

深度学习下的民居建筑自动分类量化测度研究^①

张春明¹, 谭人殊¹, 宋晨明², 赵党书³

1. 云南艺术学院, 昆明 650101; 2. 西安交通大学, 西安 710049; 3. 昆明理工大学, 昆明 650031

摘要: 针对传统人工采集和判定的方法受到诸多因素的影响, 导致效率低下且评定结果的波动性的问题, 探索了一种基于深度学习方法进行民居建筑的多项特征指标的量化评定. 采用 CNN 神经网络结构对民居建筑进行特征的自动分类. 结果表明, EfficientNet 是性能最优秀的 CNN 架构. 采用这样的研究途径有助于更为全面地了解民居建筑的演化趋势和相关保护政策的制定.

关键词: 民居建筑; 量化评测; 卷积神经网络; 风貌特征

中图分类号: TP391.41

文献标志码: A

文章编号: 1000-5471(2023)06-0001-11

Research on Quantitative Measurement of Automatic Classification of Residential Buildings Under Deep Learning

ZHANG Chunming¹, TAN Renshu¹,

SONG Chengming², ZHAO Dangshu³

1. Yunnan Arts University, Kunming 650101, China; 2. Xi'an Jiaotong University, Xian 710049, China;

3. Kunming University of Science and Technology, Kunming 650031, China

Abstract: The traditional manual collection and judgment methods are affected by many factors, resulting in slow efficiency and fluctuation of evaluation results. This study explored a quantitative evaluation of multiple characteristic indexes of residential buildings based on deep learning method. CNN neural network structure was used to automatically classify the characteristics of residential buildings. The results showed that EfficientNet was the best CNN architecture. Adopting such a research approach will help to have a more comprehensive understanding of the evolution trend of residential buildings and the formulation of relevant protection policies.

Key words: residential buildings; quantitative evaluation; convolution neural network; style characteristics

1 研究背景

从 2012 年开始, 住房和城乡建设部、文化和旅游部、财政部组织开展了全国传统村落摸底调查, 截至 2019 年第 5 次全国传统村落调查统计, 全国有 6819 个传统村落进入统计名单.

① 收稿日期: 2022-09-11

基金项目: 云南艺术学院科研基金项目(2021KYJJ05); 教育部人文社会科学研究青年基金项目(21YJC760070).

作者简介: 张春明, 副教授.

通信作者: 赵党书, 副教授.

村落的建筑环境随着经济建设的发展,传统的民居建筑正在日益遭受到逐步的蚕食,村落的风貌正在逐步的异化,所幸的是这样的问题已逐渐得到重视,但是对于传统村落的保护发展,如何评估村落民居的建筑特征风貌,对其进行分类统计管理,确实是一件非常必要且困难的工作.

对于传统村落的保护,迫切地需要对村落民居建筑的数量、风貌、建筑质量、建筑高度等信息进行量化的评估和分析,进而完善对村落规划发展管理,目前在全国开展的实用性村庄规划编制工作中,对此也提出了具体的相关要求,然而目前对于村落建筑信息的统计主要还是通过人工现场调研的方式进行,对民居进行多方位的拍照之后,对其建筑特征进行人工的判读识别.这样的方式,一方面容易受到所在地村落交通、气候和地形等因素的限制,给数据的采集带来不便,同时还会增加大量的调研成本;另一方面由于通过人工判读的方式,势必会因为判读人学科背景、生活经历、心情情绪等主观的因素影响,带去一些不确定的变化,给建筑风貌的界定的结果带去一定的扰动.

新技术新数据的不断涌现为更为细致的空间品质研究提供了丰富的数据基础,同时将机器学习、边缘计算等智能技术运用于各个行业,是一个顺应时代发展的解决方案^[1].这样的研究首先需要进行一个数据收集过程来收集所需的数据,该过程收集关键的建筑影像数据,通常依赖于现场调查.如此高度的劳动密集型和耗时的工作使得进行大规模的建筑风貌评测极其困难.在这方面,以有效的方式收集和集成建筑风貌数据仍然是目前学术界研究的挑战.

目前的研究,基于深度神经网络的信息抽取模型在公开数据集取得了不错的成绩^[2],但是很难同时保证大范围、细粒度的计算结果^[3].本研究提出了一个通用的框架来解决上述挑战,它利用现场获取的民居图像进行民居建筑的特征提取和分类.具体来说,调研现场拍摄的图像中显示的立面结构足够丰富,可以进行民居建筑特征分类(图 1),图 1 中的建筑立面显示的细节比相应之前研究通过卫星遥感影像图显示的屋顶图,揭示了不同类型的建筑更多的细节.因此,在该方法中,根据建筑现场图片对建筑进行分类训练,然后通过高分影像图空间聚类将推断出的特征标签与单个建筑连接起来,在此基础之上,利用卷积神经网络和反向传播算法,通过卷积操作对民居建筑图像的特征进行提取,对于民居的风格,卷积神经网络模型的浅层学习主要是诸如民居外观纹理、颜色等简单特征,而深层学习的是一些高纬度语义特征,将这些高纬度语义特征和民居建筑风格的标签对应,通过反向传播算法这种监督学习的方式不断迭代优化模型参数,最终建立输入(原始图像)和输出(建筑风格)映射关系.基于此原理,此次研究中,试图建立一个构建民居建筑图像的基准数据集,以训练卷积神经网络(convolutional neural network, CNN)在建制村范围内进行分类训练,进而对民居的多维度特征进行机器识别与判读,因为 CNN 已经被证明了其在这类任务中强大的能力^[4-5].

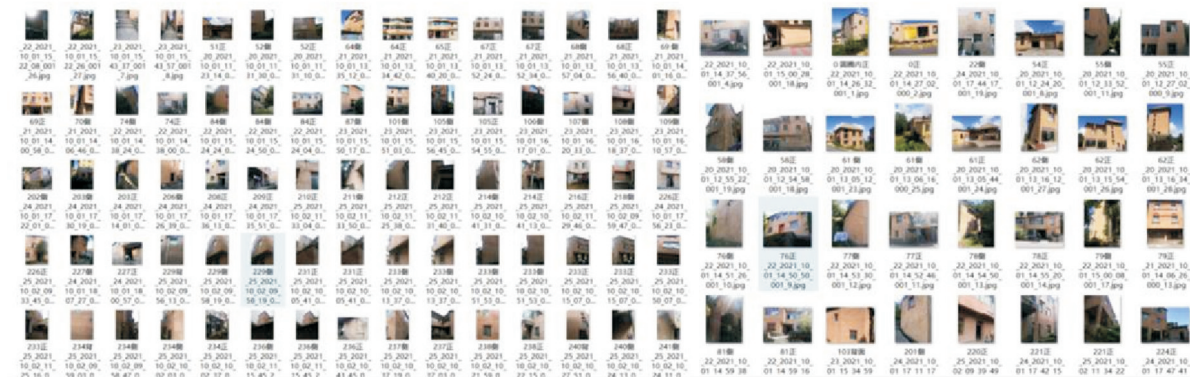


图 1 民居建筑数据集

综上所述,本研究的贡献如下:

- 1) 提出了基于单体建筑物层面上的特征分类的一般框架;
- 2) 基于立面结构构建了一个用于训练建筑实例 CNN 分类器的建筑图片基准数据集.本研究所使用的数据集主要通过研究地村落现场图片拍摄获取;
- 3) 所获得的建筑分类图显示了数据对于村落量化分析的潜力.对后续理解村落布局结构和成因量化将起到很大的帮助.

2 预备工作

随着大数据技术的推广深化以及人工智能工业应用的蓬勃发展,越来越多的研究者尝试结合人工智能手段和传统学科之间的大数据展开系统、深入的挖掘^[6-7]. 本次研究主要通过深度学习的方法对传统民居建筑风貌进行特征提取判读. 建筑单体图片是反映建筑风貌的最直观的数据,通过图片元素对建筑的风格、材质和结构形式等多个物理特征进行分辨,在过去的几十年里,对图片的机器理解仍然是模式识别和图像分类领域的一个活跃的研究课题. 具体来说,给定一个建筑单体图像,不仅检测到单个建筑立面元素以及它们的物理属性,还检测到与此更为高级相关的有意义的语义单元,可以识别出相关信息(如风格、建筑质量). 早期的工作集中于基于低级图像处理的平面图分析,如通过霍夫变换^[4]的线检测,通过字袋模型^[8]的图形符号识别. 对于建筑几何特征和外观形式及组合模式的主要依赖于图片所反映的建筑物理属性进行判断识别. 近年来,一些基于 CNN 的数据驱动技术^[9-11]已经取得了很好的效果. 然而,他们使用的公共数据集^[9,12-14]是从城市住宅中收集的,居住建筑外观的复杂性是有限的,而图形元素则是常规形式. 相比之下,农村住宅受地形、气候和居住习惯等多方面因素的影响,其建筑结构更加复杂,风格更加多元. 因此,不同农村地区的建筑有着不同的几何结构形式和建筑外观风格. 这些因素使得人工智能对于建筑的分类效果不能获得令人满意的预测结果.

本研究的主要步骤流程是对村落民居的图片采集、图片预处理、民居风貌特征分类与映射. 此次图片数据的采集主要是通过对元阳多依树村进行人工拍照的方式获取各个自然村民居建筑图片,在此基础上,对图片进行初期的筛选及图片预处理,同时结合规划文件的需要,对民居建筑的特征提出了 4 个方面的特征指标描述. 根据 4 个特征特点在多个不同卷积神经网络模型上进行随机分割图片数据的预测指标得分对比,进而选出最优模型结构. 在此基础上,采用测试集数据进行数据验证,判别模型的可行性.

2 文献综述

2.1 深度学习

CNN 可以应用在场景区分和图像分类,LeNet 是最早的 CNN 结构之一,主要用在字符分类问题(图 2). 由于在程序中使用了卷积运算,不仅可以提取图片的特征,而且卷积运算保持了像素之间的空间关系. 在 CNN 中,使用了滤波器作为特征提取器,而通过卷积得到的矩阵则称为“特征图”. 在选择特定的 CNN 时,应考虑到目标对象的图像特征,如农村建筑与城市建筑之间的差异,以及粗粒度建筑的情况. 因为现实世界的分类问题都是非线性的,而卷积运算是线性运算,所以在使用 CNN 来解决的时候,必须使用一个如 ReLU(或其他非线性函数,如 Tanh 和 Sigmoid,)的非线性函数来加入结果的非线性的性质,然后采用下采样的形式,提取经过 ReLU 的处理过后的特征值,或提取元素平均值或提取最大值,从而在保持图片重要的信息的同时降低特征图的维度. 最后通过全连接层(多层感知器),使用一个 softmax 激活函数,将前面卷积层提取到的特征结合在一起然后进行分类. 得出一个值 0-1 的向量,通过概率值来判断图片分类.

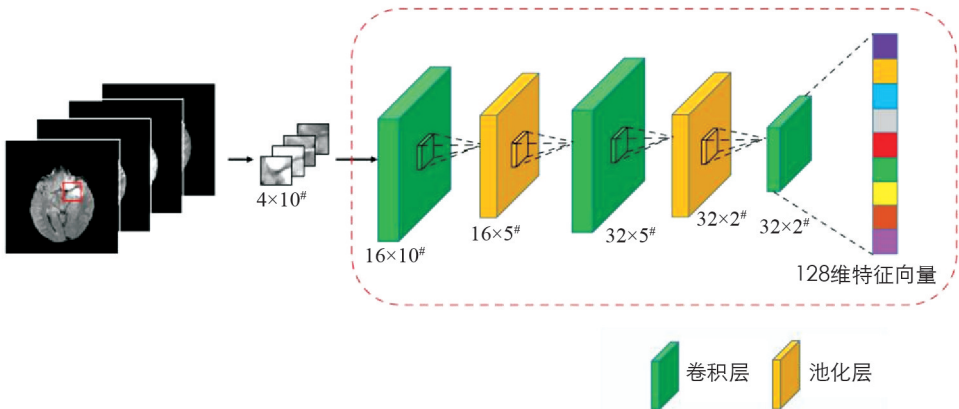


图 2 CNN 示意

近年来,深度学习方法,特别是 CNN 在各种计算机视觉任务中的表现已经超越了传统的方法,如其在目标检测、语义和图像分割方面均取得很好的研究成果.用标签标注图像像素的方法是基于图像中的语义来进行划分识别的,也就是说该算法从图像中将存在于图像中的,诸如汽车、树木或建筑物作为语义从整个图像中提取出来,并将每个语义进行标定.此外,在计算机视觉领域,有大量的研究是关于卷积神经网络中使用的各种模块,这些模块利用了“每个对象分类”的概念.这些模块,如卷积和金字塔池,提高了语义分割任务的算法性能.近年来,随着芯片处理能力(如 GPU 单元)的显著提高,计算硬件成本的显著降低,以及机器学习算法的显著进展^[15],深度学习在图像识别领域取得了迅速的进展,从而大大提高了计算机的处理能力.

2.2 深度学习在建成环境中的应用

随着城市更新从“增量时代”到“存量时代”的变迁,建成环境的数据以及与之对应的人类行为数据之间的关系愈发密切,大数据通过证明以人为本的视角、及时和实时的信息和精细分辨率的空间动态规律^[16].面对建成环境所产生的诸如遥感影像和街景图片的数据研究,经过过去几年的图像分析研究土地规划领域^[17],高分辨率(VHR)遥感图像的语义标签,为图像中的每个像素分配一个类别的任务,包括土地使用规划、基础设施管理和城市扩张检测,采用深度学习介入的方式已经被广泛采用.

随着深度学习技术研究的深入,街景图片因为其自身数量上的特征,并且其所附带的地理位置信息的特点,逐渐成为对于建成环境进行量化研究的重要数据来源.基于人本视角的空间量化评测成为研究的重要方向,包括对于街道的风貌特征、环境特征、建筑材料和功能的检测、建筑立面构件的语义分割与街景环境之间的关系. Gong 等^[18]开发了一种方法准确估计天空景观因子,使用公开的谷歌街景(GSV)图像和深度学习输出算法提取街道特征(天空、树木和建筑).此外, Gonzalez 等^[19]探索了使用 CNN 自动检测基于街道水平图像的建筑材料和横向负载抵抗系统类型的潜力.在本研究测试的 5 种网络架构中, ResNet50 的性能最好. Dai 等^[20]提出了一种新的集成模型,用于建筑立面组件的语义分割,以便对需求进行分类用于建筑改造的街景建筑立面图像数据集.

3 研究方法

本研究提出了一种新的基于深度学习的建筑分类识别框架,重点是更好地理解村落民居的建筑风貌划分.为此,首先建立了一个新的数据集,其中包含 420 个真实民居建筑风貌图片,其中每个建筑的特征属性都进行了手工标注.与之前的工作相比,本研究不仅标记了建筑的几何特征和风貌信息,还对建筑的材质特征和建筑质量特征进行了相应的信息标注.

3.1 数据收集

基于数据集的特点,设计了基于 ResNet 架构的 3 种不同的语义分割模型,并以集成的方式进行训练.

3.1.1 数据集和研究区域

民居建筑单体图片的采集主要采用 EOS 200D II(EF-S 18-55mm f/4-5.6 IS STM1)设备采集,由于多数民居均处于山地,为了更为全面的反映建筑自身特点,采用 3 个独立的方位进行拍照采集,每张图片像素均为 $2\,048 \times 2\,048$, 3 张图片形成一个覆盖建筑单体的图片数据组,反映建筑单体各项物理特征.用于构建数据集的图像是课题组在被列入世界遗产名录的红河哈尼梯田景区的核心部位——元阳多依树村进行的数据采集.目前还没有针对少数民族民居建筑类型的数据库,因此参考英国住宅建筑类型学数据库进行了特征分类.该数据库根据建筑类型对建筑特征进行分类,其定义的 3 种主要建筑类型包括单户住宅、多户住宅和排屋.此次研究形成的数据集共涵盖 420 栋独立的民居建筑,共计 1 260 张图片,其中有 133 张由于拍摄光线和镜头抖动的原因,被判为废图,有效图片共计 1 127 张.该数据集被分为训练集、验证集和测试集,其比例分别为 80%, 10% 和 10%.因此,训练集有 901 张图像,验证集和测试集各有 113 张图像.

3.2 数据预处理

结合民居建筑和深度学习特点,对图片分类的建筑特征指标包括建筑层数、建筑风格、建筑质量和建筑材料等 4 个特征进行实验.同时为了能够增强模型的泛化性,考虑到在实际的使用场景中,模型的识别率会受到天气、拍摄角度和光照等因素的影响,为了能够让数据的分布尽可能接近真实世界,本研究对数据进行了数据增强,方式有随机翻转、随机旋转、随机颜色和亮度变换等(图 3).



图 3 数据增强

3.3 基于深度学习的自动分类模型

考虑到传统民居建筑的特点和实际场景应用时对实时性的要求, 选用了 3 种卷积神经网络模型在本研究的数据集上做了充分的实验, 这 3 组模型分别是 MobileNetV3, ResNet50 和 EfficientNetB3a, 3 组模型的网络结构见表 1—表 3.

表 1 EfficientNet 网络结构

阶段	算子	分辨率/像素	# 通道/个	# 层数
1	Conv3×3	224×224	32	1
2	MBConv1, k3×3	112×112	16	1
3	MBConv6, k3×3	112×112	24	2
4	MBConv6, k5×5	56×56	40	2
5	MBConv6, k3×3	28×28	80	3
6	MBConv6, k5×5	14×14	112	3
7	MBConv6, k5×5	14×14	192	4
8	MBConv6, k3×3	7×7	320	1
9	Conv1×1&Pooling &FC	7×7	1 280	1

表 2 MobileNet 网络结构

种类/步数		过滤皿形状/像素	输入尺寸/像素
5×	Conv/s2	3×3×3×32	224×224×3
	Conv dw/sl	3×3×32dw	112×112×32
	Conv/sl	1×1×32×64	112×112×32
	Conv dw/s2	3×3×64dw	112×112×64
	Conv/sl	1×1×64×128	56×56×64
	Conv dw/sl	3×3×128dw	56×56×128
	Conv/sl	1×1×128×128	56×56×128
	Conv dw/s2	3×3×128dw	56×56×128
	Conv/sl	1×1×128×256	28×28×128
	Conv dw/sl	3×3×256dw	28×28×256
	Conv/sl	1×1×256×256	28×28×256
	Conv dw/s2	3×3×256dw	28×28×256
	Conv/sl	1×1×256×512	14×14×256
	Conv dw /sl	3×3×512dw	14×14×512
	Conv dw/sl	1×1×512×512	14×14×512
	Conv dw/s2	3×3×512dw	14×14×512

续表 2 MobileNet 网络结构

	种类/步数	过滤皿形状/像素	输入尺寸/像素
5×	Conv/sl	1×1×512×1024	7×7×512
	Convdw/s2	3×3×1024dw	7×7×1024
	Conv/sl	1×1×1024×1024	7×7×1024
	AvgPool/sl	Pool7×7	7×7×1024
	FC/sl	1024×1000	1×1×1024
	Softmax/sl	Classifier	1×1×1000

表 3 ResNet50 网络结构

层名	输出尺寸 /像素	18 层	34 层	50 层	101 层	152 层
conv1	112×112	第 49 层 7×7. 64. stride 2				
conv2x	56×56	3x3max pool. stride 2				
		3×3, 64	3×3, 64	1×1, 64	1×1, 64	1×1, 64
		3×3, 64×2	3×3, 64×3	3×3, 64	3×3, 64	3×3, 64
conv3x	28×28	1×1, 256×3				
		1×1, 256×3	1×1, 256×3	1×1, 256×3	1×1, 256×3	1×1, 256×3
		1×1, 256×3	1×1, 256×3	1×1, 256×3	1×1, 256×3	1×1, 256×3
conv4x	14×14	3×3, 128	3×3, 128	1×1, 12	1×1, 12	1×1, 12
		3×3, 128×2	3×3, 128×4	3×3, 128	3×3, 128	3×3, 128
		3×3, 128×2	3×3, 128×4	3×3, 128	3×3, 128	3×3, 128
Conv4x	14×14	1×1, 512×4				
		1×1, 512×4	1×1, 512×4	1×1, 512×4	1×1, 512×4	1×1, 512×8
		1×1, 512×4	1×1, 512×4	1×1, 512×4	1×1, 512×4	1×1, 512×8
conv5x	7×7	3×3, 256	3×3, 256	1×1, 256	1×1, 256	1×1, 256
		3×3, 256×2	3×3, 256×6	3×3, 256	3×3, 256	3×3, 256
		3×3, 256×2	3×3, 256×6	3×3, 256	3×3, 256	3×3, 256
conv5x	7×7	1×1, 1024×6				
		1×1, 1024×6	1×1, 1024×6	1×1, 1024×6	1×1, 1024×6	1×1, 1024×6
		1×1, 1024×6	1×1, 1024×6	1×1, 1024×6	1×1, 1024×6	1×1, 1024×6
conv5x	7×7	3×3, 512	3×3, 512	1×1, 512	1×1, 512	1×1, 512
		3×3, 512×2	3×3, 512×3	3×3, 512	3×3, 512	3×3, 512
		3×3, 512×2	3×3, 512×3	3×3, 512	3×3, 512	3×3, 512
conv5x	7×7	1×1. 205×3				
		1×1. 205×3	1×1. 205×3	1×1. 205×3	1×1. 205×3	1×1. 2048×3
		1×1. 205×3	1×1. 205×3	1×1. 205×3	1×1. 205×3	1×1. 2048×3
FLOPs	1×1	average pool1000-df softmax 第 50 层				
		1.8×10 ⁹	3.6×10 ⁹	3.8×10 ⁹	7.6×10 ⁹	11.3×10 ⁹
		1.8×10 ⁹	3.6×10 ⁹	3.8×10 ⁹	7.6×10 ⁹	11.3×10 ⁹

ResNet 通过残差模块解决了深度神经网络模型训练过程中梯度消失和爆炸的问题，在多个数据集上都表现出良好的性能，同样也适用于本次研究的传统民居数据集。MobileNet 使用深度可分离卷积在保证一定模型精度的前提条件下，大大降低了模型的参数，相对于本次实验的其他两组模型，参数量最少，更适用于实时的计算场景。EfficientNet 通过复合模型缩放的技术在网络结构上进行搜索，对卷积和的大小、神经网络的深度和分辨率大小进行调整，在保证一定效率的同时提高模型的精度，但是模型的参数较多，计算相对较慢，更适合于一些对实时性不高的应用场景。

本研究模型的实验环境为 Inter@core i7-9700K 处理器，显卡为 NVIDIA Geforce GTX3090 24GB，内存为 32GB，操作系统为 Ubuntu 20.04，64 位系统，CUDA11.3，CUDNN8.0.1，深度学习框架为 Pyotrchl. 8.0，编程语言为 Python3.8.5，集成开发环境为 PycharmCE2022，绘图工具为 Matplotlib1.3.1。

本次研究是基于深度卷积神经网络进行的，深度学习在训练之前需要保证超参数一致，在 4 组数据的 3 个模型实验上，测试修改了网络结构和最后的全连接层，其余的参数均保持一致。数据训练的批次大小为 4，轮数设置为 10，学习率初始为 0.001，采用余弦退火算法动态调整学习率，优化器使用的是 ADAM 优化器，采用迁移学习的方式对模型进行训练，在模型训练之前加载了模型在 Imagenet 上的预训练权重。值得一提的是，为了解决训练过程中样本类别分布不均的问题，采用了 Focalloss 的损失函数，相对于传统的交叉熵损失函数可以帮助模型快速收敛，该损失函数的公式为：

$$L = -y\log y' - (1 - y)\log(1 - y') = \begin{cases} -\log y', & y = 1 \\ -\log(1 - y'), & y = 0 \end{cases} \tag{1}$$

$$L_{\beta} = \begin{cases} -\alpha(1 - y')\gamma\log y', & y = 1 \\ -(1 - \alpha)y'\gamma\log(1 - y'), & y = 0 \end{cases} \tag{2}$$

$$L_{fl} = -(1 - y_t)\gamma \log y_t \tag{3}$$

Focal Loss 由 Facebook 人工智能研究院于 2018 年 2 月提出^[21], 本研究探讨了一阶段目标检测比二阶段目标检测精度差的原因, 即分类问题中类别不平衡以及分类难度存在差异, 并根据该原因提出了 Focal Loss 损失函数. 以二分类问题为例, 传统的交叉熵损失函数的定义见(1)–(3), 其中 y' 表示预测值, 即使用 Sigmoid 激活函数得到的预测值, y 表示的是真实标签, 取值为 0 或 1, 1 表示正样本, 0 表示负样本. 该损失函数存在的一个明显问题就是对于正样本而言, 输出概率越大损失越小, 对于负样本而言, 输出概率越小损失越小, 这样会导致模型在大量简单样本的迭代过程中参数更新缓慢, 并且无法优化到最优. 针对这一问题, 本研究在该损失函数的基础上引进了参数 γ 和参数 α , 通常 γ 取 2, α 取 0.25. γ 参数的主要作用是让模型关注难区分的样本, 以正样本为例, 当预测概率取值较大时, 函数的损失值会变得很小, 而当概率取值较小时, 损失函数值会变得很大, 从而降低了简单样本的影响, 增加了模型的训练速度和泛化能力. α 参数则主要是解决正负样本比例不均衡的问题, 用来平衡正负样本的重要性. 将二分类的 Focal loss 推广到多分类问题上, 可以得到(1)–(3)式所示的损失函数.

图 4—图 7 是 4 组模型在训练过程中验证集上 Loss 和 ACC 的变化, 其中蓝色曲线表示 MobileNetV3, 黄色曲线表示 ResNet50, 绿色曲线表示 EfficientNet. 从结果可以看出, 模型的 Loss 在第 5 个 Epoch 之后基本趋于收敛, 模型验证集上的 ACC 也在第 5 个 Epoch 之后趋于平稳, 并且从 ACC 的曲线中可以看出, 模型整体在数据验证集的表现中, EfficientNet 的效果最佳, MobileNet 的效果较差, ResNet 表现适中.

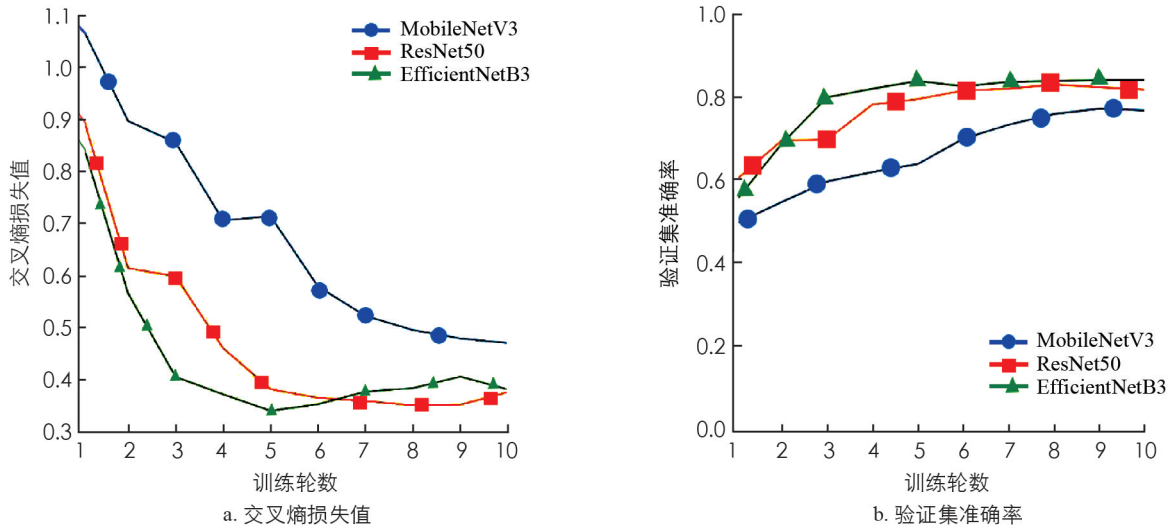


图 4 楼层实验

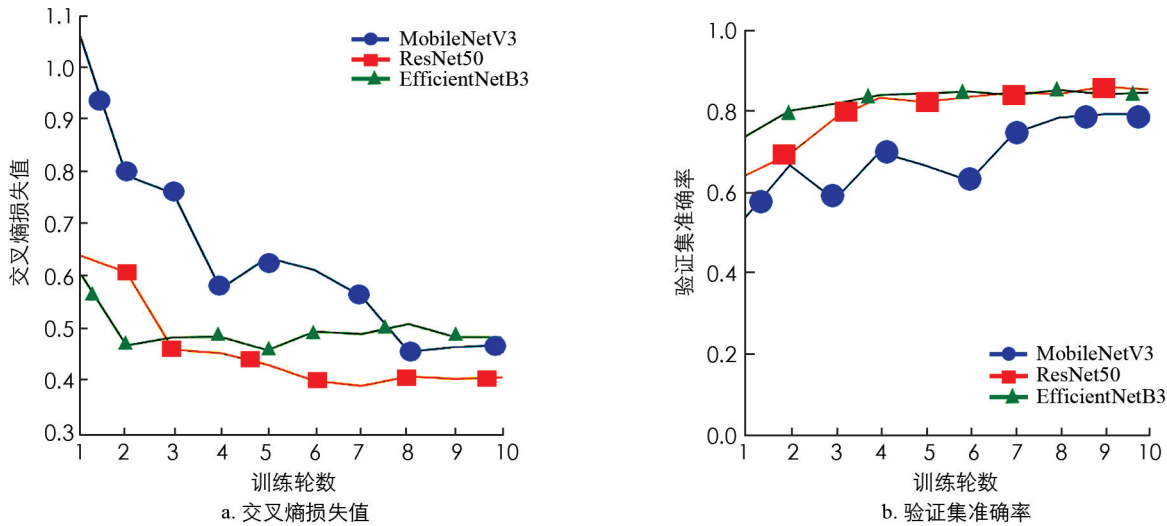


图 5 建筑风格

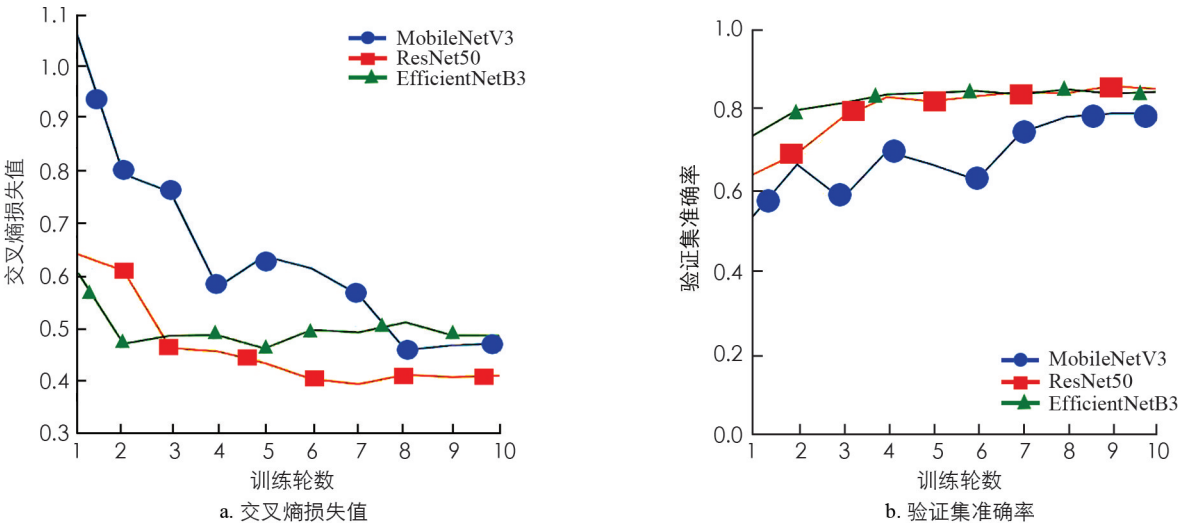


图 6 建筑质量

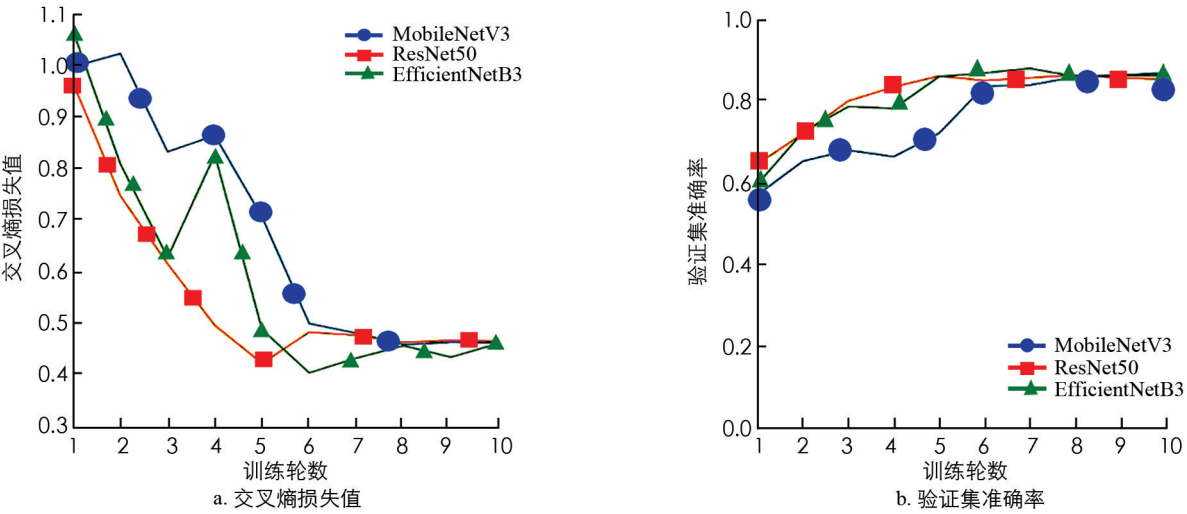


图 7 建筑材料

3.4 评价

通过本次实验对传统民居建筑特征进行量化研究，在评价指标上选用了分类模型中常用的 3 个指标，分别是准确率 P 、召回率 R 和综合评价指标 F_1 。

对于一个机器学习模型而言，可以将样本的真实类别和模型预测的类别进行组合，划分为如表 4 所示的 4 种情况，将 4 种情况下所对应的样本数相加得到的数量就是样本总数，即 $TP + FP + TN + FN =$ 样本总数。

表 4 样本真实情况和预测结果对照表

真实情况	预测结果	
	正例	反例
正例	真正例(TP)	假反例(FN)
反例	假反例(FP)	真反例(TN)

根据定义，可以衍生出查全率(recall，简写为 R)和查准率(precision，简写为 P)两个概念，计算公式为

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$R = \frac{TP}{TP + FN}$$

(5)

F_1 和 ACC 则是综合评价指标, 可以同时兼顾模型的精度和召回率, F_1 的定义如下:

$$F_1 = \frac{2 * R * P}{R + P}$$

3.4.1 性能指标

在测试集上对本研究的模型进行了测试, 测试的结果见表 5—表 7.

表 5 召回指标 R

模型	数据类型			
	建筑楼层	建筑风格	建筑质量	建筑材料
MobileNetV3	0.344	0.394	0.473	0.488
ResNet50	0.439	0.500	0.655	0.473
EfficientNetB3a	0.469	0.484	0.547	0.733
平均值	0.420	0.460	0.560	0.560

表 6 准确率指标 P

模型	数据类型			
	建筑楼层	建筑风格	建筑质量	建筑材料
MobileNetV3	0.767	0.765	0.790	0.870
ResNet50	0.829	0.872	0.851	0.865
EfficientNetB3a	0.838	0.887	0.837	0.881
平均值	0.81	0.84	0.83	0.87

表 7 F_1 分数指标

模型	数据类型			
	建筑楼层	建筑风格	建筑质量	建筑材料
MobileNetV3	0.421	0.434	0.557	0.515
ResNet50	0.454	0.474	0.556	0.487
EfficientNetB3a	0.499	0.557	0.506	0.755
平均值	0.46	0.49	0.54	0.59

3.4.1 速度指标

对训练得到的模型进行了速度上的测试(表 8), 测试的指标是 FPS , 即模型在 1s 之内能处理多少张图片, 测试的设备是 RTX3090.

表 8 速度指标

模型	数据类型				平均速度
	建筑楼层	建筑风格	建筑质量	建筑材料	
MobileNetV3	173	177	172	176	174.50
ResNet50	121	115	114	123	118.25
EfficientNetB3a	78	84	83	76	80.25

4 结果及讨论

4.1 结果

从结果上来看, 使用深度学习模型在 4 种建筑要素分类上的指标均可以达到 80 以上, 其中 Efficient-Net 模型在识别指标上表现最好, 综合指标表现最好; 而 MobileNet 在速度上表现最好, 但是精度相对较

低。从总体上看,模型在建筑风格、建筑质量和建筑材料的表现都较好,在建筑楼层的识别上表现较差,分析原因是楼层的特征对于其他 3 个建筑要素来说相对表面,深度神经网络提取的深度特征反而不利于楼层的识别,后续的研究可采用语义分割的方法来提高建筑楼层识别的准确率。另外,从总体指标上来看,本研究的指标并没有达到 90 以上,其原因是本研究的数据多采集自真实场景,在建筑周围有很多嘈杂的背景导致模型学习的相对困难,后续可以通过语义分割的方式提取干净的建筑前景区域来进一步提高识别的准确率和精度。

4.2 自动量化识别对民居建筑保护的价值

本次研究的数据来自 2021 年的元阳多依树村调研,就数据来源的时效性而言,能够很好地体现村落目前的保护情况。由于地理环境和民族地区发展的原因限制,云南本地传统村落的调研和保护受到很大的影响,现有的数据采集方法多依赖于人工现场调研,使得调研的成本变得昂贵,时间周期拉长,本次研究提出的研究方法,对采集到的数据集的分类和映射具有较高的自动化性,从目前的测试结果来看,能够对建筑的多项特征指标产生一个较好的评测结果,有助于减少未来村落民居建筑研究中因人工认知差异而造成的结果的不稳定性。云南省有 25 个少数民族,每个民族由于其自身的文化和地理位置造就了不同的建筑风格,但每种建筑风格都不同程度地因为发展的原因有了变化和异化,随着数据量的增加和具体参数的调整,本次研究的这种方法是可转移的,而且准确性会逐渐提高,可以为传统村落的保护政策研究提供数据上的量化参考。

4.3 局限性及未来的工作

本次的研究所采集的数据目前还是主要依靠人工进行现场拍照的形式进行,虽然与目前的图片数据采集技术(如 3 维激光采集、倾斜摄影数据采集)相比,人工采集的图片具有容量小、清晰度高和建筑细节准确性等特点,但在数据的采集过程中还存在一定的主观性和天气气候及光线的随机性,给数据的客观性带来了一定的波动性。此外,由于云南省的民族众多,村落所在地理环境复杂,导致民居形式丰富,对每个民族或每个区域的村落风貌的界定需要进一步地提高数据集数量,并且对模型的参数做进一步的调试,找到一定的规律性(如本次研究中,对增强数据的参数进行了多轮调试,在测试结果的 ACC 和 F_1 值上都取得了很好的提高),在此过程中,通过对参数的调试,可以看出 CNN 模式对图像的知觉认知的指标分类具有很好的潜力,这是研究下一步需要进一步加强和完善的地方。

5 结论

本研究提出了一个民居建筑风貌分类的研究框架,它能够对图片进行更丰富的分类。该方法对单体民居建筑物的特征分类具有较高的识别精度。为了完成这个任务,本研究构建了一个包含 4 个民居建筑特征类别的人工采集的基准数据集,用于训练和测试。通过研究 3 种不同的 CNN 架构,选择综合指标表现最优的 EfficientNetB3a 进行村落尺度上的民居建筑特征分类。这样的研究结果有助于了解村落民居发展和异化的规律。

为了提高模型的泛化力和对图片分类性能的准确性,未来的工作可以融合其他信息,如社交媒体图像和图像中显示的文本信息,又如图片标题和图片附带的情感语言信息。此外,为了获得更密集的民居村落图片,可以结合遥感影像图像和附带地理信息的倾斜摄影图像数据(建筑数据单体化)进行多源的数据综合测评研究。在此基础上,由于有了量化的数据测评,如果在后续的研究中加入时间维度的考量,就能够对村落民居风貌的变化有更为准确的量化描述,对于村落保护的监测具有更为重要的意义。

参考文献:

- [1] 许声宏. 在线教学中的人工智能应用初探 [J]. 西南师范大学学报(自然科学版), 2022, 47(9): 111-115.
- [2] 代建华, 彭若瑶, 许路, 等. 基于深度神经网络的信息抽取研究综述 [J]. 西南师范大学学报(自然科学版), 2022, 47(4): 1-11.
- [3] 张丽英, 裴韬, 陈宜金, 等. 基于街景图像的城市环境评价研究综述 [J]. 地球信息科学学报, 2019, 21(1): 46-58.
- [4] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet Large Scale Visual Recognition Challenge [J]. International Jour-

- nal of Computer Vision, 2015, 115(3): 211-252.
- [5] 董华珍. 基于 CNN 的普米语孤立词语谱图分类 [J]. 西南大学学报(自然科学版), 2021, 43(2): 160-168.
- [6] 吴绪玲. 基于混合优化的双模深度学习文本分类方法 [J]. 西南大学学报(自然科学版), 2022, 44(11): 234-242.
- [7] 刘伦, 王辉. 城市研究中的计算机视觉应用进展与展望 [J]. 城市规划, 2019, 43(1): 117-124.
- [8] HERAS L P, AHMED S, LIWICKI M, et al. Statistical Segmentation and Structural Recognition for Floor Plan Interpretation [J]. International Journal on Document Analysis and Recognition, 2014, 17(3): 221-237.
- [9] LIU C, WU J J, KOHLI P, et al. Raster-to-Vector: Revisiting Floorplan Transformation [C]//2017 IEEE International Conference on Computer Vision (ICCV). October 22-29, 2017, Venice, Italy. IEEE, 2017: 2214-2222.
- [10] YAMASAKI T, ZHANG J, TAKADA Y. Apartment Structure Estimation Using Fully Convolutional Networks and Graph Model [C]//Proceedings of the 2018 ACM Workshop on Multimedia for Real Estate Tech, 11 June 2018, Yokohama, Japan. New York: ACM, 2018: 1-6.
- [11] ZENG Z, LI X, YU Y K, et al. Deep floor plan recognition using a multi-task network with room-boundary-guided attention, in: Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 9096-9104.
- [12] HERAS L P, TERRADES O R, ROBLES S, et al. CVC-FP and SGT: a New Database for Structural Floor Plan Analysis and Its Groundtruthing Tool [J]. International Journal on Document Analysis and Recognition, 2015, 18(1): 15-30.
- [13] KALERVO A, YLIOINAS J, HÄIKIÖ M, et al. CubiCasa5K: A Dataset and an Improved Multi-Task Model for Floorplan Image Analysis [M]//Image Analysis. Cham: Springer International Publishing, 2019: 28-40.
- [14] WU W M, FU X M, TANG R, et al. Data-Driven Interior Plan Generation for Residential Buildings [J]. ACM Transactions on Graphics, 2019, 38(6): 1-12.
- [15] PORIKLI F, SHAN S G, SNOEK C, et al. Deep Learning for Visual Understanding: Part 2 [from the Guest Editors [J]]. IEEE Signal Processing Magazine, 2018, 35(1): 17-19.
- [16] KONG L Q, LIU Z F, WU J G. A Systematic Review of Big Data-Based Urban Sustainability Research: State-of-the-Science and Future Directions [J]. Journal of Cleaner Production, 2020, 273: 123142.
- [17] MA L, LIU Y, ZHANG X L, et al. Deep Learning in Remote Sensing Applications: a Meta-Analysis and Review [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 152: 166-177.
- [18] GONG F Y, ZENG Z C, ZHANG F, et al. Mapping Sky, Tree, and Building View Factors of Street Canyons in a High-Density Urban Environment [J]. Building and Environment, 2018, 134: 155-167.
- [19] GONZALEZ D, RUEDA-PLATA D, ACEVEDO A B, et al. Automatic Detection of Building Typology Using Deep Learning Methods on Street Level Images [J]. Building and Environment, 2020, 177: 106805.
- [20] DAI M L, WARD W O C, MEYERS G, et al. Residential Building Facade Segmentation in the Urban Environment [J]. Building and Environment, 2021, 199: 107921.
- [21] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal Loss for Dense Object Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 15-25.

责任编辑 潘春燕