

基于机器视觉与卫星图像的聚落传统价值初步判断方法研究——以鄂西南为例

苗鹏程¹, 刘小虎², 董潇晓^{3✉}

摘要: 传统聚落的发掘, 通常需大规模实地调查, 这往往花费大量时间和金钱, 尤其是处于山地环境中的鄂西南聚落。本文利用卫星遥感影像, 开发了一种初步判断聚落传统价值的评估方式。研究采用基于 YOLO 的机器视觉检测模型, 通过识别固定图像等效地理尺度的卫星图中传统建筑, 并结合数学筛查, 进而判断整个区域的聚落传统价值。本研究对咸丰县 120848 张图片进行检测, 筛查, 最终获得 883 张聚落传统价值较高的图片, 其中包含 8 个传统村落和多组未挂牌聚落。结果表明, 该模型可以准确地找到传统村落, 并有效地提升搜寻高传统价值聚落的效率, 为以后的研究打下坚实基础。

关键词: 鄂西南; 传统村落; 机器视觉; 深度学习

Keywords: Southwest Hubei; Traditional villages; Machine vision; Deep learning

资助项目情况: 国家自然科学基金资助项目, 项目名称: 乡村住宅“一键式”人工智能设计系统研究——以湘鄂赣地区为例, 项目编号: 51978295

引言

乡村在当下的发展中, 肌理和文化正在遭受着破坏。在乡村的保护与发展中, 不仅仅需要对村落产业进行升级迭代, 更重要的是要对乡土文化的传承[1], 作为乡土文化载体的传统聚落, 成为了研究乡土文化的重要参考。然而, 由于多年的粗放式发展以及对于乡土聚落的不够重视, 许多的传统聚落已经成为了工业化的村庄, 不再保有传统的聚落文化[14]。为解决这个问题, 我国着手打造中国传统村落名录, 该名录收录了具有重要保护价值的村落, 这些村落因其保存完好的传统建筑、村落风貌以及传承的历史文化价值而被列入保护名录, 以促进传统村落的保护和开发利用。

1. 研究背景

1.1. 研究问题

在过往的传统村落评价中, 发掘保存较完整传统聚落往往是依靠政府部门的推荐, 专家下乡调研的方式。这样的方式需要消耗大量的时间和精力, 且容易因为信息差异以及其他原因导致本应该调研的传统聚落无法进行有效的研究, 进而影响村落文化的保护。因此, 在正式的传统村落调研前, 需要一种快速便捷的方式对聚落传统价值进行评估, 进而作为后续调研的依据。

1.2. 文献综述

¹ 苗鹏程, 华中科技大学建筑与城市规划学院, 博士研究生

² 刘小虎, 华中科技大学建筑与城市规划学院, 教授

³ 董潇晓(✉), 华中科技大学建筑与城市规划学院, 讲师, dongxiaoxiao@hust.edu.cn

1.2.1. 实际需求

鄂西南位于中国湖北省西南部[6]，是土家族、傣族、苗族等中国少数民族的聚集处。在生活文化以及建筑、村落形式上与汉族地区有着明显的区别。建筑以适应鄂西南山地的吊脚楼为主，村落布局顺应山势，形成了一套独特的体系[8][9][13]。然而，中国的改革开放后经济快速增长以及城市化速度过快，导致传统村落大面积消亡[7]。更为严重的是，大量的特色文化也因为环境的改变而消亡[10]。因此，及时发掘和保护传统村落尤为重要。由于鄂西南以山地为主，村与村之间交通不便，难以通过逐一排查的方式调研。而政府人员推荐和文献资料调查为主的传统调研方式，难以了解村落当前状况。因此需要一种可以快速大量筛查传统村落遗存的方式。

1.2.2. 技术基础

近年来，随着高分辨率遥感技术的快速发展，卫星已经能够提供厘米级空间分辨率图像。这些图像可从 Google 地球和天地图等多个在线平台免费或低价获取。随着近年来卫星遥感与机器视觉的紧密结合，越来越多的研究以此为方法。图像的分割可以使用深度学习提取建筑物足迹和派生属性层[2][3]。Shenglong Chen 等人证实了机器视觉模型可以从卫星图中有效提取建筑物[16]。Ziyao Xing 等人通过将卫星遥感图像和街景图像相结合的方式，通过对卫星图识别，判定建筑所在位置，并与立面结合，判断建筑的洪水脆弱性[4]。在聚落研究中，Zhiling Guo 等人通过机器学习的方法识别村落建筑物[15]，Fabrice Monnaa 通过机器视觉与卫星遥感结合，对印度尼西亚松巴岛的建筑物进行检测，寻找建筑遗产[5]，这些研究证实了机器视觉可以用于在卫星图中准确识别建筑与其他信息，有快速判断聚落传统价值的潜力。

1.3. 研究目的

本研究通过机器视觉结合高清卫星遥感，开发一种可快速识别鄂西南传统村落的方法，通过训练机器视觉模型，从卫星图像中自动提取具有特定建筑特征的村落数据，过滤出符合传统村落特征的区域，从而实现大范围、批量化的识别和初步定位，完成对传统村落初步筛选的目的。进而为后续的下乡调研提供证据，确保实地调查的针对性。

1.4. 研究创新点

1. 本研究创新将机器学习算法用于村落普查中
2. 本研究创新训练了一个可识别传统建筑与现代建筑的机器视觉算法
3. 本研究验证了机器视觉识别传统聚落的可靠性

2. 数据收集与模型训练

本次实验通过卫星图寻找聚落，并判断其传统价值。具体步骤如下图 1：

第一步是通过已认证的传统村落进行标注，每张图片的地理尺度为 500m*500m。

第二步是训练 YOLOv5 模型，通过不断调整参数，最终得到可以较为准确识别传统村落的模型。

第三步是切分鄂西南样本地图，切分图片的每张图片的地理尺度为 500m*500m，切分每张重叠度为 50%，通过机器视觉检测，将识别结果通过数理计算进行区分，找到有可能是传统聚落的图片，并与中国传统村落名录中的村落进行对比，确认准确率。

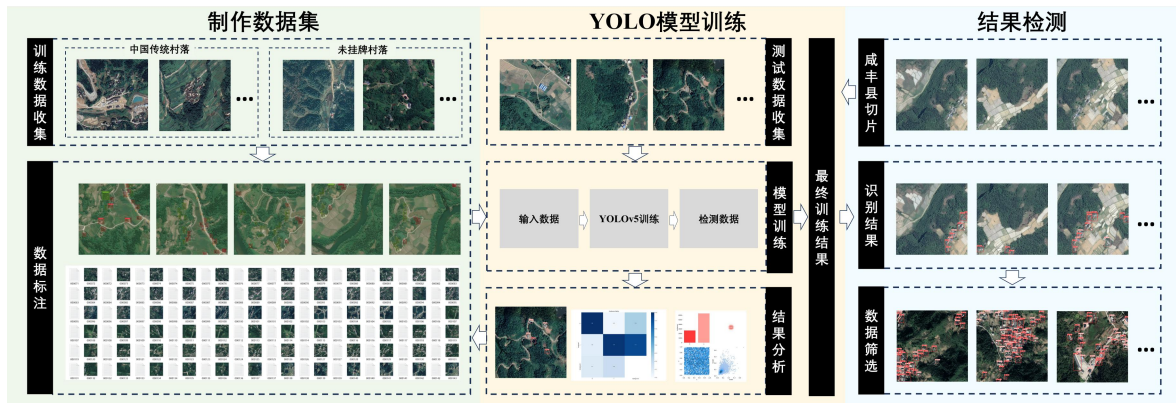


图 1. 研究步骤


2.1. 数据收集






2.1.1. 图像尺度分析

由于机器视觉在同一轮训练中，保证每组图片的像素大小一致可提高训练效率。因此，本研究通过将比例尺融入到图像大小中的数据处理方式，控制每张图像所代表的实际地理尺度（下称图像等效地理尺度）。

根据鄂西南的传统聚落特征，可以将鄂西南传统聚落分为三类，分别是条带式，组团式与散点式 [12]。在鄂西南，由于人口的流失，当前的行政村大多是由多个聚落组团构成，若以行政边界为界限，则会导致整个图面过大，进而影响机器视觉的识别效果。因此，本研究在制定图像识别尺度时，更多采用单个聚落组团。本研究分别对三种典型的传统聚落进行分析，研究发现，在鄂西南传统聚落中，聚落组团尺度差异大。散点式聚落由于布局分散，一个聚落组团的图像等效地理尺度在 $100\text{m} \times 100\text{m} - 150\text{m} \times 150\text{m}$ ，而条带式由于顺应道路线性排布，聚落尺度往往较大，为 $400\text{m} \times 400\text{m} - 700\text{m} \times 700\text{m}$ 。下表（表 1）是针对不同类型的聚落组团图像等效地理尺度。小的尺度（例如湄坪村散点组团 1， $110\text{m} \times 110\text{m}$ ）无疑会让机器视觉识别建筑的效果更好，但由于图面过小而往往难以判断整个聚落组团的保存现状。而过大的图像等效地理尺度则难以有效识别建筑。经过多次实验，证实当图像尺度为 $500\text{m} \times 500\text{m}$ 时，既可以保证机器视觉的准确度，同时也可以尽可能多的保留传统聚落的平面肌理。

表 1. 不同村落尺度对比（图片来源：<https://www.google.com/maps>）


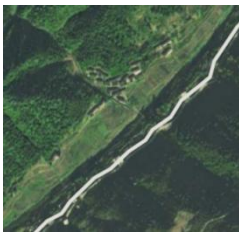

村落类型	村落名称	村落图像	图像尺度 (m)
条带式村落	老屋基村		400*400

组团式村落	大集杨村		650*650
	彭家寨		240*240
	曾家寨		240*240
散点式村落	湄坪村散点组团 2		130*130
	湄坪村散点组团 1		110*110

2.1.2. 图源选择

本次实验数据为鄂西南的卫星图。通过对比当下常用的几种图源，选择清晰度更高的谷歌卫星图作为本次实验的卫星地图（表 2）。

表 2. 不同地图源对比（图片来源：<https://www.google.com/maps>）

地图名称	百度地图	高德地图	谷歌地图
彭家寨			

徐家寨



2.2. 模型训练

2.2.1. YOLOv5[11]模型

YOLOv5 模型是一种目标检测模型,用于在视频或图像中识别和定位物体。它是 YOLO(You Only Look Once) 算法的最新版本。YOLOv5 使用卷积神经网络(CNN)来学习图像中物体的特征,并使用多尺度预测和网格分割来检测和定位目标。

该模型的优势在于其高速运行能力,并且可以在不同的图像分辨率上很好地工作。这使得 YOLOv5 成为一种高效的目标检测模型,可以应用于许多不同的场景,包括图像分析、卫星图识别等。

2.2.2. 数据集制作

本研究分为训练数据集和识别数据集两部分。训练数据集的数据来源于鄂西南传统村落以及其他不同村落的谷歌卫星图,用于训练可以识别鄂西南传统建筑的机器视觉模型。识别数据集是验证机器视觉排查效果的数据集。

鄂西南传统聚落破坏较为严重,难以从特有的形制入手大量地标记机器学习数据集,因此,直接通过聚落肌理判断传统村落是十分困难的。因此需要将聚落进行拆分,从聚落元素入手判断聚落的传统价值。由于在本研究中卫星图的特性,建筑较文化景观有着更易被识别的特征,且在中国传统村落判断标准中,聚落中传统建筑遗存是判断传统村落的重要指标之一。故而在本研究中,将通过对聚落内传统建筑遗存的分析,作为判断聚落保留了完好的传统布局,拥有更高的传统价值,值得进一步的调研分析的依据。

鄂西南地域广阔,机器视觉筛查整个鄂西南地区需处理大量的数据,进而影响实验的效率。因此,本研究挑选鄂西南典型区域——湖北省恩施州咸丰县作为检验方法有效性的样本。为保证结果有效性,在训练数据集中将删除有关咸丰县的数据。咸丰的传统民居聚落分布广泛,在咸丰十个乡镇和一个区中都有分布。传统民居主要以吊脚楼为主,历来就有“干栏之乡”的美称,具有典型的鄂西南文化特征[13],可以作为典型案例进行探讨。

2.2.3. 训练数据集

根据 YOLOv5 自身的特性,要得到好的机器识别效果,需要大量的数据作为训练集。因此难以通过特殊建筑,例如鼓楼、摆手堂等,作为传统聚落的识别依据。故本次选取识别遗存较多的传统建筑以及与之相对的现代建筑作为排查传统聚落的依据。两者在遥感图中的差异主要体现在屋顶形制的不同。现代建筑,尤其是混凝土建筑,颜色普遍为蓝色、红色和白色(图3),而传统建筑屋顶的颜色大多为灰色(图2),因此可以使用机器视觉加以区分。

由于天气状况等原因,卫星图本身存在色差,而这样的色差会很大程度上影响机器视觉的识别准

确度。因此，需要在当前标注的基础上针对不同色差进行标注。这样标注可以很大程度降低颜色差异带来的影响。



图 2. 以传统建筑为主的传统村落（彭家寨）（图片来源：<https://www.google.com/maps>） 图 3. 以现代建筑为主的村落（松坪村）（图片来源：<https://www.google.com/maps>）

本次标注共选择大约 3000 张在中国传统村落名录中的传统村落与非传统村落的图片进行标注。本次数据标注的软件为 VoTT, 对标注 Label 的命名分别为“traditional building”和“normal building”两个，分别代表传统建筑和现代建筑。标注是通过人为识别图像中传统建筑与现代建筑，并通过 Bounding Box 将其框选，最终获得与被标注图像匹配的文本数据。该数据记录了 Bounding Box 相对于整张图片尺寸的比例以及所处的相对位置。标注者是熟悉传统村落领域并在第五批、第六批传统村落普查中调研村落的专家，标注数据具有可靠性。（图 4）



图 4. 标注实例（图片来源：作者自绘）

2.2.4. 识别数据集

本次实验中，从谷歌地图中获取咸丰县全域卫星图，精度为 19，使用 GIS 将卫星图按照 500m*500m 的方格切分，为防止切分图片边缘存在建筑而导致漏检的情况，在本次切分中，每相邻两张图片的重复率为 50%。共切分出 120838 张图片，作为机器视觉最终的识别数据集。

3. 实验结果与数据分析

3.1. 机器视觉网络识别结果

从机器视觉算法本身来说，在本研究最终结果中，传统建筑识别精度为 0.81、召回率为 0.72。且从（加一段）

机器视觉识别的结果显示，面对密集的聚落，会有一定程度的遗漏。而在较为稀疏的环境中，识

别效果更好。除此之外，还存在错检现象。错检大多是由于现代住宅屋顶与传统住宅的形制和颜色相近而导致，且大多数错检都是将现代住宅识别为传统住宅。因此，对于初步筛选的需求来说，是可以满足的。

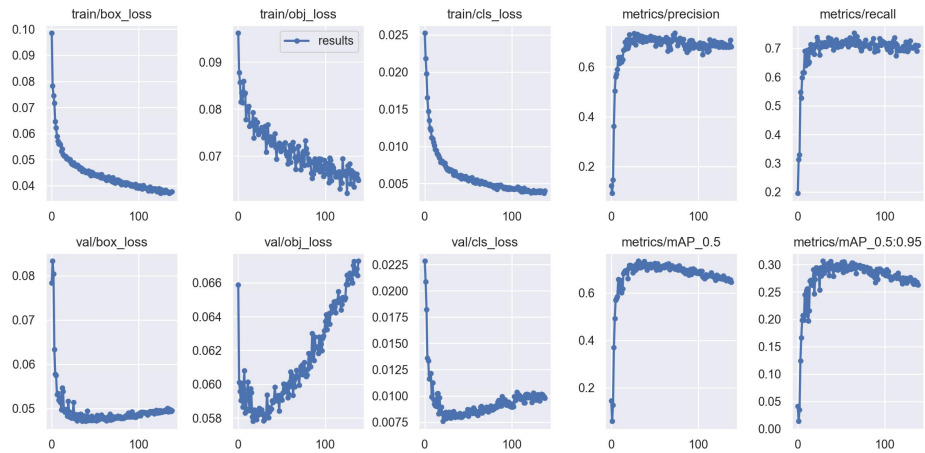


图 5.

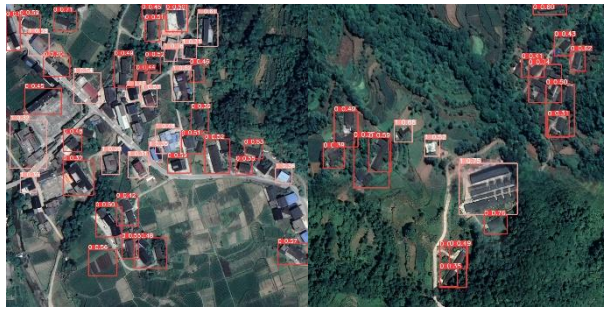


图 6. 机器视觉识别结果图（图片来源：作者自绘）

3.2. 数理计算筛查疑似传统聚落

根据中国传统村落判断标准，当一个村落中传统建筑的比例大于 60% 则可以获得最高的评分。因此，依据这个标准，对本次识别结果中，切片中识别传统建筑占总建筑比大于 60%，即为筛选的有较高传统价值的聚落。

研究对咸丰县切出的 120838 张图片进行机器视觉识别，并在此基础上通过数理计算进行筛查。当仅仅以图像中超过 60% 的建筑为传统建筑作为筛查条件时，共筛查出 32081 张图片。其中除了本身切分中存在的重叠以外，很大程度是一些图片内仅有 1-2 栋传统建筑，这样的规模难以作为本次研究所想筛选出的聚落。因此对筛选规则进行优化，规则改为三条，并最终总结成为数学公式(1)：








1. 每张图片内，筛查出建筑总数必须大于 5。
2. 传统建筑选框占图片面积大于整张图片的 0.05。
3. 图片内，传统建筑选框面积占比应大于总选框面积（传统建筑选框和现代建筑选框）的 60%。

$$\begin{cases} x_T > 5 \text{ ①} \\ \frac{S_T}{S_P} > 0.05 \text{ ②} \\ \frac{x_T}{x_T+x_N} > 0.6 \text{ ③} \end{cases} \quad (1)$$

其中， x_T 代表每张图片中识别的传统建筑数量， S_T 代表识别传统建筑选框面积， S_P 代表图片整体面积， x_N 代表每张图片中识别现代建筑数量。

最终，经过公式(1)三条规则的筛选，在咸丰县切分的 120838 张图片中，成功筛选出 883 张图片，与筛选优化前对比，图像内聚落质量有很大的提升。（表 3）。

表 3. 两次筛选结果对比（图片来源：作者自绘）

筛选规则	实验结果				
第一次筛选					
第二次筛选					

3.3. 数据检验

通过对检验出的 883 张图片进行验证，发现咸丰县的 9 个传统村落中，有 8 个被成功识别，为大路坝区蛇盘溪村、甲马池镇马家沟村、清坪镇中寨坝村郑家坝、尖山乡唐崖寺村、高乐山镇官坝村、高乐山镇牛栏界村、清坪镇大石坝村、清坪镇庄房坝村。其中坪坝营镇新场村蒋家花园由于新建筑过多，导致识别后在数理分析中无法有效地筛选。

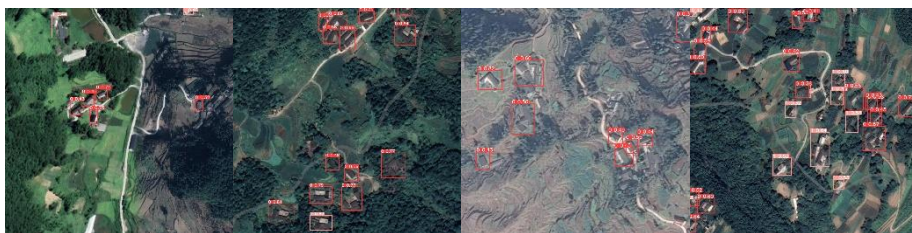









图 7. 检测出的中国传统村落示例（从左到右：大路坝区蛇盘溪村、清坪镇大石坝村、尖山乡唐崖寺村、清坪镇庄房坝村）（图片来源：作者自绘）



图 8. 坪坝营镇新场村蒋家花园 (图片来源: 作者自绘)

除了中国传统村落名录中的村落以外, 还找到了类似椿木槽村和葫芦坝村等初步筛查聚落 9 组。

表 4. 初步筛查聚落 9 组 (图片来源: 作者自绘)

村落名称	图像	村落名称	图像	村落名称	图像
椿木槽		龚家坨		坪坝营	
葫芦坝		梅家庄		老岩孔	
大坑		石柱坝村		八角庙	

4. 结果分析

从当前的实验来看，通过大量的、不同色调的样本学习，机器视觉算法可以有效地区分传统建筑与现代建筑（图 9）。但是由于近年来的各种乡村工作，使得更多的现代建筑在卫星图上与传统建筑相仿，因此会存在错检现象。由于图像上颜色的差异，也容易导致图像中一些模糊的建筑难以被准确地识别。除此之外，由于鄂西南传统村落相较于现代村落来说密度较小，因此本次训练集中缺少对高密度村落的标注数据，因此在最终的结果中，往往也会容易出现漏检现象（图 8）。

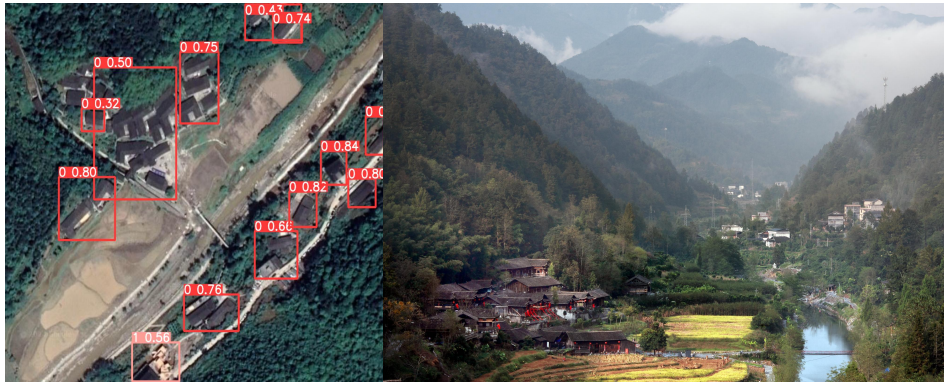


图 9. 彭家寨机器视觉识别图和村落实景图（图片来源：作者自绘、自摄）

在本次实验对咸丰县切分的 120838 张图片中，成功筛选出 883 张传统价值较高的聚落。这其中包括咸丰县 9 个在中国传统村落中的 8 个，准确率为 89%。由此可以说明，虽然当前技术存在缺陷，但通过机器视觉识别传统建筑去排查聚落传统价值的路径是可行的。而能够成功从检测数据中成功筛选，也可以说明筛查的标准的优化是有效的，相较于仅仅采用建筑遗存比例筛查的 32081 个结果，优化后的筛查标准减少了超过 97% 的低质量结果，大大优化了检测效率。

在最终的结果中，未能筛查出坪坝营镇新场村蒋家花园，是因为周边现代建筑过多，导致在后续的数理统计中，图片中传统建筑少于 60% 而被淘汰。蒋家花园入选传统村落名录更多的原因是在建筑尺度上的蒋家花园十分珍贵，而非村落尺度上整个村落的传统风貌遗存较好，属于特殊案例，不在本方法的筛查范围内。蒋家花园为一个保存完好的多进传统院落，在卫星图识别中被成功识别，这在侧面进一步说明通过卫星图筛查传统建筑的准确性以及有效性。

除了 9 个中国传统村落以外，算法还检测出多个有较高传统价值的聚落，由于数量过大且有一定比例的重合，文章节选一部分加以说明（表 4）。这说明除了当前的认定的传统村落以外，还有许多亟待发掘的聚落，这说明了在传统村落研究中，对整个区域进行排查必要性，同时为后续对这些聚落进一步研究提供了依据。

5. 结论

本研究使用机器视觉算法，可以有效地在卫星遥感图中寻找传统价值较高的乡村聚落，并作为未来下乡调研的前期依据，使得未来的调研更加有的放矢。但由于缺少立面信息，使得本方法同样存在局限性，难以在初步判断的基础上进一步对村落的传统风貌遗存进行判断。在后续的研究中，可以将

立面整合到识别系统中，或者寻找一种方式，通过屋顶平面判断立面形式，进而能对村落传统风貌进行更有效地判断。

参考文献

- [1] 冯骥才. 传统村落的困境与出路——兼谈传统村落是另一类文化遗产[J]. 民间文化论坛, 2013(01):7-12. DOI:10.16814/j.cnki.1008-7214.2013.01.002.
- [2] Jing, W., Lin, J., Lu, H., Chen, G., & Song, H. H. (2022). Learning holistic and discriminative features via an efficient external memory module for building extraction in remote sensing images. *Building and Environment*. <https://doi.org/10.1016/j.buildenv.2022.109332>
- [3] Abdollahi, A., Pradhan, B., & Al-amri, A. (2020). An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images. *Geocarto International*, 37, 3355–3370. <https://doi.org/10.1080/10106049.2020.1856199>
- [4] Xing Z, Yang S, Zan X, et al. Flood vulnerability assessment of urban buildings based on integrating high-resolution remote sensing and street view images[J]. *Sustainable Cities and Society*, 2023, 92: 104467.
- [5] Monna F, Rolland T, Denaire A, et al. Deep learning to detect built cultural heritage from satellite imagery.-Spatial distribution and size of vernacular houses in Sumba, Indonesia[J]. *Journal of Cultural Heritage*, 2021, 52: 171-183.
- [6] 李晓峰, 谭刚毅. 《两湖民居》(第一版). 北京: 中国建筑工业出版社, 2009.
- [7] 张睿婕, 高元, 李佳奇. 近 30 年来中国传统村落的研究热点、进程及展望——基于 CiteSpace 的可视化解析. *现代城市研究*, 2021, (11): 1-7+15.
- [8] 张良皋. 吊脚楼——土家人的老房子. *中国美术*, 1995, (4): 28-31.
- [9] 张良皋, 李玉祥. 武陵土家(第一版). 北京: 三联书店, 2001.
- [10] 刘馨秋, 王思明. 中国传统村落保护的困境与出路[J]. *中国农史*, 2015, 34(04): 99-110.
- [11] Zhu X, Lyu S, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 2778-2788.
- [12] 肖慧. 鄂西南土家族传统聚落形态研究[D]. 湖北工业大学, 2018.
- [13] 胡平. 鄂西传统民居聚落影响因素分析[D]. 华中农业大学, 2008.
- [14] "加快公共文化服务体系研究"课题组, 吴理财. 城镇化进程中传统村落的保护与发展研究——基于中西部五省的实证调查[J]. *社会主义研究*, 2013, (04): 116-123.
- [15] Guo Z, Shao X, Xu Y, et al. Identification of village building via Google Earth images and supervised machine learning methods[J]. *Remote Sensing*, 2016, 8(4): 271.
- [16] Chen S, Ogawa Y, Zhao C, et al. Large-scale individual building extraction from open-source satellite imagery via super-resolution-based instance segmentation approach[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, 195: 129-152.

图片来源

图 1-2, 图片来源: <https://www.google.com/maps>

图 3-6, 作者自绘

图 7, 作者自绘, 自摄