文章编号：1673-9493（2019）01-0008-10　中图分类号：TU984.13　文献标识码：A　DOI：10.22217/upi.2018.514

深度学习在城市感知的应用可能
—— 基于卷积神经网络的图像判别分析

The Latent Application of Deep Learning in Urban Perception: Image Discrimination Analysis by Convolutional Neural Network

何宛余　李春　聂广洋　杨良崧　王楚裕
He Wanyu, Li Chun, Nie Guangyang, Jackie Yong Leong Shong, Wang Chuyu

摘要：作为人工智能领域的研究重点，机器学习近年衍生出了各式各样的智能化应用，例如图像判别、语音助手和智能翻译等。尤其是图像判别技术已在各行业进行了大量的研究和实践，城市领域也不例外，这很大程度上是因为深度学习的卷积神经网络在计算机视觉领域取得了令人瞩目的成果。这也使得训练计算机判别建筑风格、城市肌理等城市特征的准确率大幅提升。本研究立足于深度学习图像判别技术，探索卷积神经网络在城市感知方面的应用。鉴于直接利用现成开源的带标签图像数据集训练个性化图像判别模型可能带来局限性和误差，本研究探索了从收集数据到自定义训练数据集，到搭建满足特定需求的图像判别模型的整体流程，并通过三个实验案例：城市风貌分析、城市问题侦测和城市肌理评估，阐明深度学习在城市感知和城市规划中的应用可能性及潜力。

Abstract: Nowadays, machine learning attracts intense attention from artificial intelligence researches and extends a variety of applications such as image discrimination, voice assistant and smart translator. In particular, image discrimination has been extensively studied and practiced in various industries, including urban field. Thanks to Convolutional Neural Network (CNN) based on Deep Learning (DL) that has made remarkable achievements in computer vision, it is more efficient to train computer to discriminate architecture styles, urban texture and other urban features. Based on image discrimination by DL, this research focuses on exploring the applications of CNN in the field of urban perception. In consideration of limits and errors brought by training customized image discrimination model with the existing open source labeled image dataset, this paper explores a whole process from collecting data, self-constructing training dataset to building a customized image discrimination model which satisfies specific requirements. The latent application of DL in urban scale are discussed through three experiment cases: the cityscape analysis, urban problem detection and urban pattern evaluation.

关键词：人工智能；深度学习；卷积神经网络；图像判别；城市感知

Keywords: Artificial Intelligence; Deep Learning; Convolutional Neural Network; Image Discrimination; Urban Perception

作者：何宛余，小库科技，创始人兼首席执行官
李春，小库科技，联合创始人兼首席技术官
聂广洋，小库科技，人工智能学家
杨良崧，小库科技，高级研究员
王楚裕，小库科技，智慧城市高级研究员

引言

传统上城市根据较为静态的规范或导则来进行规划和管理。城市化进程的加快及城市系统的日益庞大复杂，一方面不断增加规划管理多层级公共系统和城市资源的成本，另一方面暴露了自 20 世纪起建立的规划设计范式难以适应现今城市快速动态的更新变化[1]。在这样的背景下，探索并设计一个智能的城市感知计算模型具有重要意义。

智能化的城市感知计算模型离不开人工智能技术及数据的支持。人工智能自身的发展从 1960 年代的基于推理[2]，到 1980 年代的知识驱动（knowledge-driven）[3]，再到现今的数据驱动，这种范式转变也随之影响了“智能”的城市计算模型[4]。早期基于推理的人工智能依赖清晰定义的推理步骤及变量，使得它在处理复杂问题方面有很大局限性。知识驱动的人工智能将行业知识结构化、数字化，使计算机可以基于先验知识推理计算，代表性的应用有专家系统（expert system）。然而对于逻辑知识体系十分庞大的领域，依赖人为整合知识需要投入大量的时间和人力。1990 年代起，人工智能迈入数据驱动时代，出现了各类可通过数据自主学习的机器学习（machine learning）算法。近几年，作为机器学习重要分支的深度学习框架取得突破性进展，在图像识别、语音识别、无人驾驶等领域取得了令人瞩目的成果。与此同时，随着万维互联网向移动互联网转变，多维度数据大量产生，并在网络中沉淀下来，作为可用于训练机器学习模型的大规模数据集，为城市感知计算模型的智能化升级提供了契机。

基于深度学习的图像判别模型可以从大规模图像数据中学习物体特征，进而自动分类或识别图像中的物体，

这为计算机视觉技术的发展翻开了新的篇章。对于包含大量图像数据的城市而言，图像判别模型的介入将大大提高城市在管理和预测方面的效率和准确性。例如利用人脸识别技术监测闯红灯的行人[5]、通过分析摄像头影片侦测违规驾驶者或者城市安全隐患等。城市规划管理的其他方面也可借助机器学习或深度学习算法提高效率，如建立犯罪率模型来预测盗窃率，设计人流量分析模型来预测人口密度和异常事件[6]。随着大数据和人工智能技术的发展，智能的城市感知计算模型将拥有巨大的潜力和广阔的前景。

本文将在第 1 节简明介绍深度学习的计算原理，为第 3、4 节中的图像判别模型建立理论基础。在应用探索方面，第 2 节综述了训练数据集的不同类型，并在第 4 节以三个实验案例探讨第 3 节提出的图像判别模型在城市感知方向的应用。

1 深度学习神经网络

如今人工智能技术在特定的感知计算方面极为高效，甚至有超越人类的表现，这得益于三个要素。第一，计算机硬件性能极大提升，很多过去看似不能实现的、依赖大内存及高运算能力的机器学习算法变得可行。第二，在互联网、移动互联网及物联网普及下，数据的量级、丰富度和可达性大幅提升，为机器学习模型提供了大量训练数据。第三是基于人工神经网络（Artificial Neural Network）的机器学习算法的优化，其中建立在神经网络基础上的深度学习是一种对数据进行表征学习的算法[7]。其“深度”体现在神经网络的层数上，该多层结构让计算机能够对图像、音频、视频等数据进行由简到繁的感知计算。尤其是深度学习中的卷积神经网络（CNN: Convolutional Neural Network）算法在计算机视觉领域硕果累累[8]。理解当今深度学习在人工智能领域的地位及其原理有助于帮助我们洞察它在城市规划设计领域的应用可能。本节将对深度学习的发展、基本原理及卷积神经网络进行简明阐述。

1.1 深度学习网络的特点和优势

传统的人工神经网络包含三层：输入层（input layer）、隐藏层（hidden layer）和输出层（output layer）。不同层之间的神经元关系由计算机从数据中自主“学习”得出。这种数据处理模式非常依赖人为的特征表示（feature representation），即人为定义代表某概念的有效信息。

深度学习把神经网络的拓扑结构由三层扩展到多层，从而打破了三层的神经网络无法从较原始的数据中学习到未被表示的抽象特征这一局限（图 1）。在深度学习模型中被增加的是中间隐藏层的数目，用于自动提取特征。例如在运用多层网络进行图像判别时，第一层负责从输入的图像像素中学习基本的线条。第二层则利用第一层的结果，学习判别简单形状（如圆形）。每升高一层就学习更多的特征，如动物肢体的数量、皮肤的纹理等。而“学习”过程实质上是一个在各层神经元之间拟合出最适合的权重关系的过程。

除了算法的优化，深度学习近几年得到重视和极速发展还因为其在处理大体量数据上的优势。传统的机器学习算法在数据量级到达一定程度后，便会进入性能停滞期，而大型多层神经网络的性能却能随着数据量级的增长而提高（图 2）。

1.2 卷积神经网络原理

卷积神经网络（CNN）被广泛应用于当今的图像判别领域。图像数据在 CNN 中被分解成不同像素大小的层，该模型从输入的图像数据中学习输入与输出之间的映射关系，而不需要人为干预特征的精确表达过程（图 3）。输入层输入的是由图片像素转化而来的数组；隐藏层包括多个卷积层（convolution layer）、池化层（pooling layer，也称采样层）和全连接层（fully connected layer）等等，可看作由相连的神经元组成的神经网络；而输出层则为判别结果。

在经典的 CNN 模型中，来自输入层的数据首先进入第一个卷积层。卷积层带有与之对应的过滤器（filter），它是

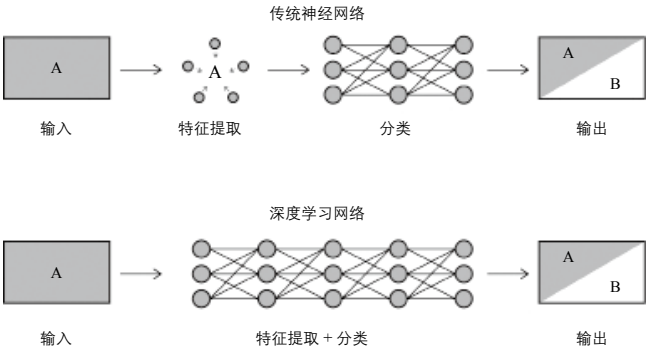

图 1 深度学习不需要人工参与特征表达环节
资料来源：作者基于参考文献 [9] 重绘

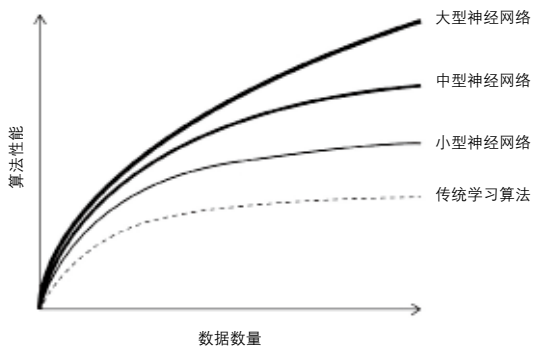

图 2 不同规模的神经网络学习算法与传统机器学习算法的性能比较
资料来源：作者基于参考文献 [10] 重绘

一个数字矩阵。在卷积层，输入矩阵与过滤器矩阵卷积相乘得到一个新的矩阵，即特征图（feature map）。卷积运算的性质使特征图保留了原图中像素与像素间的关系。特征图在池化层被降维处理（图 4），进入下一个卷积层，然后再次被池化。带有不同过滤器的卷积层实现了对图像的多种操作，例如边缘检测、轮廓检测、模糊化、锐化等等。如此经过多次卷积与池化后，数据抵达全连接层。全连接层通过激励函数（如带损耗的逻辑回归函数）对图像数据进行分类，最终输出结果以概率表示输入图像属于某个类别的可能性大小。

由于其自主进行特征提取的能力及优良的性能表现，CNN 不仅被广泛用于图像判别，也被大量应用于非视觉领域，如自然语言处理、药物结构探索和棋类攻略的学习等。基于 CNN 的模型训练仰赖一定的数据量级，所以近几年拥有足够数据量级和算法技术的城市管理部门、学术研究机构和科技企业等，纷纷开展借助机器学习算法建立城市数据分析或预测模型方面的研究。

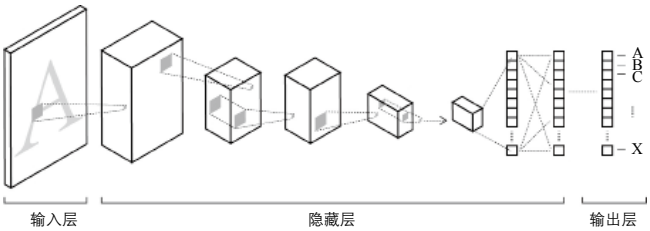

图 3 卷积神经网络结构
资料来源：作者基于参考文献 [11] 重绘

2 训练 CNN 图像判别模型的数据集类型

部分社会研究者认为，城市外观在很大程度上与人们评价城市的各项指标有着密切的关系 [12]。由于人类感知城市最直接常用的是视觉，因此近年模拟人类视觉感知城市面貌的技术引起城市研究人员的重视。然而过去依赖人为特征表示的算法模型在描述像城市面貌这样的抽象概念时往往较为困难。CNN 图像判别模型作为可以自主提取特征的学习模型，只需简单的人为干预即可在庞大量级的图像数据与抽象概念之间建立关系，从而实现自动判别。由计算机从海量数据中自主提取特征、规律、模式等，意味着对训练数据的量和质有一定的要求。目前在城市问题研究领域用于训练图像判别模型的数据集有两种类型：一是使用（包括直接使用或再进行微调）第三方已标签好的图像数据集；二是收集图像数据后，对图片进行自定义标签来构建训练数据库。

2.1 现成带标签图像数据集

遥感、低空航拍等技术的发展使得地表图像数据成为研究城市问题理想的数据源。利用有监督学习训练图像判别模型需要预先对图像数据打上标签（label），以标记图像的所属类别。然而当数据量级较大且人工和时间成本有限时，为图像标注标签是一项繁重的任务。出于提高行业效率的考虑，一些机构向大众提供带标签的开源图像数据集，如

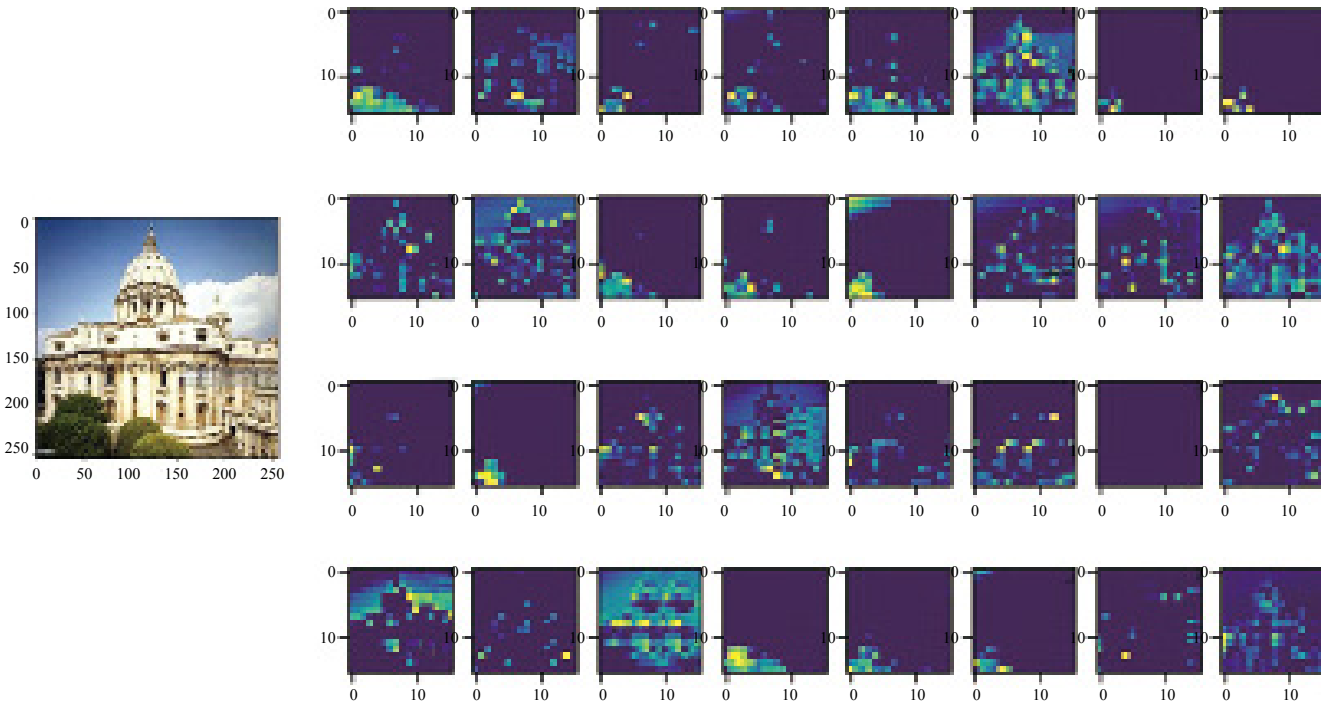

图 4 卷积神经网络中某隐藏层的图像样貌

ImageNet ①、DeepSat ②和 UC Merced ③等，其中后两者是专为研究土地利用而设的数据集。

目前有不少研究利用开源的遥感图像数据库分析城市问题。例如乌巴（N. K. Uba）在用深度学习技术分析用地类型和土地覆盖的研究中，以微调（fine tuning）方法改良了用 ImageNet 数据库训练的图像识别模型，使之在城市遥感图片的识别上获得平均高于 95% 的准确率 [13]。这种方式相当于先利用现成的带标签图像数据进行训练，然后再花相对少的时间筛选和标注部分目标图片（如森林、湖泊、建筑等的航拍图）对模型进行微调训练，最终使模型在识别目标类别上达到满意的精度。希腊国家理工大学的帕帕多曼诺拉基等（Papadomanolaki et al.）也采用了相似的方法，通过 DeepSat 数据集训练经典的 CNN 模型 AlexNet、AlexNet-small 和 VGG，验证了 CNN 模型在卫星遥感图片中识别荒地、树木、草地、公路、建筑和水系六大类别的地表事物均有超过 99% 的准确率 [14]。

裴纳缇等（Penatti et al.）则基于 UC Merced 和另一个遥感图像数据库（Brazilian Coffee Scenes）④来比较 CNN 用不同的数据库训练的表现。研究结果显示不同的训练数据集会导致 CNN 的性能表现不同，这源于不同数据集之间的差异，如 Coffee Scenes 的图像相对高清，UC Merced 卫星图像数据的噪音较大等 [15]。其中 UC Merced 的卫星图像数据集根据用地属性预设了 21 种分类组，其中某些数据组间有重叠，使得训练出来的模型存在误判的可能 [16]。

虽然通过这些数据库训练而成的图像识别模型具有不错的准确率，但这类数据库的普适性特点使得它们的分类颗粒度较大，因而面对特定的领域或者问题时，可能出现缺少图像类别或分类欠细致等问题。另一方面，对于非物质的概念，例如城市的舒适度，与审美或心智相关的抽象概念等目前仍没有普遍可用的带标签数据库。为了建立图像数据与抽象概念间的关系，往往需要自行收集图片并人为进行标签以构建合格的训练数据库。例如为了训练图像识别模型从一组街景图像中分辨出哪些元素具有“美”的价值，需要有针对性地收集、整理和标注图片，从而构成可服务于特定用途和需求的训练数据集。

2.2 自定义训练图像数据集

自定义训练数据集的特点体现在两个方面：一是收集方式，包括通过关键词从互联网中爬取收集和通过无人机航拍等方式重新收集；二是打标签的方式，包括依靠研究人员手动给数据打标签和通过评分游戏等方式让公众参与数据标签，即所谓的众包方式。自定义数据集可以很好地满足特定应用场景的分析需求，所以研究重点需集中于如何构建高质量的训练数据集及如何建立与场景匹配的分析模型。

在自定义训练图像数据集的探索上，申乔木等人试图基于城市街景了解人对城市形态的感觉，为此设计了一个可与市民互动的视觉分析系统 StreetVizor。该系统允许公众参与城市空间的评分（在系统后台可转化为对城市街景图的标签），收集了人们在城市中的感受。另一方面从谷歌街景图像（GSV: Google Street View）收集的街景图像被输入开源的图像语义分割深度网络 SegNet，以判别其中六种特征（绿植、天空、建筑、道路、交通、其他）的大致轮廓。然后每种特征元素的像素面积被计算，对这些特征在图像中所占的比例进行统计分析后便可以了解城市中各街道的特点，如某些街道的绿化较多，某些则是建筑密度较高 [17]。街景图像的分析结果与人们的评价结果相结合可以帮助研究人员发现广受好评的街道的特点。研究指出该系统在城市规划设计中还有更多的应用可能性，例如指出潜在需改善的范围、比较各城市的规划（案例）、辅助设计步行体验更好的道路、检测城市环境问题等等。

叶宇等人的研究则提出借助一个基于 SegNet 建立的图像判别模型，量化评估城市街道的绿视率 [18]。该项研究沿着城市街道选择取样点，以各取样点四个方位的 GSV 构建训练数据集。街景图在 SegNet 中被按颜色划分区块，因此绿色的树木、草地等可以被识别出来。模型进一步结合专家们的评分可实现对街道绿视率的量化衡量。

为了建立一个普适的探索城市形态与市民感知之间关系的街景判别模型，杜贝等（Dubey et al.）基于前人的实验基础（Place Pulse 1.0）继续细化训练数据集（Place Pulse 2.0）[19]。通过评分游戏的形式让公众对 10 万张横跨 56 座城市的街景图像进行 6 个抽象类别的标签工作：安全、活力、无趣、富裕、低落和美丽。纳达依等（Nadai et al.）也试图自己建立图像判别模型来研究被标记为“安全”和“活力”的城市是否具有共同或相反的特征 [20]。结果表明两种城市元素——面向街道的窗户和绿化——可使城市同时满足安全和活力两个特征，这也验证了简·雅各布斯（Jane Jacobs）所提出的“街道眼”理论。在鼓励公众参与对城市空间评分方面，塞雷辛赫（C. I. Seresinhe）团队在英国也进行了类似探索。他们借助 CNN 图像识别模型从英国各地的 20 多万张户外图片中提

① http://www.image-net.org/
② http://csc.lsu.edu/~saikat/deepsat/
③ http://weegee.vision.ucmerced.edu/datasets/landuse.html
④ http://www.patreo.dcc.ufmg.br/downloads/brazilian-coffee-dataset/

取出数百个特征，然后让公众在网上对包含这些特征的场景评分，以进一步加深对市民所认为的城市中美丽的户外空间的理解[21]。

由此可见，合适的训练数据可充分发挥 CNN 图像判别模型在城市规划管理方面的潜力。可以预见图像判别模型将成为政策制定者和城市设计者有力的辅助工具，以实现城市资源的合理分配，因此亟须提出一个从训练数据集构建，到图像判别模型搭建，再到实际应用的整体框架流程。下文除了提出搭建自定义图像判别模型的简单方法和实验，也讨论模型在多种场景的应用可能性。

3 自定义图像判别模型建立方法

由于特定需求难以依靠目前已有的带标签的数据库来建模分析，因此我们需要根据不同的任务需求来收集、清洗和进行数据标签以搭建图像判别模型。

首先要选择训练数据来源，经过对比研究数个图像数据搜索引擎谷歌、百度、必应和图像社交平台（如 Pinterest）等，选择谷歌图像搜索作为图像收集工具。针对不同的应用场景爬取相应的图像数据后，进行清洗工作（剔除不相关的图像），最后得到各场景对应的数据量从 1 500~5 000 张图片不等。训练前，根据常见的机器学习模型所需的训练数据量，每组图像数据按照以下比例划分：75% 作为训练数据集，15% 作为验证数据集，10% 则作为测试数据集。首先用训练和验证数据集作为输入来训练模型，训练完成后，输入测试数据得到的测试结果（判别结果）可用于估计模型在实际使用时的效果。

本研究中的图像判别模型利用常见的开源神经网络库 Keras 搭建。根据 CNN 原理，模型包含三个主要部分：输入层、由三组卷积池化层组成的隐藏层和全连接层。如图 5 所示，进入输入层的数据是由二维彩图（RGB 色彩模式）转化而成的数组，其大小由图像的分辨率乘以 RGB 位数决定（宽 × 高 ×3）。被输入数组先后进入三组卷积池化层，每组卷积池化层包括以下三个子层：(1) 卷积层。可根据实际应用调节过滤器大小、步长及填充方式，其中过滤器的选择决定每次取样的范围，步长决定每次取样移动的像素个数，填充方式有零填充和抛弃填充（用于处理过滤器大小与图像大小不符的情况）。(2) 激活层。可选择不同的函数对数据进行非性线处理，较为常用的是修正线性单元（ReLU: Rectified Liner Unit）。(3) 池化层。在保留图像像素间关系的同时对图像数据进行压缩，有最大池化（max pooling）、均值池化（average pooling）与求和池化（sum pooling）三种方式。最后数据通过全连接层得到代表分类结果的数值（本模型预设的数值区间介于 0~1 之间）。本研究使用的 CNN 框架，允许用户为特定的任务构建相应的训练集、测试集以及自定义需要判别的类别，从而满足不同的应用需求（图 5）。

训练验证模型的结果可用于性能评估，模型准确率的度量标准有训练阶段的准确率和损失值（表示为 acc 和 loss）和验证阶段各的准确率和损失值（表示为 val_acc 和 val_loss）。评估时，判断模型是否过度拟合可以通过分别比较训练样本与验证样本的准确率和损失值之间的差值。如果训练与验证阶段的准确率数值差异较大，意味着模型过度拟合；而训练和验证阶段的准确率越低，则意味着图像判别效果未达到令人满意的程度。模型使用二元交叉熵（binary crossentropy）为目标函数，让模型以达到损失值最小化为目标进行更新迭代。因此损失值越小意味着训练出的模型对数据拟合得就越好。在本次研究中，各案例的训练及验证结果纪录在表 1 中，测试结果则体现在以下各案例的可视化分析图中。

4 自定义图像判别模型在城市领域的前沿探索

国际上已有不少运用图像识别模型分析城市问题的案例，但它们多数采用国外的城市图像数据进行训练和测试；而国内城市拥有独特的风貌，直接运用国外的图像识别模型难以达到理想的准确度。所以，本研究通过引进国外的技术搭建适合分析国内独特城市风貌的图像判别模型，并针对三个应用场景：城市风貌分析、城市问题侦测和城市肌理评估，开展基于 CNN 的图像判别技术在城市感知方面的应用探索①。

4.1 城市风貌分析

目前的城市图像数据反映最多的是城市风貌。除了常见的街景，建筑立面也可作为分析城市市容的重要因素。例如杜尔斯等（Doersch et al.）通过分析巴黎的街景和建筑立面数

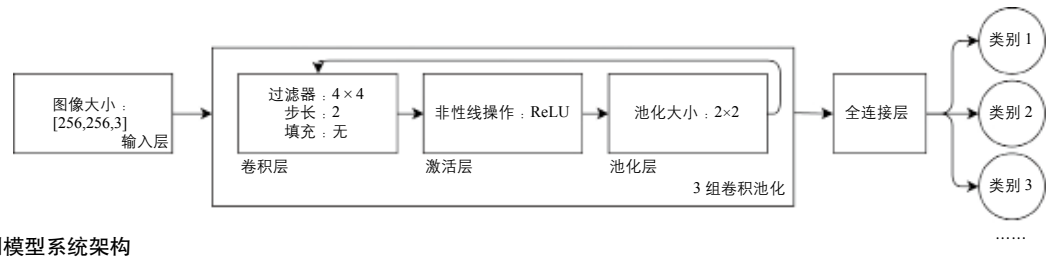

图 5 图像判别模型系统架构

① 以下案例包括小库科技在 2018 年同济大学“数字未来”活动中与学员的部分共同成果。

据，挖掘出能代表当地特色的视觉元素，并发现这些视觉元素多出现在市井的生活场景中[22]，为研究巴黎城市特色提供了方向。为验证 CNN 图像判别技术在识别城市建筑风格上的潜力，本案例选择上海作为研究对象。由于独特的地理和历史优势，上海拥有迥异的建筑风格，从西方古典到现代以玻璃幕墙为主的立面，因此是一个极佳的城市风貌研究对象。

4.1.1 模型设置

在模型设置上，利用第 3 节提出的系统架构搭建并训练一个能判别西方古典和现代幕墙这两种风格的图像判别模型。图像数据来源方面，由于上海本地的图像数量不够充足，这两种建筑立面风格的训练数据并非仅来源于上海，部分数据是借助谷歌关键字爬取的来自全球各地的建筑立面数据，以亚洲地区的数据为主。数据的筛选规则是以无人像、拥有较完整的建筑立面为主，允许旋转角度和透视变化。每种风格的图片大约各 1 000 张，共 2 000 张左右的图片被赋予“古典”和“现代”的标签，按预设比例划分训练、验证和测试集（75 ： 15 ： 10）①。

4.1.2 模型训练和测试结果

西方古典风格和现代幕墙风格的建筑立面图片经过模型处理被区分出来（图 6）；模型准确度如下文表 1 所示，训练集和验证集的准确率都达到了让人满意的程度，分别是 0.9248 和 0.8520，并且两者之间的差值较小，过度拟合的程度很低。训练和验证阶段的损失值也在满意的范围内。

得到足够精度的建筑风格判别模型后，本课题的研究人员在上海黄浦区外滩选取其中 6 个街区的建筑立面图像数据作为测试输入。分类结果为 [0~1] 之间的数值，越靠近 0 表示越可能是西方古典风格，越靠近 1 则表示越可能是现代幕墙风格。将建筑风格的判断结果按照颜色深浅进行可视化渲染，发现建筑风格的演变在一定程度上与上海的发展趋势吻合（图 7）。因此我们可以借此作出假设：上海的建筑立面可以反映其建成时间。

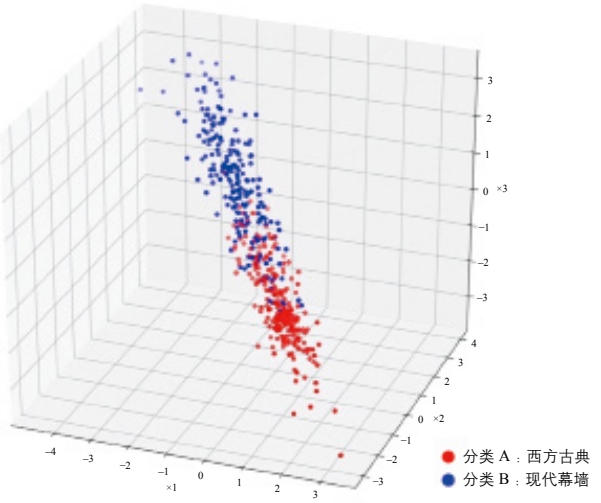

图 6 两种建筑立面的图像判别模型数据分布三维图
资料来源：赵珂、刘钰绘制

4.1.3 模型改进和应用可能

基于实验结果，该案例的优化目标可聚焦于建立建筑风格与城市演进的关系模型。实现方式是在建筑立面风格判别模型的基础上将图像与概念（可以是建成年份或年代）通过人为标签建立关联性，使得新模型接收新的输入后输出建筑建造年代，从而实现在城市空间图中呈现城市发展的时间轴。要完成这一优化，主要制约在于复古建筑风格与现代风格的时间跨度较大，但本案例中针对两类建筑风格的深度 CNN 图像判别模型尚不能实现更细粒度的建筑年代划分，因此模型的改进需要更充足的数据，同时需要将二元建筑风格数据进一步拆分细化并进行人为标签。

基于城市风貌图像的判别模型的应用场景可包括虚拟导游产品[23]、基于内容的图像检索（CBIR: Content Based Image Retrieval）、建筑数据库的指标参考、三维城市建模[24]等。

4.2 城市问题侦测

城市问题的一个重要观测因素是活力指数。假设街道活力越大，则该区域商业价值越高；反之该区域的犯罪率或者安全隐患越高。当时间的划分粒度很细时，街道或区域的活力可以直观地通过人流量反映，然而当以更大的时间跨度（几年到几十年）去分析区域的活力变化时，建筑、

图 7 建筑风格和城市发展进程比较分析
资料来源：同图 6

① 计算机的配置为：处理器（1.80 GHz Intel Core i7）；显卡（NVIDA Quadro P500）；内存（8.0 GB）。

公共设施等静态的物件可以被认为是重要的分析对象。本案例尝试对街景图像的活力度进行简单的二元分类（商业活力与消极）。

4.2.1 模型设置

本案例的数据集有两个来源：百度地图的人类视角街景图（对街道店面图进行个别裁剪）和谷歌图像数据。谷歌图像搜索关键词以英文为主，将关键词分为三类进行交叉组合搜索：活力度、空间承载和地理位置。活力度方面的关键词包括知名的（popular）、开放的（opened）、商业（commercial）等积极词汇，而消极词汇有空置的（vacant）、关闭的（closed）、拆毁的（demolished）等。空间承载词汇包括店面（store front）、底层立面（ground façade）、街道层级（street level）等，而地理词汇以上海地区为主（为了达到一定的数据量级，也收集了东京等个别亚洲城市的数据）。图像筛选规则允许人像的存在（因为人的数量也可反映活力）。活力与消极的街景图数量各为约 1 500 张，在进行标签后以 75 ： 15 ： 10 的比例划分训练、验证和测试集[①]。

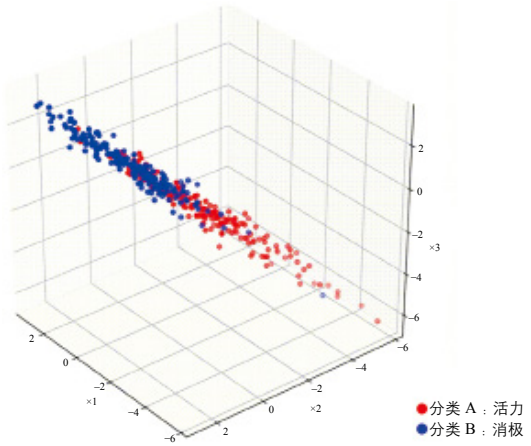

图 8 两种活力属性立面的图像判别模型数据分布三维图
资料来源：侯赛因·瓦基普瓦拉（Husain Vaghjipurwala）、张思玉、洪逸伦绘制

4.2.2 模型训练和测试结果

本案例模型的图像数据分类结果如图 8 所示，模型准确度如本节末尾表 1 所示。训练集准确率与验证集准确率分别达到 0.983 2 和 0.862 4，测试集中最后约 300 张图像的测试结果准确率达到约 88%。损失值则均在可接受的范围内。图 9 显示了部分测试结果，模型输出结果依然是以 [0,1] 为值域的概率，越靠近 0 代表越可能被判断为“活力”类别，越接近 1 则越可能被判断为“消极”类别。

4.2.3 模型改进和应用可能

该模型的一个重要改进方向是与地理信息及犯罪率数据结合，通过比较模型得出的街道活力判别结果和实际的犯罪率地图信息，来验证是否能根据街道的活力度来预测安全性。

图 9 部分活力属性立面测试结果（以东京街景为例）
资料来源：同图 8

① 计算机的配置为：处理器（2.40 GHz Intel Core i7）；显卡（NVIDA GeForce GT750 M）；内存（8.0 GB）。

此外，本案例设想通过定期更新输入活力判别模型的图像数据，来侦测街道的商业活力变化，例如建立城市各区域不同时期的街景图数据库，借助该模型判断城市哪些地段的活力在萎缩或增强（人口变得密集），从而实行空间改善措施或调整巡逻资源。更多拓展的应用可能性包括与城市空间三维数据相结合研究空间和活力的关系、侦测街道的业态变化等（图 10）。

4.3 城市肌理评估

城市用地分类对于城市区域规划、地产建设、商用许可和基础设施发展来说具有重要的参考价值[16,25]。城市的用地分类会反映为城市的肌理特点，而城市肌理则反映了生活在其中的人们所创造的区域文化环境特征，从而为下一步的城市空间结构规划提供数据依据。依此，本案例借助卫星图像建立判别模型尝试对城市肌理分类。

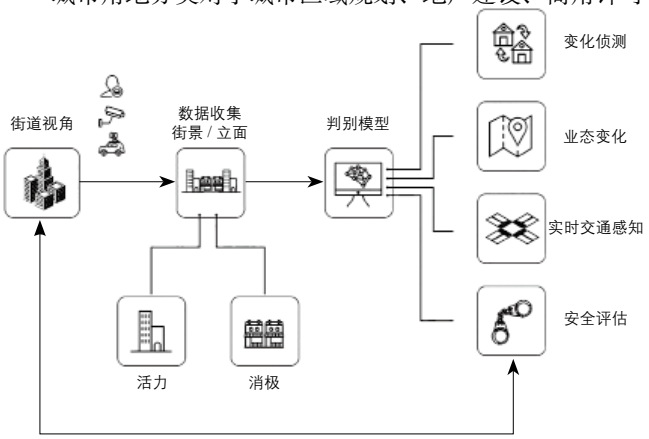

图 10 对图像判别模型在城市问题侦测应用的设想
资料来源：同图 8

4.3.1 模型设置

此实验以城市肌理中的六种特定要素（高密度片区、中等密度片区、低密度片区、公共建筑、交通、旧城区）作为判别类型，使该模型可从大尺度的城市卫星图中快速分辨出这六种用地类型并计算各类区域的占比。

模型的训练数据来源皆是百度的二维卫星图片（经过颜色处理）。选取的城市覆盖了国内一线城市（如：上海）到欠发达城市（如：台州）。模型的训练数据选取了范围尺寸为 1 km×1 km，数量共 2 000 张的上海卫星图像（图 11）。模型所需的训练数据在进行标签后以 75 ： 15 ： 10 的比例划分训练、验证和测试集①。

4.3.2 模型训练和测试结果

模型的训练结果如图 12 所示，准确度如表 1 所示，训练集准确率与验证准确率分别达到 0.952 3 和 0.757 1。结果显示了验证阶段的准确率偏低且损失值偏高，需要警惕模型

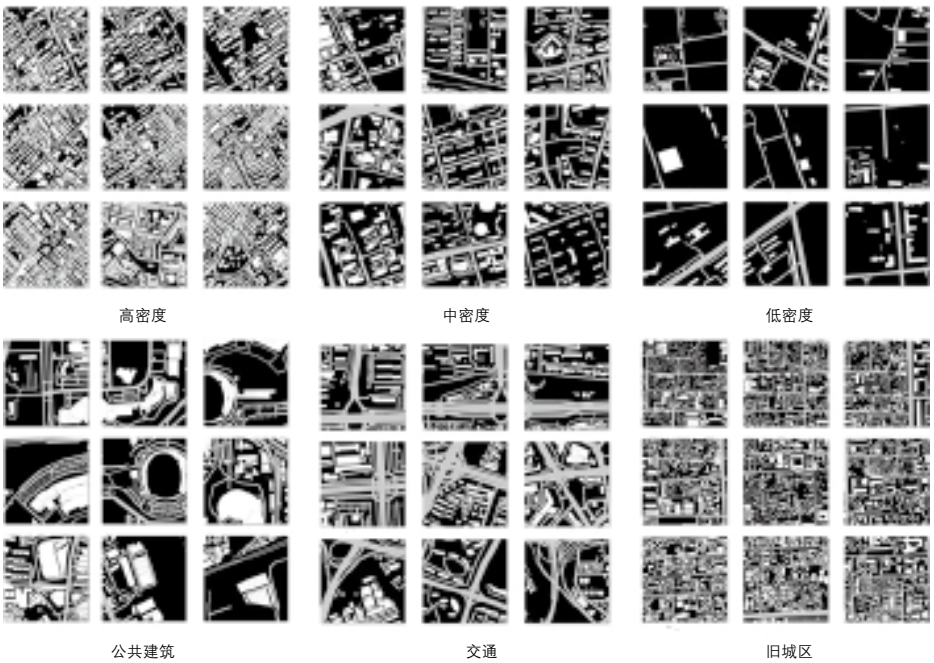

图 11 城市肌理第一阶段的分类示例
资料来源：贺斌、黄辰宇、刘鹏坤、彭茜、陈一宁绘制

① 计算机的配置为：处理器（4.20 GHz Intel Core i7）；显卡（NVIDA GTX 1070）；内存（16.0 GB）。

过拟合问题。解决方案是采用更高清、更大量的遥感图像进行训练，以修正模型偏差。

模型训练完成后，以上海某区域的卫星图作为测试对象输入模型，部分分析结果如图 13 所示。操作过程是首先将分析区域的卫星图按 1 km×1 km 范围（与训练集单张卫星图的尺寸相符）的网格划分，然后把划分好的图片数据接入判别模型进行肌理类型判别。输出结果以 [分类 A 的概率；分类 B 的概率；……分类 F 的概率] 的格式表达。其中概率以 [0,1] 为值域，各类别之后的概率越靠近 0 代表输入图片属于该分类的可能性越小，越接近 1 则表示该图片越可能属于该分类。然后根据分类结果计算该区域各类肌理元素的占比并可视化展现结果（图 13）。

4.3.3 模型改进和应用可能

该图像判别模型可结合城市规划设计的需求，增加更多的判别类型进行优化，例如识别商业片区、工业片区等。另一方面，卫星图划分尺寸的选择也有更多的可能性，除了规整的矩形或正方形网格划分，也可以考虑根据现有路网进行不规则地块形状划分（但可以预见后者的识别准确度会受影响并且模型训练难度也会增加）。

在设计或规划前期，场地分析是关键环节之一。当文本和数字数据（如城市规范或用地属性资料）缺乏时，很难对设计场地的周边区域做出快速准确的分析。而且区域评估分析并不仅是城市规划设计部门的需求，城市管理者、开发商等也需要高效精准的评估工具来辅助战略研究。然而大多数发展中国家（或城市）所面临的问题是没有经济技术基础开发专属的城市用地分析模型 [26]，因此利用现有的卫星图像数据进行评估是一种经济有效的方式。基于卫星图像的评估模型的建模成本低，可为建筑或城市设计师（或其他相关利益者）提供在可接受的准确率范围内的场地周边用地分析。同时，本模型也可发展成普适性的评估系统，供相对落后的地区使用。

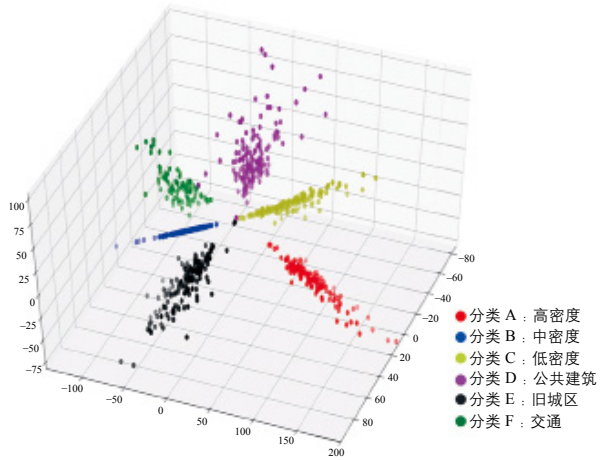

图 12 六种城市肌理的图像判别模型数据分布三维图
资料来源：同图 11

4.4 案例总结

从上述的全部案例中，基于 Keras 框架的图像判别模型展现出不俗的准确率，预示着自定义图像识别模型在分析评估城市方面具有较高的可行性。二元分类模型的训练时间平均不超过 2 分钟（样本数量平均为 1 500），而多元分类模型的训练则需花费 15 分钟左右。训练好后的模型对新图像输入的判别时间在毫秒级别，体现出其相比传统的人力分析具有极高的效率优势。

图 13 城市肌理判别模型测试结果（以上海部分区域为例）
资料来源：同图 11

表 1 各运用场景图像判别模型度量标准

课题	分类数目	训练准确率（acc）	训练损失值（loss）	验证准确率 (val_acc)	验证损失值 (val_loss)
城市风貌分析	2	0.924 8	0.192 1	0.852 0	0.353 2
城市问题侦测	2	0.983 2	0.283 9	0.862 4	0.275 3
城市肌理评估	6	0.952 3	0.184 1	0.757 1	0.622 5

5 总结

人工智能发展至今，机器从海量数据中自主“学习”规律模式的能力日趋完善。在大数据时代，城市数据的规模也与日俱增。同时，深度学习算法的发展允许计算机通过构建层级化的模型来学习复杂的抽象概念，使得很多无法通过人力抽取图像特征的问题由计算机自主学习解决。而日益成熟的 CNN 技术将发挥城市图像数据的巨大潜力，成为城市决策管理者在分析、侦测和评估方面的得力助手。

在应用方面，基于 CNN 的图像判别技术正在各行业垂直渗透，城市规划设计领域也不例外。通过数个课题的探索，本研究认为自定义的图像判别模型可在城市风貌分析、问题侦测和城市评估几个方面发挥可观的作用，并对当下城市的设计方法论有深刻的借鉴意义。进一步的探索除了关注图像识别技术在城市感知方面更多的应用可能性，也需要在技术方面进行提升，例如尝试不同类型的 CNN 框架，尝试其他深度学习的神经网络计算框架，优化标签工作，增加现有判别模型的丰富度和提升数据集的质量等。

值得注意的是，通常图像识别模型的准确率达到 80% 即可接受，大约 1 500 张训练图片可达到这个精准度。然而，若想获得类似人类的感知表现（在有监督学习的情况下），至少需要 1 000 万个被标注的样本数据用于模型训练 [7]18。城市问题的复杂性要求全面而细致的数据筛选与标注，因此训练数据库的建立成为需要突破的瓶颈。这将仰仗深度学习或其他感知技术的进一步突破，以实现用更小的数据集获得更高的性能表现。UPI

注：未注明资料来源的图表均为作者绘制。

参考文献

[1] 尼格尔·泰勒 . 1945 年后西方城市规划理论的流变 [M]. 李白玉，陈贞，译 . 北京：中国建筑工业出版社，2006.

[2] LUSK E, BOYLE J, WOS L, et al. Automated reasoning: introductions and applications[M]. New Jersey: Prentice Hall, 1984: 4.

[3] FLASIŃSKI M. Introduction to artificial intelligence[M]. Switzerland: Springer Nature, 2016: 5.

[4] 朱玮，王德 . 大尺度城市模型与城市规划 [J]. 城市规划，2003, 27(5): 47-54.

[5] 陈苏雅 . 5 月 1 日起深圳交警试点“刷脸”执法（附电子警察位置）[N/OL]. (2018-04-19)[2018-10-22]. http://www.sznews.com/news/content/2018-04/19/content_18921886.htm.

[6] ARIETTA S M, EFROS A A, RAMAMOORTHI R, et al. City forensics: using visual elements to predict non-visual city attributes[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 2624-2633.

[7] COURVILLE A, GOODFELLOW I, BENGIO Y. Deep learning[M]. Cambridge: MIT Press, 2016: 9.

[8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C] // PEREIRA F, BURGES C J C, BOTTOU L, et al. Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 1. New York: Curran Associates Inc., 2012: 1097-1105.

[9] GILL J K. Automatic log analysis using deep learning and AI for microservices[EB/OL]. (2017-07-21)[2018-10-20]. https://www.xenonstack.com/blog/data-science/log-analytics-deep-machine-learning-ai/.

[10] NG A. Machine learning yearning (draft version)[M/OL]. (2018-09-29)[2018-09-28]. https://www.deeplearning.ai/machine-learning-yearning/.

[11] PATEL S, PINGEL J. Introduction to deep learning: what are convolutional neural networks?[EB/OL]. (2017-04-24)[2018-09-28]. https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html.

[12] NAIK N, KOMINERS S D, RASKAR R, et al. Computer vision uncovers urban change predictors of physical urban change[J]. Proceedings of the National Academy of Sciences, 2017, 114(29): 7571-7576.

[13] UBA B K. Land use and land cover classification using deep learning techniques[D/OL]. Arizona: Arizona State University, 2016. [2018-11-20]. https://repository.asu.edu/attachments/170740/content/Uba_asu_0010N_15901.pdf.

[14] PAPADOMANOLAKI M, VAKALOPOULOU M, ZAGORUYKO S, et al. Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data[C]. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences Volume III-7. Göttingen: Copernicus Publications, 2016: 83–88.

[15] PENATTI O A B, NOGUEIRA K, SANTOS J A D. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops. New York: Curran Associates Inc., 2015: 44-51.

[16] ROMERO A, GATTA C, CAMPS-VALLS G. Unsupervised deep feature extraction for remote sensing image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(3): 1349-1362.

[17] SHEN Q, ZENG W, YE Y, et al. StreetVizor: visual exploration of human-scale urban forms based on street views[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(1): 1004-1013.

[18] YE Y, RICHARDS D, LU Y, et al. Measuring daily accessed street greenery: a human-scale approach for informing better urban planning practices[J]. Landscape and Urban Planning, 2018. https://doi.org/10.1016/j.landurbplan.2018.08.028.

[19] DUBEY A, NAIK N, PARIKH D, et al. Deep learning the city: quantifying urban perception at a global scale[C]. LEIBE B, MATAS J, SEBE N, et al, eds. Computer Vision – ECCV 2016. Cham: Springer, 2016: 196-212.

[20] NADAI M D, VIERIU R L, ZEN G, et al. Are safer looking neighborhoods more lively? a multimodal investigation into urban life[C] // HANJALIC A, SNOEK C. Proceedings of the ACM Multimedia Conference 2016. New York: ACM, 2016: 1127-1135.

[21] SERESINHE C I, PREIS T, MOAT H S. Using deep learning to quantify the beauty of outdoor places[J/OL]. London: Royal Society Open Science, 2017, 4(7). (2017-07-01)[2019-09-28]. https://royalsocietypublishing.org/doi/full/10.1098/rsos.170170.

[22] DOERSCH C, SINGH S, GUPTA A, et al. What makes Paris look like Paris?[J]. Communications. New York: ACM, 2015, 58(12): 103–110.

[23] SHALUNTS G, HAXHIMUSA Y, SABLATNIG R. Architectural style classification of building facade windows[C] // BEBIS G, BOYLE R, PARVIN B, et al. ISVC 2011: Advances in Visual Computing. Berlin: Springer, 2011: 288.

[24] ESLAMI S M A, REZENDE D J, BESSE F, et al. Neural scene representation and rendering[J]. Science, 2018, 360(6394): 1204–1210.

[25] SCOTT G J, ENGLAND M R, STARMS W A, et al. Training deep convolutional neural networks for land-cover classification of high-resolution imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(4): 549-553.

[26] JEAN N, BURKE M, XIE M, et al. Combining satellite imagery and machine learning to predict poverty[J]. Science, 2016, 353(6301): 790-794.

（本文编辑：张祎娴）