样图像中，但尚未记录在参考数据中。这一发现表明，这类未标记建筑物的存在可能会影响评估过程。我们已在 GitLab 项目仓库中提供性能最佳的模型、测试图像以及案例研究的图块。

4 聚合网格中的预测质量及基于夜间灯光数据的稳健性检查

在过去二十年里，夜间灯光（NTL）卫星（遥感）影像已成为城市与区域研究、人口学和经济学中的常用测量工具。夜间灯光数据被广泛用于测量城市结构[16, 24], 人口密度 [11, 41], 经济活动和增长 [4, 8, 14, 19, 40], 以及贫困与财富 [2, 10]。夜间灯光数据之所以能用于测量和估算一系列社会人口与社会经济因素，主要有两个原因：（i）它们在全球范围内均可获取，且时间跨度长达数十年；（ii）它们在空间和时间上的测量精确度具有可比性或完全一致。相比之下，更精确的人口学、经济活动或城市与区域发展指标往往测量频率低得多（例如，每十年才公布一次的人口普查数据），覆盖范围具有选择性（抽样时设有报告门槛，或只有部分司法辖区报告特定数据），且不同司法辖区的报告质量和时间并不一致。与本文生成的数据相比，夜间灯光数据的一个缺点是，其空间分辨率不如哨兵卫星及其他日间卫星影像那样精细。Henderson 等人 [19]

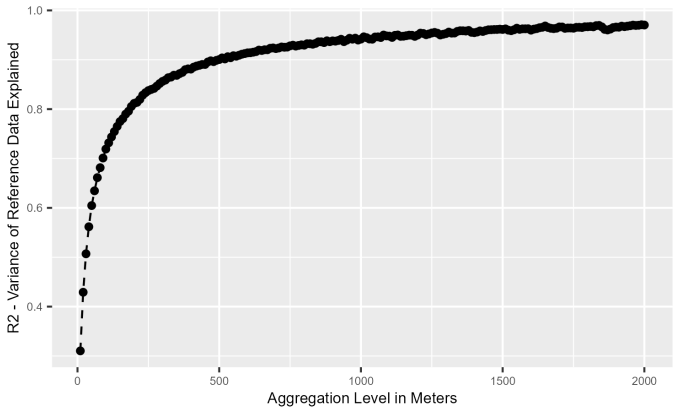


图 4. 聚合层级不同的情况下，参考数据中建筑高度由模型预测的建筑高度解释的方差份额。横轴：网格单元边长（米）。纵轴：相关模型基于模型预测结果，对可信参考数据中建筑高度变异性的解释比例。

使用了近 1000×1000 网格分辨率，而 Gibson 等 [14], 作为较新的例子，则使用了分辨率为 700×700 米的夜间灯光数据。

NTL 数据的广泛使用及其可获得性，使其成为与我们的楼面面积预测结果进行比较的自然候选数据源。原因在于，人们通常会预期，人口密集地区（即拥有高层建筑以及单位覆盖土地面积楼面面积密度较高的地区）在夜间会发出更多光线。为验证这一推测，我们将我们的楼面面积预测结果与 2019 年深圳市珞珈一号01星夜间灯光数据中的夜间灯光辐射进行了比较，所用网格边长为 120×120 米。在研究建筑楼面面积与夜间灯光数据之间的相关性之前，我们先考察了聚合网格中预测结果与可信数据之间的相关性（见第 4.1 节）。这使我们能够在一个尺度上解读预测质量，而基于 NTL 数据的城市发展指标通常正是在这一尺度上计算的。随后，在第 4.2 节中，我们比较了表现最佳的模型、参考数据和夜间灯光数据对预测结果的建筑占地面积和高度覆盖情况。我们通过将模型得到的楼面面积估计值与 NTL 数据进行比较，评估前者的优势。具体而言，最大的优势显然在于，本文提供的预测结果揭示了 NTL 数据无法呈现的城市密度，因为这些预测结果受到建筑物以外其他来源的光排放影响（例如交通，或除建筑物之外的照明基础设施）。

4.1 聚合网格中的预测与可信参考数据

图4展示了2019年深圳地区，预测建筑高度与参考数据之间简单相关模型的均方误差随空间聚合（空间网格单元的边长）变化的情况。此处所示数据仅包含深圳参考数据中可信的部分。横轴表示网格单元的边长，每个点代表增加10米。证据表明，从 10×10米网格开始，随着空间分辨率降低，模型的预测能力显著提升。请注意，在 10×10米网格上，我们的模型仅能解释参考数据中真实建筑高度约31%的变化。在 200×200米网格上，解释能力提升至80%以上。经过 2000×2000米空间聚合后，决定系数趋近于98%。这反映了所提出的模型在空间中观尺度上预测实际建筑高度的效果。鉴于第3.3节讨论的哨兵卫星影像质量，人们通常会关注楼面空间密度在中度聚合级别（例如 10×10米网格）上的表现，因为后者能提供足够精确的城市发展指标。

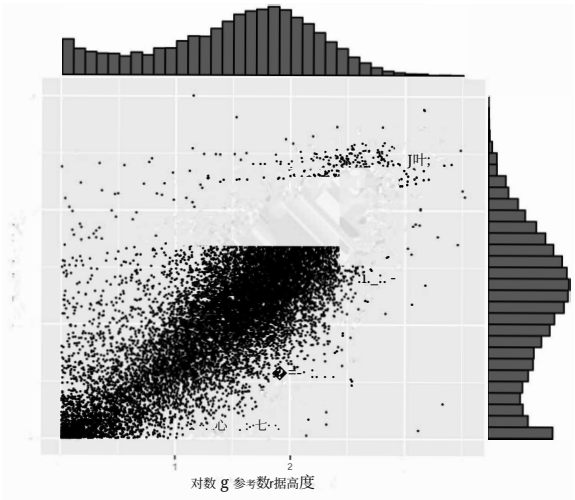


图5. 聚合至200米网格单元的对数平均高度加一散点图。可信参考数据位于横轴，预测高度位于纵轴。图中仅显示两个维度均为非零值的网格单元。

我们在图5中展示了可信参考数据中预测高度与目标高度在对数尺度上的相关性，其中高度分布已聚合至 200×200米网格单元。观察该图可见，这印证了图4所示的强相关性。

4.2 与夜间灯光数据的相关性

在此，我们考察深圳的所有参考数据点，以及模型对整个城市区域的预测结果。在图6所示的建筑轮廓预测中，我们的模型覆盖了当前参考数据中缺失的所有城市区域。

图7展示了深圳三组数据样本的比较，包括深圳的所有参考数据、表现最佳模型的预测结果，以及珞珈一号01星夜间灯光数据 [7] 辐射强度。为与夜间灯光数据保持一致，参考数据和我们的预测结果均以聚合至统一 120×120米尺度的网格单元平均高度表示。图示清楚表明，这三组数据描述了相似的城市结构。夜间灯光辐射与预测的区域建筑体量之间存在稳健的相关性。

图8展示了我们的预测值乘以相应相关系数后，与深圳的所有参考数据（上图）及夜间灯光数据（下图）之间的差异。两种差异均按其标准差缩放并取对数。下图还显示了 OpenStreetMap (OSM) 中的公路（黑色）、部分基础设施要素（白色）以及高尔夫球场（绿色）。红色区域表示，与参考数据相比，我们的预测值偏高，而夜间灯光数据的预测值偏低；紫色区域则表示，相对于参考数据，预测值偏低，而夜间灯光数据的预测值偏高。上图显示，香港（位于下方中部）等区域，以及地图边缘的外围区域，虽然夜间灯光数据清楚地显示了这些区域，但参考数据中却没有记录，而我们的预测则涵盖了这些区域。下图揭示了夜间灯光数据的一些不足之处。夜间灯光数据会严重高估夜间有照明的大型基础设施，例如港口、海关检查站、机场和公路，以及高尔夫球场等休闲设施。

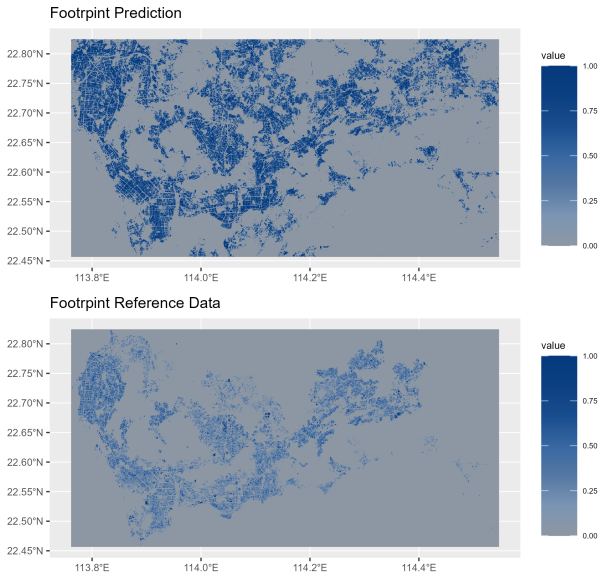


图6. 深圳2019年所有参考数据和模型预测的建筑占地面积。

相反，夜间灯光数据会低估中央商务区 (CBD) 中的建筑体量；这些区域高层建筑较多，其他城市环境也往往十分密集。

建筑体量与夜间灯光强度之间的偏差凸显了采用此类稳健处理流程的必要性，以改进迄今主要依赖夜间灯光数据的研究。从图8可以看出，我们的指标与夜间灯光数据反映了城市环境的不同方面。前者能更好地展示建筑物和楼面空间密度，而后者则着重呈现大型基础设施和道路。尤其是，人口密度的估算以及需要更精细空间分辨率的建筑结构指标，若仅基于夜间灯光数据，则难以避免较大偏差。总之，通常由卫星影像获取的夜间灯光数据能够揭示经济活动活跃的区域，但可能无法准确反映基于居住地的指标，以及与建筑物和楼面空间相关的城市密度指标。

5 相关工作

在此，我们总结了遥感领域相关的早期研究，这些研究启发了我们的工作 .

二维制图。 自 1972年首批陆地卫星任务启动以来，遥感方法长期应用于生态学、冰川学和地质学等地球科学领域。进入 21 世纪后，得益于数据质量的提升和人工智能方法的改进，人们重新燃起了对利用遥感研究城市结构的兴趣。绘制全球土地利用类别地图逐渐成为可能，不过分辨率相对较低，仅为 500米网格 [32]。相关研究致力于绘制不同的土地利用类型，其中城市用地只是其中之一。全球制图的分辨率逐步提高到 13米 [12]和 10米 [15] 。在这样的精细度下，单体建筑变得清晰可见，因此可以合理地讨论建筑物轮廓，而不再只是土地利用分类。前一类研究中的分类基于

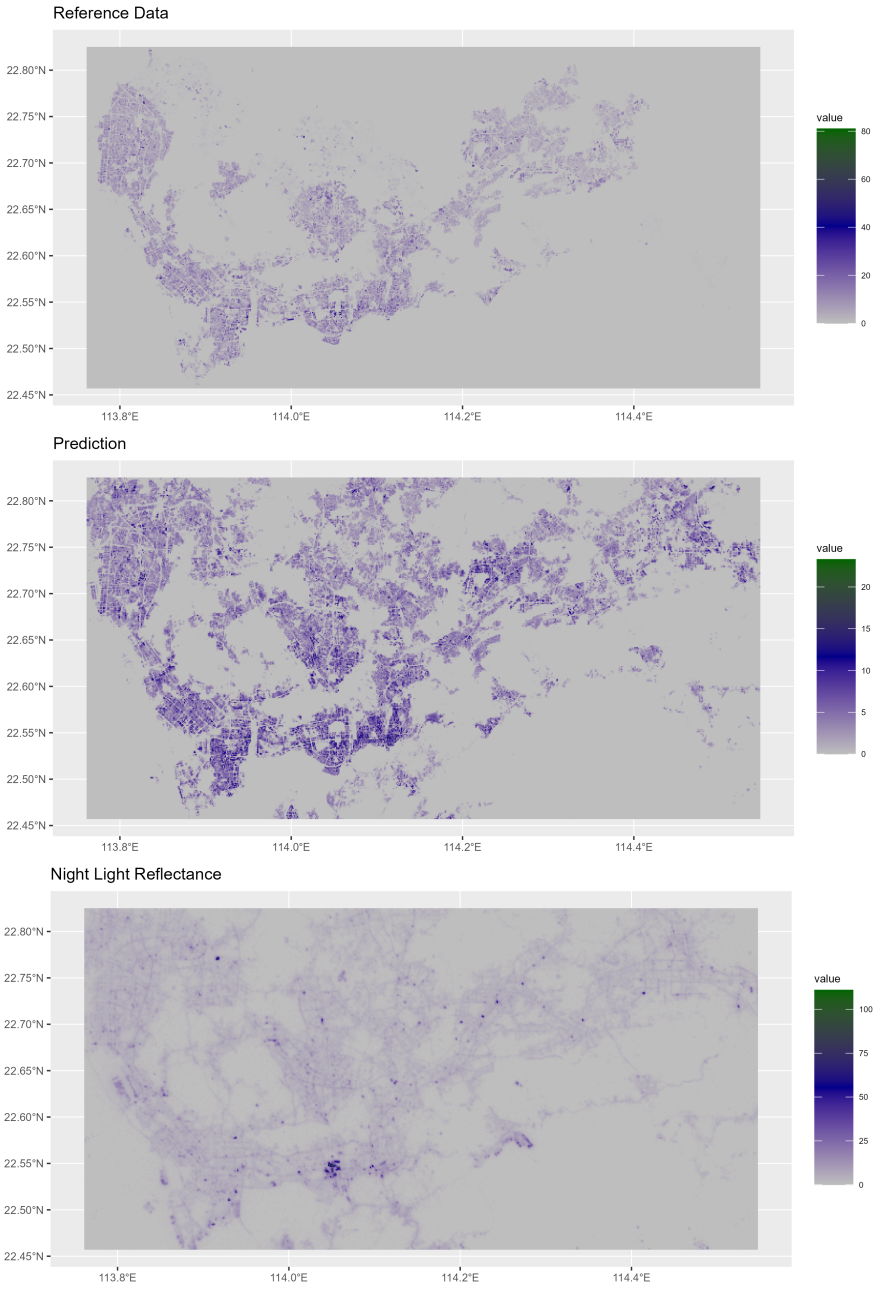


图7。深圳2019年的所有参考数据、模型预测和 LuojiaNTL 数据。

无监督学习方法，而后一类则采用随机森林框架。Wu 等人 [37] 基于 U-Net 深度学习架构，为中国选定的一组城市提供了建筑物轮廓映射。

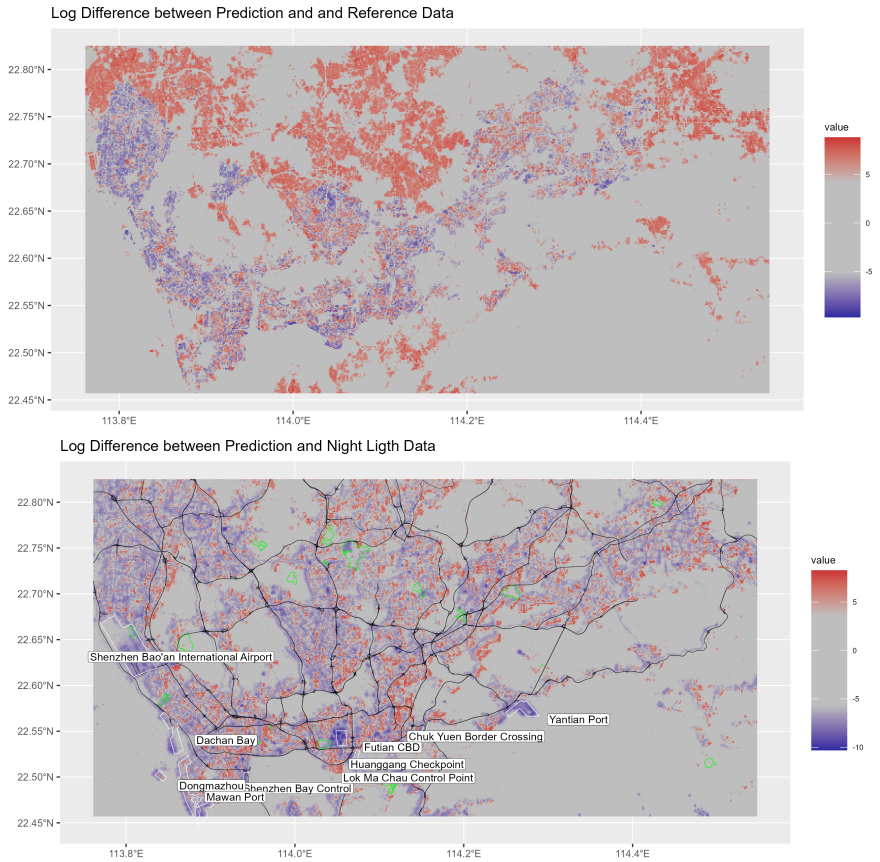


图 8. 模型预测与全部参考数据之间，以及模型预测与夜间灯光数据之间的对数差值图。红色区域表示，与参考数据相比，我们的模型存在高估；相对于我们的模型，夜间灯光数据存在低估。紫色区域表示，相对于数据，我们的模型存在低估；相对于我们的模型，夜间灯光数据存在高估。黑线：来自 OpenStreetMap（OSM）的公路；绿线：高尔夫球场；白线：夜间灯光辐射对应的基础设施选区。

建筑高度估算。过去几年，一批新的研究文献不断涌现，旨在估算建筑高度和土地利用强度。Li 等人 [26] 使用随机森林模型，对北美洲、欧洲和中国进行三维大陆制图。估算的主要输入数据源为 Landsat 8（光学数据）和 Sentinel-1（合成孔径雷达数据），最终数据集的分辨率约为 1×1 千米。Frantz等人 [13] 采用基于 Sentinel-1（合成孔径雷达数据）和 Sentinel-2（光学数据）的支持向量机方法，根据阴影长度和雷达反射率获取建筑高度估计值。后一个项目的成果是德国全国地图，分辨率为 10×10米。Cao 和 Huang [6] 进一步利用分辨率高得多的中国资源三号卫星多视角图像数据。通过深度学习方法，他们为42座中国城市制作了分辨率为 2.5×2.5米的三维地图。然而，卫星影像和最终制图成果均未以开源形式提供。近期其他关注中国建筑高度估计的研究还包括 Ji 和 Tang [22], 、Yang 和 Zhao [38], 以及 Yang 和 Zhao [38]。Yu 等人 [39] 仅针对深圳生成了建筑高度面板数据，该数据始于

1986年，止于2017年。他们以高分辨率激光雷达（LiDAR）地图为起点，并利用低分辨率的陆地卫星档案影像进行逆向更新。Li 等人 [27] 在多任务设置下使用卷积神经网络，研究全球38座城市的建筑物轮廓和建筑高度，这些城市拥有良好的三维参考数据。他们报告称，建筑高度和建筑轮廓预测的 R^2 分别比传统机器学习模型高出0.27–0.63和0.11–0.49。

6 讨论、未来工作与结论

讨论：数据质量。采用分辨率为10米的中等粒度数据，具有大范围横截面覆盖和时间序列覆盖范围广的优势。然而，这不可避免地会牺牲精确率。现阶段，我们将数据集和处理流程作为开源平台提供。未来，我们将整合包括Planet在内的各类卫星图像，这将显著提高建筑物的可见度。 [4]

讨论：更优质的建筑物轮廓和高度参考数据。我们的团队正在积极开发一款工具，用于抓取最新的高德地图三维重建数据。[5] 附录E中简要介绍了该工具。这些参考数据质量应该更高，因此在缺乏开源参考数据的情况下，可以减少图像标注工作。

讨论：标注。我们的设想是，对于没有参考数据（通过高德地图获取）的感兴趣区域，我们需要获取人工标注。我们正在与具有卫星图像标注经验的专业人员联系。我们可以借助他们的专业知识以及用于标注森林数据的工具。[6]

讨论：稳健性检验。 我们将利用人口普查数据中城市的人口统计指标，在汇总的空间层级开展稳健性检验，例如考察中国县市的建筑覆盖率。

讨论：多任务学习器中的任务系数。 未来值得探索如何从数据中学习各种任务的系数。我们目前的方法依赖于对预定义系数进行网格搜索，并根据各项任务的数值尺度来选择系数。Cao 和 Huang [6] 提出使用可学习系数（任务不确定性）对每项任务进行参数化，这一方法源自 Kendall 等人 [23]。这种方法能够让我们的网络自动、动态地学习分配给各项任务的权重，无需手动调参。

讨论：时间序列数据。 所提出方法的一项优势在于，它可以与中分辨率遥感数据结合使用，从而追踪更长时间窗口内楼面面积的变化。这不仅对于检验城市研究中的假设至关重要（涉及集聚的决定因素及其对经济结果的影响），对于公共经济学和交通经济学也同样重要（涉及基础设施发展和集聚的影响）；此外，对于统计学和计量经济学而言，如果研究旨在根据数据变化而非横截面状态识别变量之间的关系及因果效应，这一点也至关重要。我们计划结合多个来源的卫星影像，整合横截面数据并拼接时间序列数据，以生成时间跨度超过卫星生命周期的一致面板数据。

结论。 本文在生成关于建筑特定楼面面积（即建筑占地面积和高度）的丰富数据集，并建立用于训练和实证评估楼面面积的处理流程方面迈出了第一步。我们使用单一的多任务目标学习器，同时学习建筑物轮廓和高度。我们还分析了汇总到城市单元后的预测质量，以及其与夜间灯光等其他城市发展数据的相关性。本文为进一步的实证研究奠定了基础，这些研究将以定量城市经济模型的理论基础为依托。未来，我们计划将处理流程扩展至中国所有城市，覆盖2009—2022年。

[4]我们的团队得到苏黎世联邦理工学院图书馆的支持，该图书馆已从Planet Labs PBC获得学术许可。我们无法开源Planet影像，但会提供使用协调后的Planet与Sentinel影像训练的预训练模型。[5]请参阅高德地图提供的官方静态演示：https://lbs.amap.com/demo/javascript-api/example/3d/fixed-view。[6]我们在Restor的联系人，请参阅https://restor.eco/about/team/。

致谢

我们衷心感谢研究助理 Muyang Jiang、Sapar Charyyev、Yuru Jia 和 Dal Fergus 在整个项目期间给予的大力帮助和出色支持。特别感谢我们的合作者、Restor/DS3Lab 的 Josh Veitch Michaelis 博士，感谢他富有启发性的讨论和头脑风暴。

参考文献

[1] Gabriel Ahlfeldt、Stephen Redding、Daniel Sturm 和 Nikolaus Wolf。2014年。密度经济学：来自柏林墙的证据。NBER 工作论文 20354。美国国家经济研究局。https://EconPapers.repec.org/RePEc:nbr:nberwo:20354[2] Sam Asher、Tobias Lunt、Ryu Matsuura 和 Paul Novosad。2021年。高地理分辨率下的发展研究：利用 shrug 开放数据平台分析印度的夜间灯光、企业与贫困。《世界银行经济评论》35, 4（2021年），845–871。[3] Nathaniel Baum-Snow。2007年。公路是否导致了郊区化？《经济学季刊》122, 2（2007年），775–805。http://www.jstor. org/stable/25098858[4] Nathaniel Baum-Snow、Loren Brandt、J Vernon Henderson、Matthew A Turner 和 Qinghua Zhang。2017年。中国城市的道路、铁路与分散化。《经济学与统计学评论》99, 3（2017年），435–448。[5] Gharad Bryan、Edward Glaeser 和 Nick Tsivanidis。2020年。发展中世界的城市。《经济学年度评论》12, 1（2020年8月），273–297。https://doi.org/10.1146/annurev-economics[6] Yinxia Cao 和 Xin Huang。2021年。利用高分辨率多视角影像估算城市地区建筑高度的深度学习方法：42座中国城市案例研究。《环境遥感》264（2021年），112590。https://doi.org/10.1016/j.rse.2021.112590[7] Li Deren、Zhang Guo、Shen Xin、Zhong Xing、Jiang Yonghua、Wang Taoyang、Tu Jianguang 和 Li Zhijiang。2019年。LJ-1 01卫星夜间灯光遥感的设计与处理。《遥感学报》23, 6（2019年），1011–1022。[8] Yufeng Duan、Xiaoxiao Jiang 和 Yaoqiu Kuang。2018年。通过将 DMSP/OLS 夜间灯光与梯度提升机相结合，理解经济增长与城市扩张之间的联系机制。《遥感》10, 3（2018年），384。[9] Peter Egger、Susie Xi Rao 和 Sebastiano Papini。2023年。中国建筑楼面面积：数据集与学习流程。*arXiv* 预印本 *arXiv:2303.02230*（2023年）。[10] Christopher D Elvidge、Kimberly E Baugh、Sharolyn J Anderson、Paul C Sutton 和 Tilottama Ghosh。2012年。夜间灯光发展指数（NLDI）：基于卫星数据的人类发展空间显式测度。《社会地理学》7, 1（2012年），23–35。[11] Christopher D Elvidge、John Safran、Ilya Nelson、Benjamin T Tuttle、Kimberly E Baugh、Janice B Dietz 和 Richard Tresch。2001年。美国及国际人口密度的1公里分辨率数据集：Landscan 2000。*ORNL*橡树岭国家实验室（2001年）。[12] Thomas Esch、Wieke Heldens、Andreas Hirner、Manfred Keil、Mattia Marconcini、Achim Roth、Julian Zeidler、Stefan Dech 和 Emanuele Strano。2017年。开拓人类居住地太空制图新局面——全球城市足迹。《*ISPRS*摄影测量与遥感学报》134（2017年），30–42。https://doi.org/10.1016/j.isprsjprs.2017.10.012[13] David Frantz、Franz Schug、Akpona Okujeni、Claudio Navacchi、Wolfgang Wagner、Sebastian van der Linden 和 Patrick Hostert。2021年。利用 Sentinel-1 和 Sentinel-2 时间序列进行全国尺度的建筑高度制图。《环境遥感》252（2021年），112128。https://doi.org/10. 1016/j.rse.2020.112128[14] John Gibson、Susan Olivia、Geua Boe-Gibson 和 Chao Li。2021年。经济学研究应使用哪种夜间灯光数据？应在哪里使用？《发展经济学杂志》149（2021年），102602。[15] Peng Gong、Han Liu、Meinan Zhang、Congcong Li、Jie Wang、Huabing Huang、Nicholas Clinton、Luyan Ji、Wenyu Li、Yuqi Bai、Bin Chen、Bing Xu、Zhiliang Zhu、Cui Yuan、Hoi Ping Suen、Jing Guo、Nan Xu、Weijia Li、Yuanyuan Zhao、Jun Yang、Chaoqing Yu、Xi Wang、Haohuan Fu、Le Yu、Iryna Dronova、Fengming Hui、Xiao Cheng、Xueli Shi、Fengjin Xiao、Qiufeng Liu 和 Lianchun Song。2019年。有限样本下的稳定分类：将2015年采集的30米分辨率样本集迁移至2017年10米分辨率全球土地覆盖制图。《科学通报》64, 6（2019年），370–373。https://doi.org/10.1016/j.scib.2019.03.002[16] Marco Gonzalez-Navarro 和 Matthew A Turner。2018年。地铁与城市增长：来自地球的证据。《城市经济学杂志》108（2018年），85–106。[17] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。2016年。用于图像识别的深度残差学习。载于*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。770–778。[18] Stephan Heblich、Stephen J Redding 和 Daniel M Sturm。2020年。现代大都市的形成：来自伦敦的证据*。《经济学季刊》135, 4（2020年05月），2059–2133。https://doi.org/10.1093/qje/qjaa014 arXiv:https://academic.oup.com/qje/article-pdf/135/4/2059/33668557/qjaa014.pdf[19] J Vernon Henderson、Adam Storeygard 和 David N Weil。2012年。从外太空衡量经济增长。《美国经济评论》102, 2（2012年），994–1028。[20] Vernon Henderson。2002年。发展中国家的城市化。《世界银行研究观察》17, 1（2002年），89–112。http://www.jstor.org/ stable/3986401

[21] Jie Hu、Li Shen 和 Gang Sun。2018年。挤压激励网络。载于*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。7132–7141。[22] Chao Ji 和 Hong Tang。2020年。利用改进的 Mask R-CNN 从单目卫星图像估算建筑层数。《遥感》12, 22（2020年）。https://doi.org/10.3390/rs12223833[23] Alex Kendall、Yarin Gal 和 Roberto Cipolla。2018年。利用不确定性为场景几何和语义的损失加权的多任务学习。载于*IEEE*计算机视觉与模式识别会议论文集。7482–7491。[24] Feng Lan、Huili Da、Haizhen Wen 和 Ying Wang。2019年。中国大陆城市群空间结构演变及其驱动因素：从单中心到多中心维度。《可持续性》11, 3（2019年），610。[25] Chuangnong Li、Lin Fu、Qing Zhu、Jun Zhu、Zheng Fang、Yakun Xie、Yukun Guo 和 Yuhang Gong。2021年。基于 Google 和 WorldView-2 遥感影像的农田建筑物提取注意力增强 U-Net。《遥感》13（2021年11月），4411。https: //doi.org/10.3390/rs13214411 [26] Mengmeng Li、Elco Koks、Hannes Taubenböck 和 Jasper van Vliet。2020年。大陆尺度的三维建筑结构制图与分析。《环境遥感》245（2020年），111859。https://doi.org/10.1016/j.rse.2020.111859[27] Ruidong Li、Ting Sun、Fuqiang Tian 和 Guang-Heng Ni。2023年。SHAFTS（v2022. 3）：一个基于深度学习的 Python 软件包，可从哨兵卫星影像中同步提取建筑高度和建筑占地面积。《地球科学模型开发》16, 2（2023年），751–778。[28] Yilong Liu、Hongbin Cui 和 Qiong Liu。2021年。合成孔径雷达图像中建筑物轮廓分割的数据增强：一项实证研究。《遥感》13, 9（2021年），1701。[29] Maxar。2021年。Maxar 的 WorldView-2 影像。https://resources.maxar.com/data-sheets/worldview-2。访问日期：2023-05-01。[30] Daniel P. McMillen 和 John F. McDonald。1998年。芝加哥郊区的人口密度：竞租方法。《城市研究》35, 7（1998年），1119–1130。https://doi.org/10.1080/0042098984510 arXiv:https://doi.org/10.1080/0042098984510[31] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox。2015年。U-Net：用于生物医学图像分割的卷积网络。载于医学图像计算与计算机辅助干预国际会议。Springer，234–241。[32] Annemarie Schneider、Mark A. Friedl 和 David Potere。2010年。利用 MODIS 500米数据绘制全球城市地区：基于“城市生态区”的新方法与数据集。《环境遥感》114, 8（2010年），1733–1746。https://doi.org/10.1016/j.rse.2010.03.003[33] 深圳市政府。2023年。深圳市政府数据开放平台。https://opendata.sz.gov.cn/。访问日期：2023-05-01。[34] Mingxing Tan 和 Quoc V Le。2019年。EfficientNet：重新思考卷积神经网络的模型缩放。载于第*36*届国际机器学习会议（*ICML*）论文集。6105–6114。[35] John Nicholas Tsivanidis。2018年。城市公共交通基础设施的总体效应与分配效应：来自波哥大 TransMilenio 的证据。[36] Dawei Wen、Xin Huang、Anlu Zhang 和 Xinli Ke。2019年。利用多视角卫星影像监测中国特大城市中心城区的三维建筑变化和城市再开发模式。《遥感》11, 7（2019年），763。[37] Fan Wu、Chao Wang、Hong Zhang、Juanjuan Li、Lu Li、Weirong Chen 和 Bo Zhang。2021年。基于深度学习框架的 GF-3 合成孔径雷达影像中国建成区制图。《环境遥感》262（2021年），112515。https://doi.org/10.1016/j.rse.2021.112515[38] Chen Yang 和 Shuqing Zhao。2022年。通过空间信息方法估算的2017年中国建筑高度数据集。9, 1（2022年），76。https://doi.org/10.1038/s41597-022-01192-x[39] Wenjuan Yu、Chuanbao Jing、Weiqi Zhou、Weimin Wang 和 Zhong Zheng。2021年。用于城市历史三维重建的 Landsat 时间序列数据。《遥感》13, 21（2021年）。https://doi.org/10.3390/rs13214339[40] Tianyu Zhu、Hao Wang、Mengyuan Wang、Yufei Huang 和 Bing Xue。2019年。中国铁路运输系统的经济绩效及空间溢出效应评估：一种夜间灯光方法。交通研究 *A* 部分：政策与实践122（2019年），1–18。[41] L Zhuo、T Ichinose、J Zheng、J Chen、PJ Shi 和 X Li。2009年。基于未经辐射校准的 DMSP/OLS 夜间灯光图像进行中国像素级人口密度建模。《国际遥感杂志》30, 4（2009年），1003–1018。

A 模型架构

在第3节所述的 U-Net 骨干网络实验中，ResNet-34 的表现最佳，其模型架构示意如下。

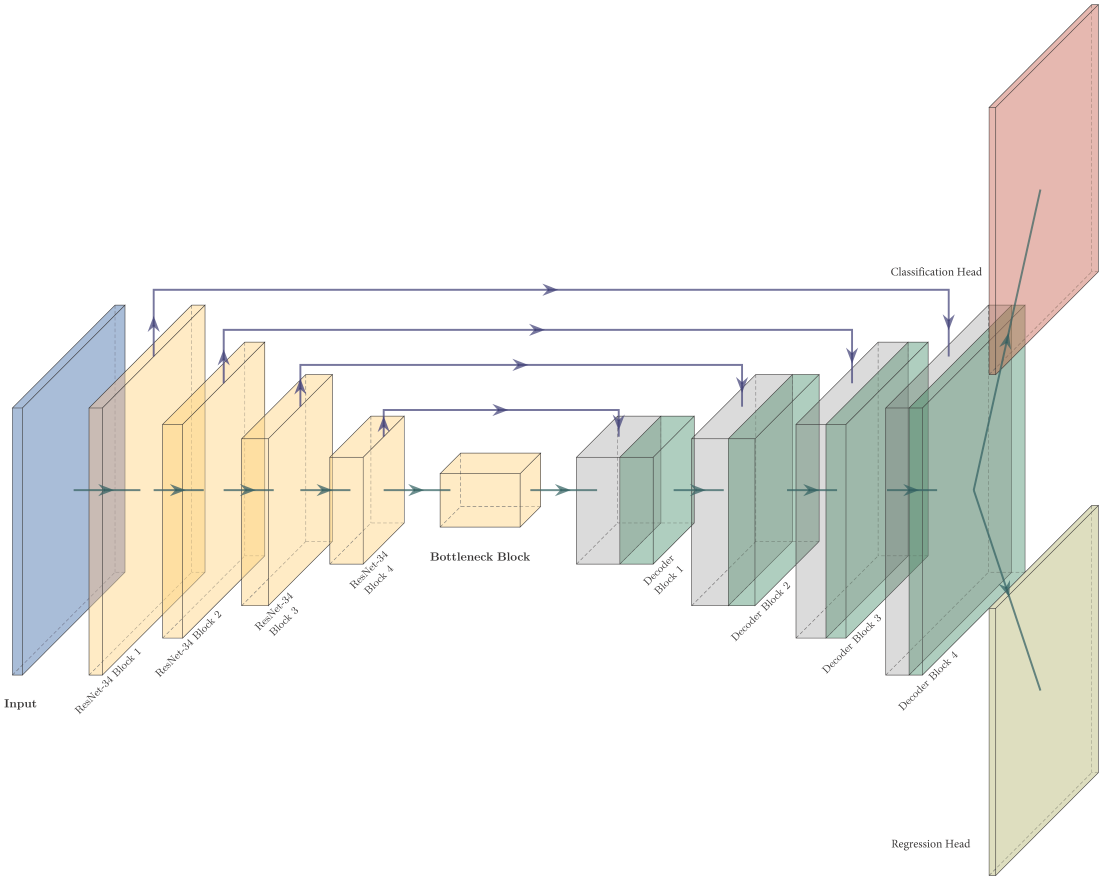


图9. 采用 ResNet-34 骨干网络的模型架构。

B 建筑物占地预测的更多结果

在我们之前的研究中 [?]egger2023building，我们使用不同配置的 U-Net 进行了实验，以预测建筑物轮廓，结果如表2所示：(a) Sentinel-1（合成孔径雷达数据）；(b) Sentinel-2（光学图像）；以及 (c) Sentinel 1+2（Sentinel-1 和 Sentinel-2 的叠加）。我们发现，在 (c) 中将合成孔径雷达数据与光学图像相结合，在所有指标上都取得了最佳性能。softmax 函数中有四个类别，即无建筑物、建筑高度低于 6 米的建筑物、高度为 6 至 27 米的建筑物，以及高度超过 30 米的建筑物。[7] 这些分界点的选择旨在将建筑高度的多模态分布划分为

[7]不过，我们测试了不同的截断值，例如低于10米、10米到20米以及高于20米，结果非常接近。

众包数据集中的全部 77 个城市分为局部最大值的三类（52%、39%、8%）。请注意，在我们的多任务学习器中，我们没有使用截断值，而是构建了一个基于回归的模型来预测高度。

	(a) Sentinel-1			(b) Sentinel-2			(c) 哨兵卫星 1+2		
	准确率	精确率	召回率	准确率	精确率	召回率	准确率	精确率	召回率
建筑占地面积	0.722	0.320	0.598	0.741	0.336	0.560	**0.762**	**0.375**	**0.617**
高度	0.670	0.194	0.364	0.703	0.229	0.382	**0.724**	**0.261**	**0.449**

表2. 使用 Sentinel-1 和 Sentinel-2 的 U-Net 性能指标。

C 两阶段模型的稳健性检验

为验证 [9], 中所述的单阶段模型的稳健性，我们使用哨兵卫星 1 + 2 影像测试了两阶段模型。在两阶段模型中，我们首先仅预测建筑物的建筑轮廓（第一阶段），然后将预测得到的建筑轮廓输入高度预测模型（第二阶段）。在两阶段模型的第二阶段进行高度预测时，我们验证了三种架构：（A1）将建筑轮廓预测结果与哨兵卫星影像堆叠，作为高度预测模型的输入；（A2）在将哨兵卫星影像输入高度预测模型之前，使用建筑轮廓预测结果对其进行裁剪；（A3）在编码器部分并行训练建筑轮廓预测和高度预测，并按照 Cao 和 Huang [6]的方法，使用跳跃连接将二者拼接起来，用于 U-Net 模型的瓶颈层和解码器部分。这三种架构的第一阶段（建筑轮廓预测）完全相同。每个阶段的损失函数和激活函数与单阶段模型中使用的类似；对于建筑轮廓预测和高度预测，分别使用 *sigmoid*函数 和 *softmax*函数 作为激活函数；损失函数仍为加权Dice系数。

两阶段模型的结果显示，建筑物轮廓预测略有提升，如表 3 所示；但在我们的所有尝试中，高度预测性能几乎保持不变。第一阶段（建筑物轮廓）预测平均每个训练轮次耗时 10 秒，而 A1、A2 和 A3 在第二阶段（高度预测）分别耗时 17.5、20 和 18.75 秒。与我们在 [9] 中介绍的分割器（每个训练轮次耗时 11 秒）相比，该分割器通过单阶段模型学习建筑物轮廓和高度，两阶段模型的计算成本显著更高。

	(c) 哨兵卫星 1+2						
	建筑占地面积			高度			
	准确率	精确率	召回率	准确率	精确率	召回率	微平均 F1_
单阶段	0.762	0.375	0.617	0.724	0.261	**0.449**	0.330
两阶段（A1）				**0.726**	0.261	0.433	0.326
两阶段（A2）	**0.773**	**0.442**	**0.673**	0.725	0.253	0.428	0.318
两阶段（A3）				0.725	**0.266**	0.446	**0.333**

表 3. U-Net 在哨兵卫星 1+2 两阶段模型中的性能指标稳健性检验。

D 加权对数损失及其结果

我们比较了两种方案：ResNet-34（尺度400损失）和 SE-ResNet-50（对数尺度损失），以预测深圳的建筑轮廓和高度。我们考察了它们对深圳不同高度类别和建筑轮廓覆盖率的预测（图 10）、将预测高度和目标高度聚合到 200米网格后的相关性（图 11），以及

模型预测的建筑高度对参考数据中建筑高度的解释方差份额随聚合层级的变化（图12）。

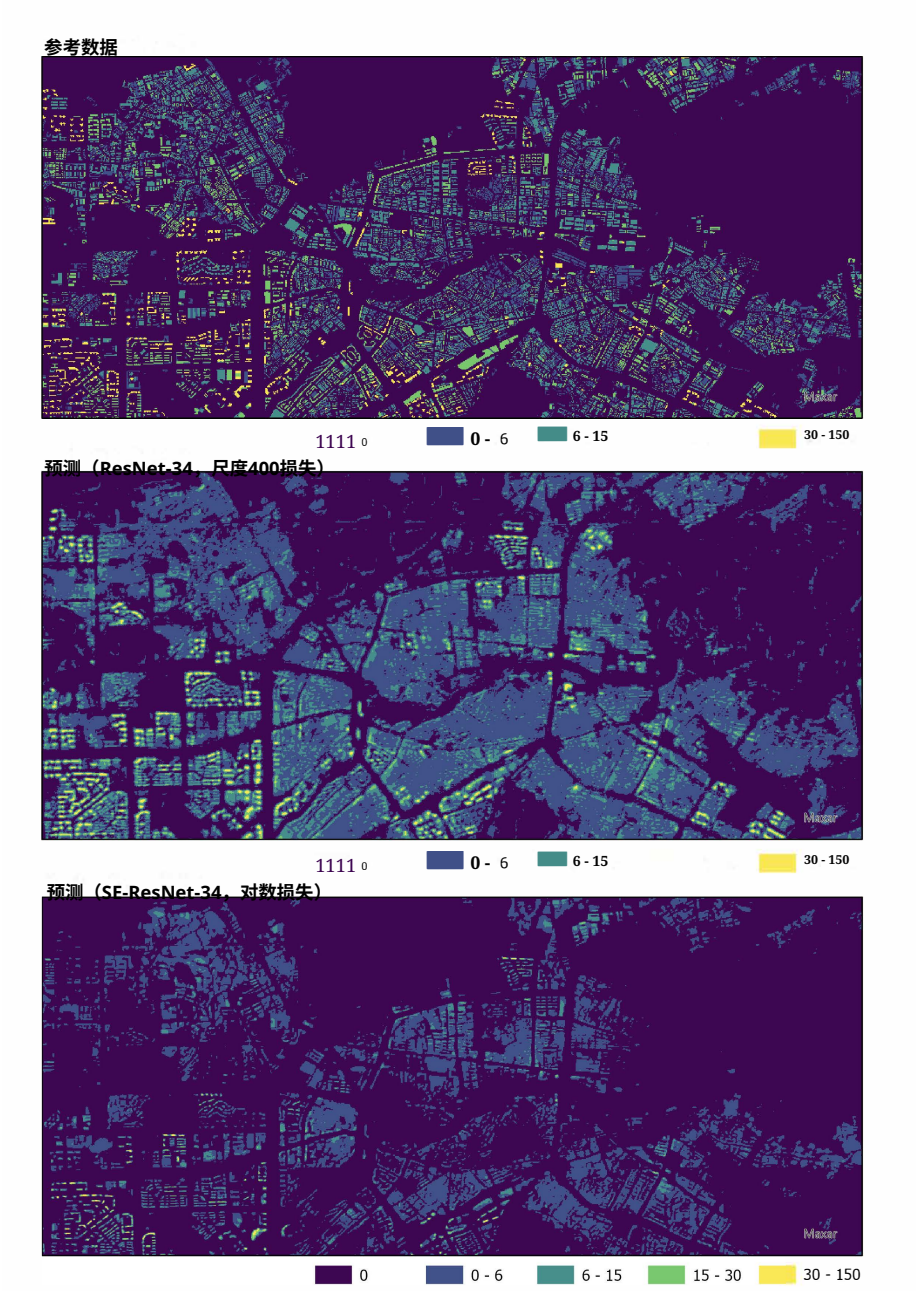
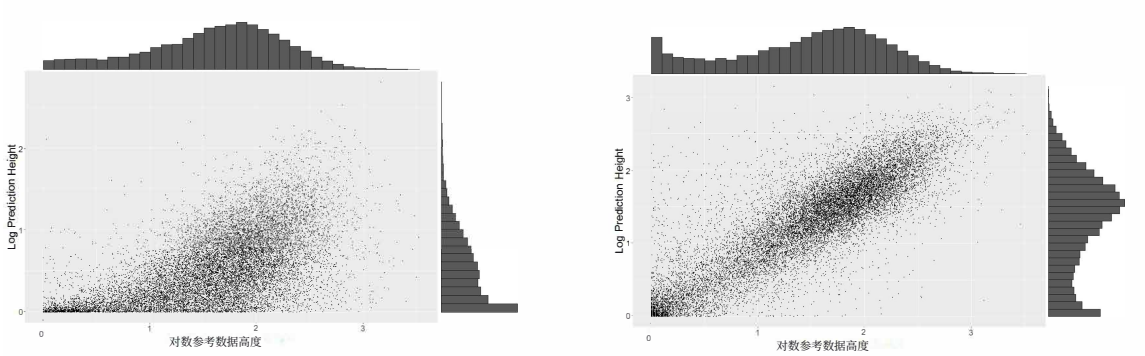
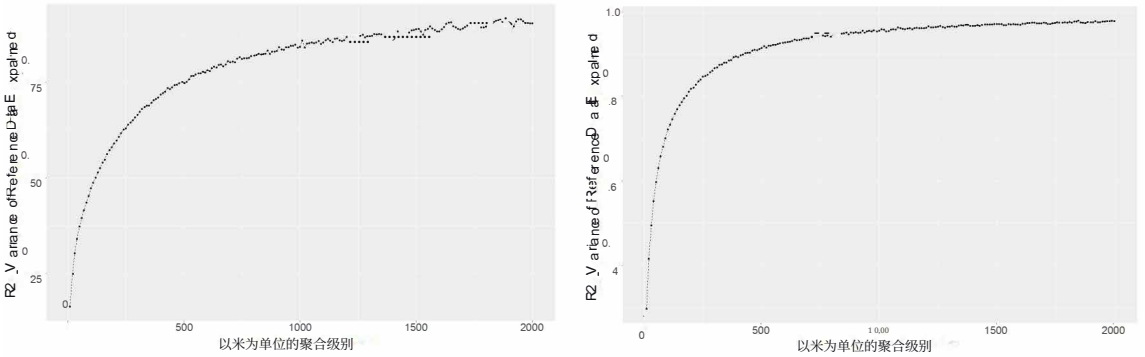


图10。两个模型在深圳的预测可视化。



对数损失，SE-ResNet-50 主干网络　　尺度400损失，ResNet-34 主干网络

图11. 比较聚合结果与200米网格的散点图：对数平均高度加一。



对数损失，SE-ResNet-50 骨干网络　　尺度400损失，ResNet-34 骨干网络

图12. 按聚合层级比较参考数据中建筑高度解释方差份额与模型预测建筑高度的关系。

使用高德地图改进参考数据

图13展示了我们基于高德地图应用程序接口构建的可定制工具的截图。通过查询多边形 JSON，我们获得了高德地图提供的实际建筑占地面积真实值。在图14中，我们对比了高德地图实际提供的真实值（黑色）与众包数据集中的参考数据（红色，据称约来自2017年）。这两组参考数据叠加在某个采样区域最新高分辨率卫星图像的卫星图像上。此处使用的两幅卫星图像由高德地图于2021年至2022年期间提供。据高德地图介绍，数据源为 DigitalGlobe&spaceview，城市地区的更新频率为0.5至1年，偏远地区则更长。正如我们在不匹配区域所看到的，实际真实值数据能够更好地反映建筑覆盖率。在未来的工作中，我们

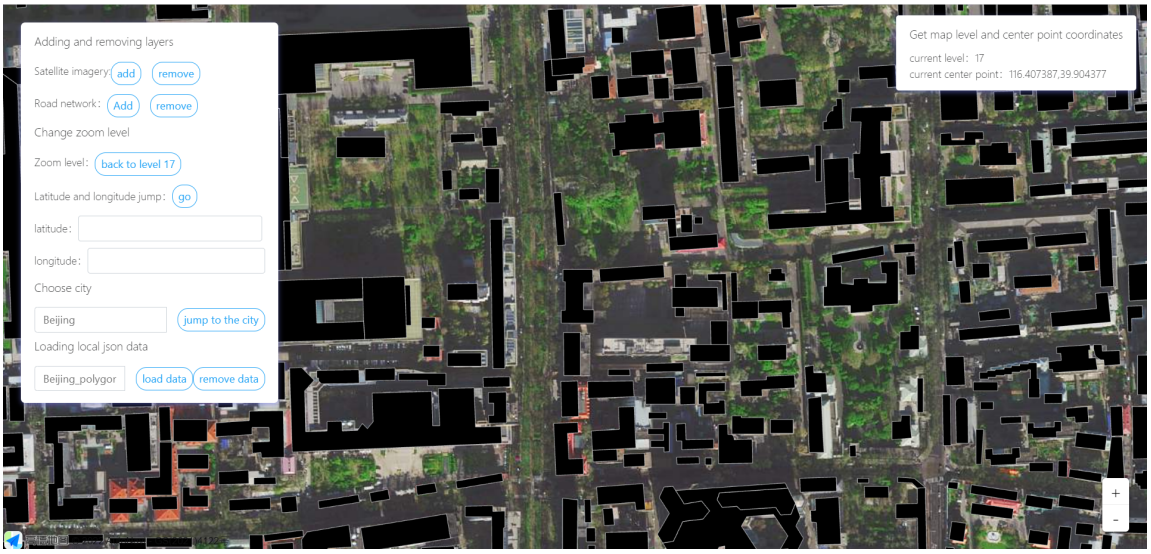


图13。 基于高德地图应用程序接口构建的定制工具箱，用于获取高德地图的建筑物轮廓和卫星影像数据集。

将系统评估各种参考数据的质量。此外，我们还看到一些没有可用参考数据的感兴趣区域；因此，正如我们在第6节中所讨论的，只需投入最少的人工进行标注，预计就能带来益处。

收稿日期：2023年5月15日



Fi图 14. 众包数据集中的建筑物轮廓（红色）与最新高德地图数据集中的建筑物轮廓（黑色）对比 k).