计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210500220

基于多模态多层级数据融合方法的城市功能识别研究

周新民[1,2]　胡宜桂[2]　刘文洁[2]　孙荣俊[2]
1 湖南工商大学新零售虚拟现实技术湖南省重点实验室　长沙 410205
2 湖南工商大学计算机与信息工程学院　长沙 410205

摘　要　城市功能区的划分与识别对分析城市功能区的分布现状和了解城市内部空间结构具有重要意义。这激发了多源地理空间数据融合的需求,特别是城市遥感数据与社会感知数据的融合。然而,如何有效实现城市遥感数据与社会感知数据的融合是一个技术难题。为了实现城市遥感数据与社会感知数据的融合,提高城市功能识别精度,以遥感图像和社会感知数据为例,引入多模态数据融合机制,提出了一种联合深度学习与集成学习的模型来推断城市区域功能。该模型分别利用 DenseNet 和 DPN 网络,从多源地理空间数据中提取城市遥感图像特征和社会感知特征,并进行特征级融合、决策级融合以及混合融合的多层级数据融合,对城市功能进行识别。所提模型在 URFC 数据集上得到了验证,其混合融合总体分类准确度、Kappa 系数和平均 F1 值 3 个评价指标值分别为 74.29%,0.67,71.92%。相比单模态数据的最佳分类方法,所提融合模型的 3 个评价指标值分别提高了 18.83%,0.24,35.46%。实验结果表明,该数据融合模型具有更好的分类性能,能有效融合遥感图像数据和社会感知数据,实现城市区域功能的精准识别。

关键词:城市功能区识别;多模态数据融合;深度学习;集成学习;社会感知
中图法分类号　TP391

Research on Urban Function Recognition Based on Multi-modal and Multi-level Data Fusion Method

ZHOU Xin-min[1,2],HU Yi-gui[2],LIU Wen-jie[2] and SUN Rong-jun[2]
1 Key Laboratory of Hunan Province for New Retail Virtual Reality Technology,Hunan University of Technology and Business,Changsha 410205,China
2 School of Computer and Information Engineering,Hunan University of Technology and Business,Changsha 410205,China

Abstract　The division and identification of urban functional areas is of great significance for analyzing the distribution status of urban functional areas and understanding the internal spatial structure of cities. This has stimulated the demand for multi-source geospatial data fusion,especially the fusion of urban remote sensing data and social sensing data. However,how to realize the fusion of urban remote sensing and social sensing data is a technical problem effectively. In order to realize the fusion of urban remote sensing and social sensing data and improve the accuracy of urban function recognition,taking remote sensing images and social sensing data as examples,introducing a multi-modal data fusion mechanism,and proposing a joint deep learning and ensemble learning model to infer urban regional functions. The model uses DenseNet and DPN network to extract urban remote sensing image features and social sensing features from multi-source geospatial data,and carries out multi-level data fusion of feature fusion,decision fusion and hybrid fusion to identify urban functions. The proposed model is verified on the URFC dataset,and these three evaluation index values of hybrid fusion overall classification accuracy,Kappa coefficient and average F1 are 74.29%,0.67,71.92%,respectively. Compared with the best classification method of single modal data,the three evaluation indexes of the proposed fusion model are increased by 18.83%,0.24,35.46% respectively. The experimental results show that the data fusion model has better classification performance,so that it can effectively fuse remote sensing image data and social sensing data,and realize the accurate identification of urban regional functions.

Keywords　Urban function recognition,Multi-modal data fusion,Deep learning,Ensemble learning,Social sensing

1　引言

城市发展一般要经历规划-发展调整-再规划的动态演变过程[1]。了解城市功能区的变化对于城市发展规划、自然资源配置和生态系统管理至关重要。对城市功能区进行及时划分,是检验智慧城市规划合理性以及指导未来城市建设的重要参考依据。城市功能区的识别大多基于单一的数据特征,如遥感图像数据。Cao 等[2]、Núez 等[3]、Rasheed 等[4]均利用高分辨率遥感图像来对城市功能相关场景进行识别。然而,遥感图像数据在进行城市功能识别时仅参照地表物理属性进行识别,并不能表征城市内部的功能属性,这使得在高分辨率下很难区分物理属性相同但功能属性不同的场景。为了挖掘

到稿日期:2021-05-30　返修日期:2021-07-27
基金项目:国家自然科学基金重大项目(72091515)
This work was supported by the Major Program of the National Natural Science Foundation of China(72091515).
通信作者:周新民(zhouxinmin2699@163.com)

更多城市内部功能属性，一些学者开始研究社会感知数据对城市功能识别的影响[5]。例如 Gao 等[6]、Xiao 等[7]利用车辆轨迹数据来识别城市区域功能，证实了公共交通数据对城市功能识别的作用。Yao 等[8]、Kang 等[9]利用 POI 数据进行城市功能识别，用社交媒体数据来描述城市功能。此外，手机数据也被有效用于城市功能分析，Jiang 等[10]、Jin 等[11]分别利用手机定位、手机信令数据对城市功能进行识别分析。这些研究表明，社会感知数据能有效表征城市的内部功能属性。

单一的遥感图像和社会感知数据能分别从城市功能的物理属性和社会属性进行功能识别。然而，在进行单一数据特征的识别时，其精度往往不高。为提升单一特征的识别精度，多特征数据融合方法应运而生。Hoffmann 等[12]利用街景图像和航空图像的融合来进行城市识别，Du 等[13]对土地 HSIS 图像和 LiDAR 图像进行特征提取，并构造多模态图进行识别分类。Xing 等[14]整合从众包数据中提取的景观指标和社会经济特征，来识别城市功能。Tu 等[15]、Qi 等[16]利用社交媒体数据和遥感图像数据融合来识别城市功能区的内部变化。而 Xu 等[17]提出一种结合建筑、景观、语义等度量方式的集成框架来绘制城市功能区画像以进行识别。以上研究均从多特征数据融合角度出发，提高了城市功能的识别精度。

由于遥感图像数据和社会感知数据存在模态差异，实现数据特征融合具有挑战性。已有研究大多采用深度学习的方法进行。如 Zhao 等[18]利用基于对象的卷积神经网络（OCNN）处理从 OSM 数据中派生出来的丰富语义元素，实现高分辨率遥感图像与 OSM 数据的集成。Bao 等[19]提出了一种深层特征卷积网络（DFCNN），用于处理 POI 数据和遥感图像数据，该方法结合了物理语义和社会功能语义，能够有效、快速、准确地识别城市功能区。Wang 等[20]提出了一种部分监督的跨域深度学习模型（CDCNN），用于从手机信号数据中识别城市居民的属性，保证了城市功能区识别的准确性。可见，基于深度学习的方法能够克服不同模态数据间的局限性，在融合遥感图像和社会感知数据方面具有很大的潜力。

当前，对于社会感知数据的研究大多停留在语义层面，对于城市功能的时间感知特征考虑较少。城市功能具有时空特性，加大对社会感知数据的时空特征的研究，具有提升城市功能分类性能的潜力。本文将更多地考虑社会感知数据的时间感知特征，结合遥感图像数据的空间特征推断城市功能。基于深度学习强大的特征表达能力以及集成学习的模型学习和泛化能力，从多模态数据融合角度出发，兼顾特征级融合、决策级融合以及混合融合等方式，将城市遥感图像数据和社会感知数据进行融合，以达到更加准确的城市功能识别分类效果。具体来说，分别利用 DenseNet 网络从城市功能遥感图像中提取场景空间特征，利用 DPN 网络从社会感知数据中提取场景时间感知特征，将两者特征进行多层级数据融合，进而实现城市功能区的识别分类。该模型充分利用了数据时空特性，通过多模态深度融合的方式来解决单一数据源可能存在的分类偏差问题，在公开数据集 URFC 上进行了相关对比实验。实验结果表明，本文提出的基于数据融合分类方法的准确性优于传统的单一特征分类方法。

2 多模态多层级数据融合的城市功能识别

2.1 基本框架

为了提高城市功能的识别精度，本文提出了一种基于多模态多层级数据融合的城市功能识别模型，以提高城市区域的功能识别精度。其模型框架如图 1 所示。该模型的关键在于建立一个联合嵌入空间，在该空间中遥感图像和社会感知数据可以结合在一起进行识别分类预测。该模型框架的识别分类步骤如下：

（1）输入遥感图像数据和社会感知数据，并进行数据选择与处理。

（2）对数据进行特征提取，采用深度学习方法 DenseNet 和 DPN 网络分别对遥感图像数据和社会感知数据进行全局和局部特征提取。

（3）将提取到的特征分别用于特征级融合模块和决策级融合模块，并将两模块的结果通过混合融合模块进行自适应地再融合，从而得到最终结果。在特征级融合模块，采取级联融合函数 Concat 对所提取的特征进行特征级融合，其融合结果输入全连接层和 Softmax 层进行分类，得到特征级融合结果；在决策级融合模块，将特征提取模块所提取的特征输入全连接层和 Softmax 层分别进行初次分类，其分类结果进行 Stacking 集成学习，得到决策级融合结果；在混合融合模块，将特征级、决策级融合结果进行 Stacking 集成学习，并输出最终的预测分类结果。

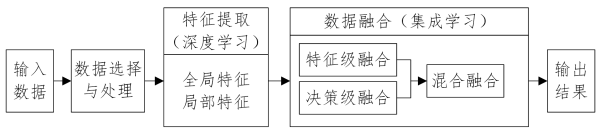

图 1 多模态多层级城市功能识别框架图

Fig. 1 Multi-modal and multi-level urban function recognition framework diagram

2.2 数据选择与处理

城市遥感图像和社会感知数据反映了城市区域功能内外部的直观特征和功能属性，选择这两种模态数据进行研究将有效把握城市功能的现状。为了验证模型的有效性，选择联合国教科文组织国际工程科技知识中心（IKCEST）举办的首届“一带一路”国际大数据公开竞赛数据集（URFC）进行相关实验[21]。该数据集共有 40 000 对带有真实场景标签的遥感图像和社会感知数据文件，记录了城市居住区（res.）、学校（sch.）、工业园区（ind.）、火车站（rail.）、飞机场（air.）、公园（park）、商业区（shop.）、政务区（adm.）、医院（hosp.）等九大功能区的遥感图像和社会感知数据。每个区域的遥感图像数据和社会感知数据分别记录在一个文件中，且文件名一一对应。例如文件名 000020_008.jpg，文件中包含 000020 区域的遥感图像，008 表示地区标注为政务区。对应文件名 000020_008.txt 记录了用户的社会感知数据，例如 c9927d3d5664a079,20181201&00|01|02|03|04|05,20181206&13|14，该记录表示用户 ID 为 c9927d3d5664a079 的用户在 2018 年 12 月 1 日的 0,1,2,3,4,5 点以及 2018 年 12 月 6 日的 13,14 点时访问过该地区。原始社会感知数据记录了 2018 年 10 月 1 日至 2019 年 3 月 31 日半年（182 天，26 周）内某区域用户的访问情况。时间精确到小时，即某个小时内的多次访问只记录一次。每个功能区的类别编号和数据量分布如表 1 所列，从表中可以看出数据集的样本分布并不均匀。

遥感图像数据为城市地区采集的影像数据，图 2 列举了城市九大功能区的遥感图像样本，图像尺寸大小均为 100 * 100 像素。从图中可看出图像质量较低，图像尺寸也限制了图像的可识别内容，各类图像之间的辨识度不高，这意味着难以对功能区图像进行准确划分。

表 1 URFC 数据集分布

Table 1 Categorical data distribution of URFC datasets

CategoryID	Functions of Areas	Num
001	res.	9542
002	sch.	7538
003	ind.	3590
004	rail.	1358
005	air.	3464
006	park	5507
007	shop.	3517
008	adm.	2617
009	hosp.	2867

图 2 城市九大功能区遥感图像样本

Fig. 2 Remote sensing image sample of nine urban functional areas

社会感知数据具体为用户访问某区域的日志性文件记录。用户在某区域通过移动设备访问本区域服务器，以服务器响应的准确时间标记本区域的用户访问行为，内容为用户到访功能区的身份 ID 和到访时间记录。需将原始数据重构为访问频度立方，使得重构后的数据能够被送入网络模型，完成特征提取及分类。本文利用用户访问的单变量时间签名数据，即每小时累计用户访问数的时间序列特征来进行实验，将用户访问数据聚合获得的每小时用户访问数据作为时间感知特征来衡量用户的活动强度。如图 3 所示，统计了用户访问数据九大功能区的周平均用户访问数，从图中可以看出不同类别的平均访问次数具有很大差异。学校、工业园区、政务区、医院在工作日(周一至周五)和周末的访问次数均有显著差异；公园和商业区在一周内用户访问次数差异不明显；火车站和机场访问情况基本相似，周末的访问数量较少。而在工作日，居住区表现出了明显的三峰值型；政务区和医院均为双峰值型；学校与工业园区类别呈现出了明显的单峰值型。以上统计结果与现实各功能区的实际情况基本相似，表明社会感知数据的时间序列特征在识别城市功能方面具有潜在价值。

为提高训练数据的多样性，减小模型泛化误差，将数据进行数据增强、数据标准化等操作，具体由以下几部分组成。

(1)数据清洗：城市遥感图像数据集含有部分无效图片(如全黑图)，需进行剔除。

(2)重采样：城市功能类别数量分布并不均衡，少量功能类别需进行重采样。

(3)数据标准化：读取城市遥感图像数据时需将图片尺寸大小统一设为[3,100,100]，并随机翻转图片角度，同时需将社会感知数据处理为维度[26,7,24]的张量，其中[x,y,z] 表示第 y 周、第 x 天、第 z 小时的到访人数。

(4)数据脱敏：社会感知数据需进行脱敏处理，将用户个人 ID 的真实信息隐藏，仅用一般字母、数字替代。

该数据集具有规模大、类别不均衡、图片辨识度低等特点，因此应用此类数据进行模型训练和测试更能体现其效果。

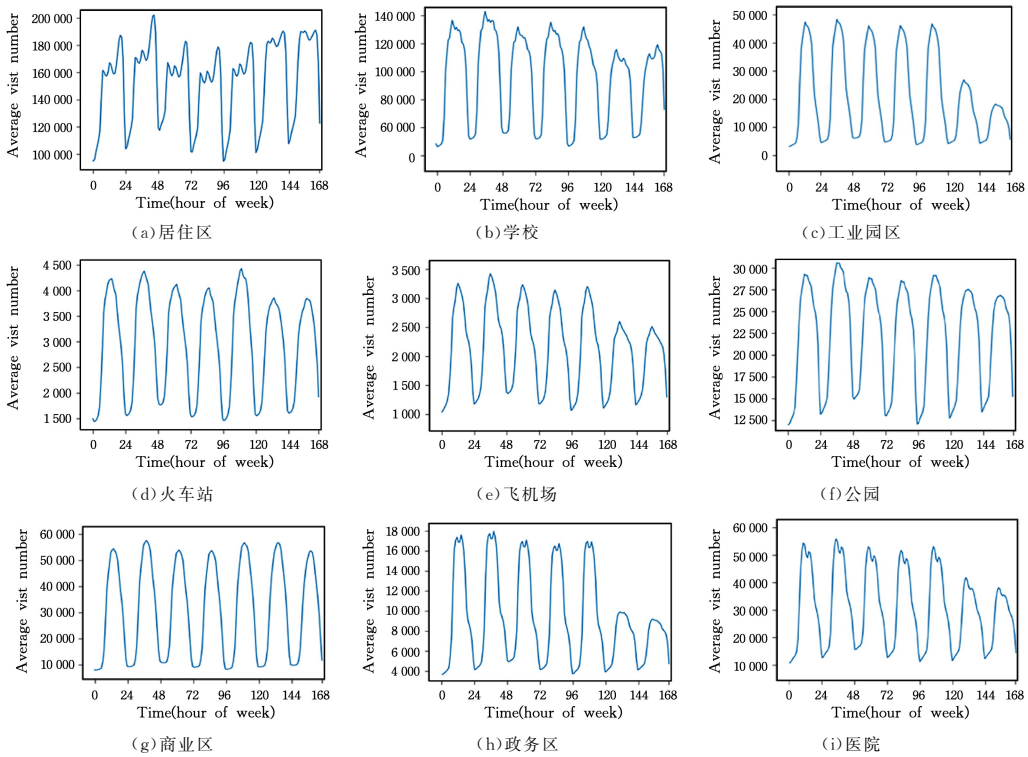

图 3 城市九大功能区每周平均访问次数汇总

Fig. 3 Summary of average number of visits per week in nine urban functional areas

2.3 特征提取

深度学习在处理时空大数据方面具有广阔的前景，特别是卷积神经网络(CNNs)，在提取时空序列数据局部特征、组合和抽象高层特征时具有一定优势[22]。选择 DenseNet 和 DPN 网络分别作为图像特征提取器和时间序列特征提取器，从遥感图像数据和社会感知数据中提取出与空间和时间相关的特征。

2.3.1 基于 DenseNet 的遥感图像特征提取

密集连接网络(DenseNet)通过加强 ResNet 网络中的跳接，能增强特征传递，减缓梯度消失，从而取得良好的特征利用效果[23]。遥感图像分辨率低，浅层次的特征提取难以准确把握遥感图像信息，DenseNet 网络能够加强对遥感图像特征的挖掘。同时，因其密集的连通性和深层次的结构，它能够有效避免模型产生的不良收敛、过拟合和梯度消失等问题[24]。DenseNet 为全连接形式，它使用密集连接路径将输入特征与输出特征进行级联。每个网络的输入特征是前面所有层输入特征的级联，当前层的输出特征又会被送到后面所有层作为输入，并进行全连接。其数学表达式如式(1)所示：

$$x_i = H_i([x_0, x_1, \cdots, x_{i-1}]) \tag{1}$$

其中，x_i 为第 i 层的输出；H_i 为非线性转换函数，该函数主要包含 BN，ReLU，Conv 层等组合操作；$[x_0, x_1, \cdots, x_{i-1}]$代表第 i 层前所有层的输出。

本文模型采用 DenseNet121 网络进行遥感图像特征提取，具体参数如表 2 所列，网络主要由 4 个密集连接模块和 3 个过渡层组成，各结构采用前馈全连接的方式进行连接。首先将遥感图像输入到 7 * 7 大小的卷积层，进行大尺度卷积，然后进行最大池化操作；输出特征被馈入 3 组含有密集连接模块(dense block)和过渡层(transition layer)的网络；再通过 7 * 7 大小的全局平均池化层，将输出特征映射进一步简化为特征向量；最后附加全连接和分类层来产生输出。这个输出结果将被用作最终提取的图像特征。

表 2 DenseNet 网络参数

Table 2 DenseNet network parameter

神经网络层	输出特征图尺寸	DenseNet121
卷积层	112×112	7×7 卷积
池化层	56×56	3×3 最大池化
Dense Block(1)	56×56	[1×1 卷积; 3×3 卷积]×6
Transition Layer(1)	56×56	1×1 卷积
	28×28	2×2 平均池化
Dense Block(2)	28×28	[1×1 卷积; 3×3 卷积]×12
Transition Layer(2)	28×28	1×1 卷积
	14×14	2×2 平均池化
Dense Block(3)	14×14	[1×1 卷积; 3×3 卷积]×24
Transition Layer(3)	14×14	1×1 卷积
	7×7	2×2 平均池化
Dense Block(4)	7×7	[1×1 卷积; 3×3 卷积]×16
分类层	1×1	7×7 平均池化
	—	softmax 分类

从表 2 可以看到，第一个 Dense Block 共包含 6 个[1 * 1，3 * 3]的卷积层，这个卷积层即为瓶颈结构，第二个到第四个 Dense Block 分别含有 12，24，16 个瓶颈结构，各瓶颈结构模块都串联在一起，紧密连接的 Dense Block 模块将更有利于特征提取。Transition Layer 模块包含有卷积和池化操作，主要用于对 Dense Block 模块传入的特征进行压缩降维。DenseNet 网络能够强化对遥感图像特征的复用，提高特征的识别精度，在解决图像处理过程中信息丢失的问题上取得了不错的效果[25]。

2.3.2 基于 DPN 的社会感知数据特征提取

双路径网络 DPN(Dual Path Network)结合了残差分组卷积和稠密连接两种思想，充分利用了残差连接和密集连接的优势进行互补，提高了特征提取的性能[26]。社会感知数据繁冗复杂，有效信息少，特征提取难度大，但 DPN 网络通过残差连接和密集连接双路径提升了对时空信息的提取能力，实现了旧特征的重用和新特征的探索，对于利用社会感知数据实现城市功能识别具有重大价值。DPN 网络数学形式如式(2)－式(5)所示：

$$x^k \triangleq \sum_{t=1}^{k-1} f_t^k(h^t) \tag{2}$$

$$y^k \triangleq \sum_{t=1}^{k-1} v_t(h^t) = y^{k-1} + \phi^{k-1}(y^{k-1}) \tag{3}$$

$$r^k \triangleq x^k + y^k \tag{4}$$

$$h^k = g^k(r^k) \tag{5}$$

其中，x^k 和 y^k 表示从单个路径第 k 步提取的信息；$f_t^k(\cdot)$指以隐藏状态为输入，输出提取信息的特征提取函数；$v_t(\cdot)$是作为 $f_t^k(\cdot)$的特征学习函数；$g^k(\cdot)$表示将收集到的信息转换为当前的隐藏状态。

式(2)为能够发现新特征的密集连接路径；式(3)为公共特征重用的剩余路径；式(4)定义了集成式(2)、式(3)的对偶路径，并将它们送入变换函数；式(5)利用最终的变换函数 h^k 的当前状态，进行下一步的预测。

本文利用 DPN92 网络对社会感知数据进行处理，提取相关时间序列特征。网络参数如表 3 所列，同 DenseNet 一样，DPN 通过堆叠多个模块来实现。

表 3 DPN92 网络参数

Table 3 DPN92 network parameter

神经网络层	输出特征图尺寸	DPN92
卷积层	112×112	7×7 卷积
池化层	56×56	3×3 最大池化
Conv1	56×56	[1×1 卷积; 3×3 卷积; 1×1 卷积]×3
Conv2	28×28	[1×1 卷积; 3×3 卷积; 1×1 卷积]×4
Conv3	14×14	[1×1 卷积; 3×3 卷积; 1×1 卷积]×20
Conv4	7×7	[1×1 卷积; 3×3 卷积; 1×1 卷积]×3
分类层	1×1	7×7 全局平均池化
	—	softmax 分类

首先进行大尺度卷积并进行池化操作，采用瓶颈结构的设计，分别经过 1 * 1，3 * 3，1 * 1 卷积层，最后一个 1 * 1 大小的卷积层的输出分为两部分，一部分以求和的形式添加到剩余路径中，另一部分与密集连接的路径进行连接。

2.4 数据融合

数据融合模块由特征级融合、决策级融合以及混合融合三大子模块组成。对特征提取模块所提取特征进行特征级与决策级的融合操作，将两个模块的结果进行混合融合，以提高决策分类精度。图 4 为多层级数据融合模块图。

在特征级融合模块，如图 4(a)所示，将遥感图像和社会感知数据利用深度学习方法进行特征提取后，在特征级上进行融合得到融合特征，并将融合后的特征输入到全连接层和 Softmax 层进行分类，得到特征级分类结果。

在决策级融合模块，如图 4(b)所示，先将特征提取模块提取特征输入到全连接层和 Softmax 层进行初次分类，再将分类结果输入到决策级进行决策级融合。决策级融合模块选择 Stacking 方式集成遥感图像和社会感知数据特征。

在混合融合模块，如图 4(c)所示，将特征级、决策级融合结果进行决策再融合。将特征级和决策级融合结果输入到 Lightgbm 中进行 Stacking 集成学习，以端到端的方式实现城市功能的准确识别。

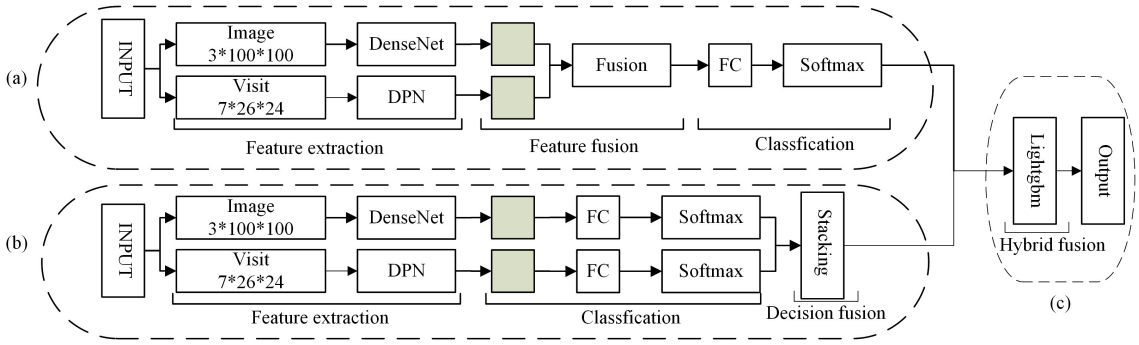

图 4 多层级数据融合模块图

Fig. 4 Multi-level data fusion module diagram

2.4.1 特征级融合

特征级融合是在特征层进行的，即从两个不同数据源中提取特征，在训练最终分类器之前进行融合。利用级联融合方法融合遥感图像和社会感知数据特征，能有效保留两种模态数据特征的优势，并进行充分融合[27]。图 5 为遥感图像数据与社会感知数据的特征级融合的具体实例分类流程图。

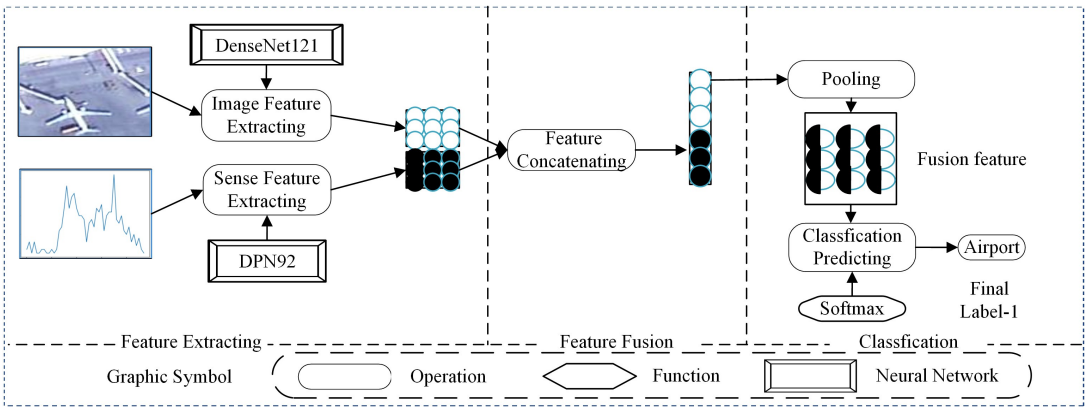

图 5 特征级融合实例识别

Fig. 5 Feature level fusion instance recognition

利用级联融合函数对特征提取阶段所提取出的图像特征和时间序列特征进行融合，将融合结果堆叠并池化，选择 Softmax 函数进行分类预测，得到特征级最终的分类标签 Final Lable-1。在级联融合函数中保留了两种模态特征图的结果，构建新的模型，将多种神经网络模型提取出的单一特征进行融合，从而形成新的融合特征。级联融合方法在神经网络 DenseNet 和 DPN 中分别提取出图像特征 f^i 和时间序列特征 f^t 进行特征拼接，即 $F=concate(f^i+f^t)$，融合后的特征 F 含有两种数据的完整特征信息，将其输入到全连接层和 Softmax 层，使网络得到更准确的分类结果。

2.4.2 决策级融合与混合融合

决策级融合模块利用遥感图像与社会感知数据两种模态的单独分类结果进行决策，实现从固定尺寸图片的静态匹配到社会感知时间序列的实时匹配。图 6 为遥感图像数据与社会感知数据的决策级融合具体实例分类流程图。先将特征提取模块所提取的特征分别输入到全连接层和分类层进行初次分类预测，将两种模态预测结果进行 Stacking 集成学习，并得出最后的预测结果 Final Lable-2。而在混合融合模块，将特征级融合分类结果与决策级融合分类结果输入到 Lightgbm 中进行 Stacking 集成学习并分类，得到最终的预测结果。

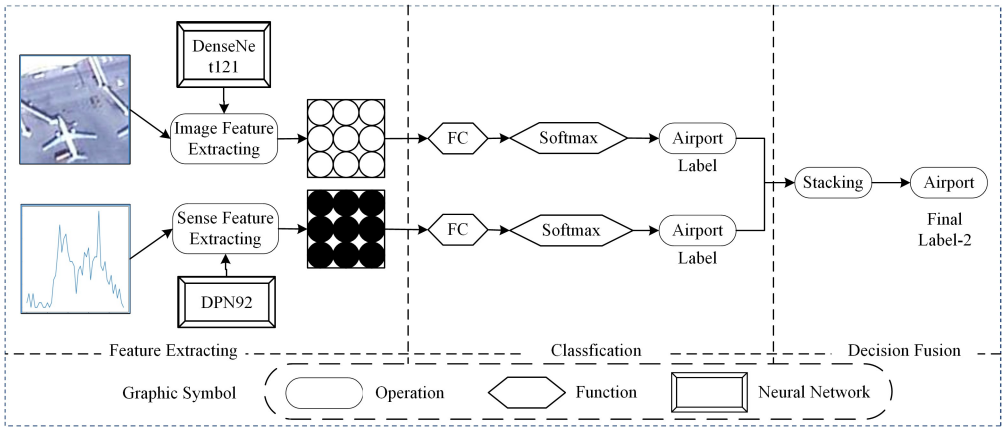

图 6 决策级融合实例识别

Fig. 6 Decision level fusion instance recognition

决策级融合与混合融合均采用 Stacking 方法进行，该方法是一种基于多级分类思想的集成学习方法，在处理不确定性预测、表示和组合方面有其独特的优势[28]。利用原始训练数据学习出低级别分类器即基学习器，并将基学习器预测结果作为新的特征，同原始训练数据一起训练高级别分类器即元学习器。Stacking 算法的流程如算法 1 所示。

算法 1 Stacking 算法

输入：数据集 $D=(x_i,y_i),i=1,2,\cdots,n$；

由若干分类器组成的基学习器 $H=\{H_1,H_2,\cdots,H_m\}$；

元学习器 L

输出：预测分类结果 $P,P=L(x'),(x',y)\in D_H$

步骤 1 利用数据集 D 训练基学习器 H：

for t=1 to m do：

Train H_t using D

end

步骤 2 利用数据集 D 和基学习器 H 生成新的数据集 D_H：

$D_H=\varnothing$

for i=1 to n do：

for t=1 to m do：

$z_{it}=H_t(x_i)$

end

$D_H=D_H\cup\{(z_{1i},z_{2i},\cdots,z_{Hi}),y_i\}$

end

步骤 3 利用生成的数据集 D_H 训练元学习器 L。

算法 1 的训练过程如下：假定一组数据集为 $D=\{(x_i,y_i),i=1,2,\cdots,n\}$，其中 x_i 为样本特征，y_i 为样本类别，n 为样本数量。当基学习器 H 对数据集 D 进行 k 折交叉验证时，对于原始数据的每个样本特征 x_i，分类器的预测结果为 z_{Hi}，将基分类器的 k 次测试结果取平均值与原始数据一起构成元学习器 L 的输入向量，即 $D_H=\{(y_i,z_{1i},z_{2i},\cdots,z_{Hi}),i=1,2,\cdots,n\}$，元学习器 L 通过学习最终输出样本分类属性。本文采用支持 GBDT 算法的 Lightgbm 框架作为元学习器，实验数据均采取五折交叉验证，其分类训练过程如图 7 所示。分类训练过程的主要步骤如下：

步骤 1 将数据集按照 3∶1 的比例划分为训练集和测试集，将训练集按照五折交叉验证方法随机分为 5 个子集（S_1，S_2，S_3，S_4，S_5），依次选取一个子集 $S_i(i=1,2,3,4,5)$ 作为验证子集，将其他 $S_{-i}=S_{\text{train}}-S_i$ 作为训练子集，进行五折交叉验证模型训练。

步骤 2 选定基学习器。选择神经网络 DenseNet 和 DPN 作为基学习器。S_{-i} 作为基学习器的训练集，将 S_i 作为测试集，输出测试结果 A_i，同时对测试集 S_{text} 进行预测，输出预测结果 B_i。

步骤 3 对步骤 2 循环 5 次得到训练集测试结果（A_1，A_2，A_3，A_4，A_5），将这 5 次结果纵向重叠合并得到 α_1，对测试集测试结果（B_1，B_2，B_3，B_4，B_5）取平均值得到 β_1。

步骤 4 在另一个学习器上执行以上步骤得到训练集产生的结果 α_2 和测试集产生的结果 β_2。

步骤 5 将 α_1，α_2 和原始数据集的标签 X 合并得到新样本训练集 $m=\{\alpha_1,\alpha_2,X\}$，将 β_1，β_2 合并得到新的测试数据集 $n=\{\beta_1,\beta_2\}$，将 m 作为元学习器 Lightgbm 的输入特征，将 n 作为元学习器的测试集来生成最终结果。

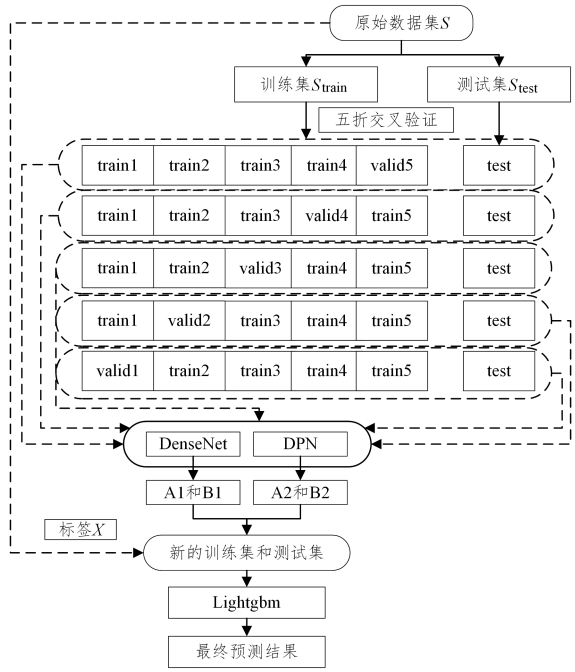

图 7 Stacking 框架下的城市功能识别分类

Fig. 7 Urban function recognition and classification using Stacking framework

3 实验结果与分析

本文实验所有网络都基于 Pytorch 框架实现，利用 sklearn 和 numpy 等机器学习库进行实验。实验设置训练集和测试集分别占原始数据集的 75%和 25%，且均采用五折交叉验证。使用 Adam 优化器对网络进行训练，学习速率为 0.0001，训练批大小为 128，最大训练迭代设置为 20 个 epoch。

3.1 评价指标

本文采用总体准确度(Accuracy)、Kappa 系数、平均 F1 值对模型结果进行评价，假设 $\boldsymbol{x}_{ij}$ 表示混淆矩阵，第 i 行、第 j 列的混淆矩阵，即第 i 类样本被预测为第 j 类样本数，n 为类别数量，N 为所有样本的总数。具体计算方式如式(6)－式(9)所示。

(1)Accuracy：

$$P_0=\sum_{i=1}^{n}x_{ii}/N \tag{6}$$

(2)Kappa 系数：

$$K=\frac{P_0-P_e}{1-P_e}, \tag{7}$$

其中，$P_e=\sum_{i=1}^{n}(\sum_{j=1}^{n}x_{i,j}\sum_{j=1}^{n}x_{j,i})/N^2$

(3)平均 F1 值：

$$F1_i=\frac{2P_iR_i}{P_i+R_i} \tag{8}$$

其中，P_i，R_i 为第 i 类的精确率和召回率，$P_i=x_{ii}/\sum_{j=1}^{n}x_{ij}$，$R_i=x_{ii}/\sum_{j=1}^{n}x_{ij}$。$F1_i$ 衡量某类的分类结果，而平均得分 $\overline{F1}$ 是不同类别所有 $F1$ 得分的平均值，可以衡量所有 n 个类别的总体分类结果：

$$\overline{F1}=\frac{1}{n}\sum_{i=1}^{n}F1_i \tag{9}$$

3.2 性能分析

为评估本文所提融合模型的分类性能，通过搭建相关实验环境，并使用训练集和测试集对模型分类能力进行评估。本文采用基于 DenseNet 和 DPN 网络的单模态数据分类方法和基于特征级融合、决策级融合以及混合融合的多模态数据融合的分类方法对城市功能区进行分类，各方法的训练损失和验证精度分别如图 8、图 9 所示。图 8 中，训练损失随着迭代次数不断增加而减少，并且在 20 个 epoch 后，训练损失开始小于 1。而在图 9 中，验证精度学习准确度随着训练次数增加而不断提升，并逐渐趋于稳定，这充分证明了本文模型的有效性。

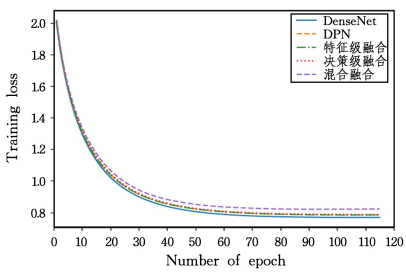

图 8 训练损失学习曲线图

Fig. 8 Training loss learning curve diagram

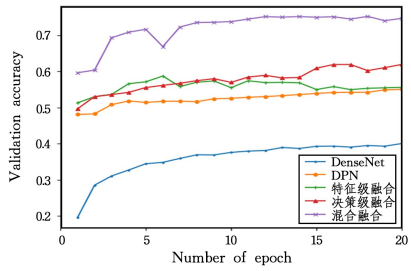

图 9 验证精度学习曲线图

Fig. 9 Validation of precision learning curve diagram

3.3 对比实验分析

为了验证数据融合模型在城市功能分区上的有效性以及融合分类效果，在 URFC 数据集上进行对比实验分析，从模型总体准确度(Accuracy)、Kappa 系数以及平均 F1 值 3 个指标进行评价。选择未经融合处理的单模态数据直接分类方法、多模态数据融合分类方法进行对比实验。将单模态分类的结果作为衡量各模型分类效果的基线，同时，对已有研究中采用相关机器学习和深度学习方法进行多模态数据融合分类的准确率进行对比分析，对比各单模态分类的结果与多模态数据融合分类的结果之间的差异。

3.3.1 单模态分类对比实验

单一图像分类方法采用传统神经网络 ResNet、VGG、Inception、Xception 以及本文所采用的 DenseNet 网络进行分类；单一社会感知数据时间序列特征分类方法采用 LSTM 网络以及本文所采用的 DPN 网络。实验性能如表 4 所列，在单一特征分类结果中，单独使用遥感图像数据进行分类的各分类方法的分类准确率均高于 38%。在传统神经网络中，ResNet 网络的分类准确率最高，为 39.48%，而本文提出的 DenseNet 网络在对遥感图像进行分类时的分类准确率最高，为 40.17%，高于传统的神经网络方法。对于仅使用时间序列特征的数据，传统时间序列网络 LSTM 和本文提出的 DPN 网络的测试精度均达到 54%以上，与使用单类遥感图像数据相比，分类结果均提高了 15%左右，本文采用的 DPN 网络也略高于 LSTM 网络。此外，单类遥感图像数据和社会感知数据的 Kappa 系数分别在 0.21 和 0.40 左右，这说明单类遥感图像数据在进行功能识别时，其模型预测结果和实际分类结果的一致性较低，而单类社会感知数据则保持中等一致性。在平均 F1 值上，社会感知数据比单类遥感图像数据高 14%左右，综合分类性能更佳。以上指标说明，时间序列特征在区域分类中比图像特征更容易区分，也进一步说明了社会感知数据在区域功能识别中的重要性。

此外，单独使用图像特征分类的结果明显低于单独使用时间序列特征的分类结果，可能的原因有两个：1)小尺度的遥感图像由于其空间覆盖范围有限，类内多样性和类间相似性高，只能提供有限的区域功能信息；2)遥感图像更多地只能反映地理属性，而时间序列特征数据更能直接反映人类的社会动态。这进一步说明，社会感知数据的时间序列特征有助于城市功能识别。同时，将遥感图像数据与社会感知数据进行融合，在提高城市功能识别性能上具有较大潜力。

表 4 单模态分类结果

Table 4 Result of single-mode classification

Data	Method	Validation			Testing		
		Accuracy/%	Kappa	Avg. F1/%	Accuracy/%	Kappa	Avg. F1/%
Image	ResNet	39.57	0.22	22.87	39.48	0.22	23.03
	VGG	38.85	0.21	22.46	38.68	0.21	22.32
	Inception	39.43	0.22	22.64	38.66	0.21	22.28
	Xception	39.02	0.21	21.73	38.42	0.21	21.61
	DenseNet	**40.61**	**0.23**	**23.40**	**40.17**	**0.23**	**23.22**
Social sensing	LSTM	54.63	0.41	36.21	54.06	0.40	36.08
	DPN	**55.05**	**0.42**	**36.35**	**55.46**	**0.43**	**36.46**

3.3.2 数据融合分类对比实验

本文采用的多模态数据融合方法分为特征级融合、决策级融合以及混合融合。从表 5 可以看出，本文提出的多模态数据融合方法均不同程度地提高了单一数据特征的分类精度。其中特征级融合分类准确率为 59.27%，决策级融合和混合融合分类准确率分别为 61.46%和 74.29%。混合融合模块相比特征级、决策级融合，其准确率提高超过 10%，比单一分类结果显著提高了 20%左右。使用数据融合方法后，特征级融合和决策级融合 Kappa 系数在 0.5 左右，说明其模型预测结果和实际分类结果保持中等一致。而混合融合 Kappa 系数接近 0.7，与其分类结果保持高度一致。而在平均 $F1$ 值指标上，特征级融合、决策级融合以及混合融合的 $F1$ 值分别达到了 46.69%，57.06%和 71.92%，数据融合模型的平均 $F1$ 值远高于单模态分类模型的平均 $F1$ 值，该模型具有良好的综合分类性能。混合融合分类效果优于特征级融合和决策级融合结果，这进一步说明了混合融合能够实现特征级与决策级融合之间的优势互补，从而可获得更好的分类效果。

表 5 多模态数据融合分类结果

Table 5 Results of multi-modal data fusion classification

Data	Method	Validation			Testing		
		Accuracy/%	Kappa	Avg. F1/%	Accuracy/%	Kappa	Avg. F1/%
Image+ Social sensing	特征级融合	58.72	0.48	46.82	59.27	0.49	46.69
	决策级融合	63.91	0.54	59.37	61.46	0.51	57.06
	混合融合	**75.16**	**0.68**	**72.74**	**74.29**	**0.67**	**71.92**

此外，本文对比了已有的相关研究[29]，已有研究采用 MLP，SVM，RF 等传统的机器学习分类方法进行城市功能识别的准确率分别为 52.75%，52.90%和 57.28%，采用深度学习方法中的 ResNet＋SPPNet，ResNet＋LSTM 分类方法的准确率分别为 56.36%和 60.59%，均比本文所采用的决策级融合(61.46%)和混合融合(74.29%)方法的准确率低，这说明本文所提的融合方法具有较好的分类效果。

特征级融合方法能够在训练过程中从两个特征中找到最大化的有用信息。决策级融合方法能从多个决策结果中进一步训练学习得到更加准确的分类结果。两种层级的融合方法各有其优势。与特征级融合相比，决策级融合更容易解释，因为在决策融合前可以提取单模态分类器的预测分数，从而直接检验不同输入数据的贡献。由于单模态数据本身只能提供相对有限的场景描述信息，因此无法取得满意的分类结果。而混合融合方法能够充分利用各特征之间的互补信息，获得足够的场景描述信息，从而较为明显地提升分类精度。总的来说，利用兼顾多模态多层级的数据融合分类方法进行城市功能识别均取得了不错的效果。

结束语 研究两种异构数据的有效融合，是探索城市遥感图像数据与社会感知数据的内在联系、把握社会感知数据对城市功能分区影响的关键。本文提出了一种新的多模态融合网络模型，该模型将遥感图像和社会感知数据进行融合，提升了城市功能识别精度。本文具有如下创新：从社会感知数据的时间感知特征角度出发，研究时间序列特征对城市功能识别的影响；利用深度学习网络 DenseNet 和 DPN 分别对两种数据进行特征提取，提高了数据的时空特征提取能力；提出了一种深度多模态融合模型，该模型兼顾特征级融合、决策级融合以及混合融合多层级数据融合方式，实现了多模态数据的有效融合。

本文展示了多模态数据融合在利用遥感图像数据和社会感知数据进行城市区域功能识别时的强大作用，为相关城市研究提供了有效途径。未来，我们计划考虑更多数据类型对于城市功能识别的影响，同时加大对多模态异构数据融合的关注，并进一步将本文方法应用到更多的现实场景中。

参 考 文 献

[1] EAGLE N，PENTLAND A S. Reality Mining：Sensing Complex Social Systems[J]. Personal and Ubiquitous Computing，2006，10(4)：255-268.

[2] CAO Y G，WANG Z P，YANG L. Research progress on road extraction methods from high-resolution remote sensing images [J]. Remote Sensing Technology and Application，2017，32(1)：20-26.

[3] NÚEZ J M，MEDINA S，VILA G，et al. High-Resolution Satellite Imagery Classification for Urban Form Detection[M]// Urban Form and Productivity in Mexico. NewYork：IntechOpen，2019：1-9.

[4] RASHEED S，ASGHAR M A，RAZZAQ S，et al. High-Resolution Remote Sensing Image Classification through Deep Neural Network[C]// 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2). IEEE，2021：1-6.

[5] YU L, XI L, SONG G, et al. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments[J]. Annals of the Association of American Geographers, 2015, 105(3): 512-530.

[6] GAO Q, FU J, YU Y, et al. Identification of urban regions' functions in Chengdu, China, based on vehicle trajectory data [J]. PLoS ONE, 2019, 14(4): e0215656.

[7] XIAO F, WANG Y, MEI Y N, et al. Urban functional area discovery method based on travel pattern subgraph[J]. Computer Science, 2018, 45(12): 268-278.

[8] YAO Y, LI X, LIU X, et al. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model[J]. International Journal of Geographical Information Science, 2016(4): 1-24.

[9] KANG X, PAN J J, ZHU Y X, et al. An urban core area identification method based on POI big data[J]. Remote Sensing Technology and Application, 2021, 36(1): 237-246.

[10] JIANG G L, HU F Y, SHI L X. Urban functional area identification based on call detailed record data[J]. Computer Applications, 2016, 36(7): 2046-2050.

[11] JIN P, CHEN M, SUN Z H. Research on the Recognition Method of Urban Land Function Area Based on Mobile Phone Signaling Data[J]. Information and Communication, 2018(1): 268-270.

[12] HOFFMANN E J, WANG Y, WERNER M, et al. Model fusion for building type classification from aerial and street view images [J]. Remote Sensing, 2019, 11(11): 1259.

[13] DU X, ZHENG X, LU X, et al. Multisource Remote Sensing Data Classification With Graph Fusion Network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021(99): 1-11.

[14] XING H, YUAN M. Integrating landscape metrics and socioeconomic features for urban functional region classification [J]. Computers Environment and Urban Systems, 2018, 72: S0198971518300462.

[15] TU W, HU Z, LI L, et al. Portraying urban functional zones by coupling remote sensing imagery and human sensing data[J]. Remote Sensing, 2018, 10(1): 141.

[16] QI L, LI J, WANG Y, et al. Urban observation: Integration of remote sensing and social media data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(11): 4252-4264.

[17] XU N, LUO J, WU T, et al. Identification and portrait of urban functional zones based on multisource heterogeneous data and ensemble learning[J]. Remote Sensing, 2021, 13(3): 373.

[18] ZHAO W, BO Y, CHEN J, et al. Exploring semantic elements for urban scene recognition: Deep integration of high-resolution imagery and OpenStreetMap (OSM)[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 151: 237-250.

[19] BAO H, MING D, GUO Y, et al. DFCNN-Based Semantic Recognition of Urban Functional Zones by Integrating Remote Sensing Data and POI Data[J]. Remote Sensing, 2020, 12(7): 1088.

[20] WANG J Y, HE X, WANG Z, et al. CD-CNN: a partially supervised cross-domain deep learning model for urban resident recognition[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2018: 192-199.

[21] 2019 The 5th Baidu & XJTU Big Data Contest The First IKCEST "The Belt and Road" International Big Data Contest [EB/OL]. (2019-04-25)[2021-07-14]. https://dianshi.bce.baidu.com/competition/30/data.

[22] TZIRAKIS P, CHEN J, ZAFEIRIOU S, et al. End-to-end multimodal affect recognition in real-world environments[J]. Information Fusion, 2021, 68: 46-53.

[23] HUANG G, LIU Z, VAN D, et al. Densely connected convolutional networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.

[24] CHEN B, ZHAO T, LIU J, et al. Multipath feature recalibration DenseNet for image classification[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(3): 651-660.

[25] WANG L F, WANG R F, LIN S Z, et al. Multimodal medical image fusion based on dual residual super-dense networks[J]. Computer Science, 2021, 48(2): 160-166.

[26] HUANG Z, LI W, LI J, et al. Dual-path attention network for single image super-resolution[J]. Expert Systems With Applications, 2021, 169(1): 114450.

[27] LIU X, WANG Z, WANG L. Multimodal Fusion for Image and Text Classification with Feature Selection and Dimension Reduction[C]// Journal of Physics: Conference Series. IOP Publishing, 2021: 012064.

[28] CHATZIMPARMPAS A, MARTINS R M, KUCHER K, et al. StackGenVis: Alignment of Data, Algorithms, and Models for Stacking Ensemble Learning Using Performance Metrics[J]. arXiv: 2005.01575, 2020.

[29] CAO R, TU W, YANG C, et al. Deep learning-based remote and social sensing data fusion for urban region function recognition [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 163: 82-97.

ZHOU Xin-min, born in 1977, Ph. D, professor, is a member of China Computer Federation. His main research interests include New Smart City and business intelligence and Big Data.