

MtSCCD: 面向深度学习的土地利用场景分类与变化检测数据集

周维勋^{1,2}, 刘京雷¹, 彭代锋¹, 管海燕¹, 邵振峰³

1. 南京信息工程大学 遥感与测绘工程学院, 南京 210044;
2. 北京师范大学 遥感科学国家重点实验室, 北京 100875;
3. 武汉大学 测绘遥感信息工程国家重点实验室, 武汉 430079

摘要: 利用遥感影像识别土地利用类型及监测其变化情况在城市规划和土地利用优化等领域发挥着重要作用。当前, 相关数据集存在样本量少、类别划分不合理、数据不开源等局限, 难以满足样本驱动的深度学习的遥感信息提取范式的需求。本文构建了一个面向深度学习的大规模场景分类与变化检测数据集 MtSCCD (Multi-temporal Scene Classification and Change Detection)。该数据集包括 MtSCCD_LUSC (MtSCCD Land Use Scene Classification) 和 MtSCCD_LUCD (MtSCCD Land Use Change Detection) 两个子数据集, 分别用于土地利用场景分类与变化检测任务。该数据集具有以下特点: (1) MtSCCD 是目前规模最大的公开的土地利用类型识别与检测数据集, 包含 10 种土地利用类型共 65548 幅图像, 并且样本覆盖中国 5 个城市的中心区域; (2) 由于 MtSCCD 数据集根据城市划分训练集、验证集以及测试集, 对于新增的城市土地利用数据, 可以根据需求划分为训练集与验证集或测试集, 因此可扩展性较高; (3) MtSCCD 数据集中测试集与训练集的样本来自不同的城市, 因此符合实际业务需求, 且能够验证模型的泛化性能。基于 MtSCCD_LUSC 和 MtSCCD_LUCD 两个子数据集, 本文评估了多个深度学习网络的分类与变化检测效果, 为后续的相关研究提供了参考。

关键词: 土地利用, 场景分类, 变化检测, 数据集, 信息提取, 特征提取, 深度学习, 卷积神经网络

中图分类号: TP79/P2

引用格式: 周维勋, 刘京雷, 彭代锋, 管海燕, 邵振峰. 2024. MtSCCD: 面向深度学习的土地利用场景分类与变化检测数据集. 遥感学报, 28(2): 321-333

Zhou W X, Liu J L, Peng D F, Guan H Y and Shao Z F. 2024. MtSCCD: Land-use scene classification and change-detection dataset for deep learning. National Remote Sensing Bulletin, 28(2): 321-333 [DOI: 10.11834/jrs.20243210]

1 引言

高空间分辨率遥感影像能够获取地物的纹理、结构等细节信息, 因此广泛应用于土地利用/土地覆盖相关任务, 其中土地利用场景分类与变化检测是广受遥感领域研究人员关注的研究热点。土地利用场景分类是从高分影像中提取图像的语义信息从而识别场景对应的土地利用类型 (白坤等, 2022; 钱晓亮等, 2018), 土地利用场景变化检测则是利用多时相的遥感影像监测土地利用类型的变化情况 (眭海刚等, 2018; 张良培和武辰, 2017),

二者对于城市发展规划和土地利用优化具有重要的指导意义 (袁静文等, 2020)。

土地利用场景分类与变化检测的关键在于获取表征能力强的图像特征。深度学习通过层次化的深层网络结构能够实现特征的自适应学习 (LeCun等, 2015), 因此面向自然图像的众多方法和模型被借鉴并用于遥感领域, 形成一种新的基于数据驱动的遥感信息提取范式 (Ma等, 2019; Zhang等, 2016; Zhu等, 2017)。对于一个有效的深度学习模型来说, 高质量、大规模的遥感图像标注样本是必不可少的 (冯权泷等, 2022)。在这

收稿日期: 2023-06-13; 预印本: 2023-08-29

基金项目: 国家自然科学基金 (编号: 42001285); 江苏省自然科学基金 (编号: BK20200813); 遥感科学国家重点实验室开放基金 (编号: OFSLRSS202215); 自然资源部国土卫星遥感应用重点实验室开放基金 (编号: KLSMNR-G202202)

第一作者简介: 周维勋, 研究方向为遥感图像智能解译。E-mail: zhouwx@nuist.edu.cn

MtSCCD 数据集: <https://sites.google.com/view/zhouxw/dataset>

一背景下,国内外学者发布了多个面向深度学习的遥感数据集,包括遥感图像军用飞机目标识别数据集 MAR20 (禹文奇等, 2022)、SAR 建筑数据集 SARBuD1.0 (吴樊等, 2022)、点云基准数据集 WHU-TLS 和 WHU-MLS (杨必胜等, 2021)。对于土地利用场景分类,当前有多个包含土地利用场景的数据集,如 UC Merced (Yang 和 Newsam, 2010)、WHU-RS19 (Xia 等, 2010)、RSSCN7 (Zou 等, 2015)、AID (Xia 等, 2017)、NWPU-RESISC45 (Cheng 等, 2017)、PatternNet (Zhou 等, 2018)、天宫一号高光光谱数据集 (刘康等, 2020)。但上述数据集仅包含少部分土地利用类别,且数据集侧重的是目标而非土地利用类型。例如, PatternNet 数据集中包含飞机、飞机跑道、桥梁、篮球场、足球场等目标,不能直接反应土地利用类型。作为土地利用类型变化的有效监测方法,当前的变化检测研究大多是像素级的 (Cheng 等, 2022)。然而,从土地利用的角度来说,场景内部一些地物的变化并不会直接导致场景类别发生变化。例如,一幅居住用地场景内部分房屋变成了裸地,虽然发生了像素级变化,但该场景仍然属于居住用地。因此,从标注成本和实际需求的角度来说,图像级的变化检测数据集更适合土地利用类型监测。现有的图像级土地利用变化检测数据集包括 MtS-WH (Wu 等, 2017, 2016) 和 WH-MAVS (Yuan 等, 2022)。其中, MtS-WH 数据集各时相包括 190 幅训练集影像、1920 幅测试集影像,由于样本数量较少,不适合基于深度学习的土地利用变化检测研究。与 MtS-WH 不同, WH-MAVS 数据集的样本数量更多,各时相包括 16496 幅训练集影像、4713 幅验证集影像和 2356 幅测试集影像,且包含的更多的土地利用类型。但 WH-MAVS 数据集存在以下几个局限:(1) 部分土地利用类型划分粒度过细。例如,一类、二类和三类居住用地的主要区别在于建筑物的高度或者密度,无论是前期样本构建还是后期的土地利用类型识别,三者都是比较容易混淆的,因此将 3 个类别合并为居住用地更为合理。(2) 数据集中的样本仅包括武汉市中心城区的土地利用场景,由于不同城市的土地利用规划与景观格局存在差异,因此构建覆盖更多城市的土地利用数据集有利于训练泛化性能更好的深度学习模型。(3) 数据集中训练集、验证集与测试集是按一定

比例随机划分的,这种划分方式不符合真实的业务需求,且不能评估模型的泛化性能。例如,实际业务应用中,往往是将训练好的深度学习模型用于和训练集不重合的区域,这样不仅可以实现模型的重复利用,而且可以验证模型的泛化性能。(4) 数据集不开源,不能用于模型训练和算法评估。

为了推动高分辨率土地利用场景分类与变化检测的研究进展,针对现有土地利用分类与变化检测数据集存在的局限性,本文利用高分辨率遥感影像构建了面向深度学习的大规模场景分类与变化检测数据集 MtSCCD (Multi-temporal Scene Classification and Change Detection)。该数据集具体包括 MtSCCD_LUSC (MtSCCD Land Use Scene Classification) 和 MtSCCD_LUCD (MtSCCD Land Use Change Detection) 两个子数据集,分别用于土地利用场景分类与变化检测研究。基于两个子数据集,本文进一步评估了多个场景分类与变化检测深度学习方法,为后续研究提供了参考基准。MtSCCD 的两个子数据集 MtSCCD_LUSC 和 MtSCCD_LUCD 分别与现有场景分类和变化检测数据集的对比如表 1 所示。可以看出,对于土地利用场景分类和变化检测两个任务来说, MtSCCD 数据集在图像分辨率、样本数量、数据组织方式、开放获取等方面具有明显的优势。

2 MtSCCD 数据集

2.1 数据来源与标注

为了保证 MtSCCD 数据集中土地利用场景图像的多样性以便更好地评估土地利用场景分类与变化检测算法,本文选择杭州、合肥、南京、上海和武汉 5 个城市中心区域的高分影像作为数据源,并且每个城市的数据均包含同区域两个时相的影像。高分影像来自 World Imagery (<https://livingatlas.arcgis.com/wayback>[2023-06-13]), 空间分辨率大约为 1 m, 包括 R、G、B 这 3 个波段。

对大尺寸高分辨率影像按照 300×300 像素大小进行裁剪得到土地利用场景图像,图像的命名格式为: xx_yyyymm_nx_ny_c。其中, xx 表示城市的首字母缩写, yyyymm 表示影像获取的年份和月份, nx 和 ny 表示图像裁切时在原影像中的位置编号, c 表示场景类别。

表1 MtSCCD数据集与现有的场景分类与变化检测数据集对比
Table 1 Comparison between MtSCCD and the existing datasets

	数据集	分辨率/m	图像尺寸	类别数目	图像数目	数据集是否划分	是否开源
场景分类	UC Merced(Yang和Newsam,2010)	0.3	256×256	21	2100	否	是
	WH-RS19(Xia等,2010)	最高0.5	600×600	19	1005	否	是
	RSSCN7(Zou等,2015)	未知	400×400	7	2800	否	是
	AID(Xia等,2017)	8—0.5	600×600	30	10000	否	是
	NWPU-RESISC45(Cheng等,2017)	30—0.2	256×256	45	31500	否	是
	PatternNet(Zhou等,2018)	4.69—0.06	256×256	38	30400	否	是
	RSI-CB(Li等,2020)	3—0.22	256×256	35	24000	否	是
	MtSCCD_LUSC	1	300×300	10	65548	是	是
变化检测	MtS-WH(Wu等,2016,2017)	1	150×150	8	4220	是	是
	WH-MAVS(Yuan等,2022)	1.2	200×200	14	47134	是	否
	MtSCCD_LUCD	1	300×300	10	65548	是	是

本文参考城市用地分类与规划建设用地标准(GB50137-2011)([http://www.risn.org.cn/Xxbz/ShowForceStandard.aspx?Guid=61387\[2023-06-13\]](http://www.risn.org.cn/Xxbz/ShowForceStandard.aspx?Guid=61387[2023-06-13]))和现有公开的城市土地利用数据(Wu等,2016;Yuan等,2022),确定了MtSCCD数据集中土地利用场景的分类体系,具体包括居住用地(residential land)、公商用地(public service and commercial land)、教育用地(educational land)、工业用地(industrial land)、交通用地(transportation land)、农业用地(agricultural land)、水体(water body)、绿地(green space)、林地(woodland)、裸地(bare land)共10种类别。杭州、合肥、南京、上海和武汉5个城市的土地利用场景经目视解译后分别划分到上述10个土地利用类别。为了保证数据集质量,人工标注

过程中丢弃成像质量不高的场景图像(如不清晰、云雾遮挡等),并对难以确定类别的土地利用场景借助Google Earth影像进行解译。

MtSCCD数据集中5个城市两个时相各类别场景的数目、标签(类别编号)、影像的获取时间如表2所示。由表2可知,各城市两个时相的影像的获取月份比较接近,这样可以充分降低季节因素对地物的影响,便于后续构建土地利用场景分类与变化检测数据集。但由于裁剪获取土地利用场景样本时去掉了样本的坐标信息,因此无法绘制各城市的采样点分布图,导致不能展现采样时的空间分异性以及评估样本空间分布的合理性,这是MtSCCD数据集的一个缺陷。

表2 MtSCCD数据集各城市土地利用数据概况

Table 2 Overview of the land use data of each city in the MtSCCD dataset

类别	标签	杭州		上海		武汉		南京		合肥	
		时相1	时相2	时相1	时相2	时相1	时相2	时相1	时相2	时相1	时相2
居住用地	1	1103	1227	2447	2517	1019	1292	1187	1339	855	1017
公商用地	2	134	176	354	417	163	256	65	79	248	274
教育用地	3	203	249	275	303	207	247	206	247	204	293
工业用地	4	355	260	613	488	342	405	500	436	396	411
交通用地	5	1892	2079	1447	1455	923	1269	758	910	1357	1466
农业用地	6	2484	2108	45	22	987	346	489	176	1016	593
水体	7	629	636	352	357	2161	2067	521	493	412	402
绿地	8	414	394	453	438	619	732	373	398	250	334
林地	9	2003	1977	9	9	199	246	814	794	76	141
裸地	10	114	225	206	195	482	242	149	190	264	147
图像数目		9331	9331	6201	6201	7102	7102	5062	5062	5078	5078
影像获取时间		2013-10	2018-10	2010-07	2019-11	2013-07	2021-09	2013-05	2019-03	2014-12	2021-10

2.2 数据集的构建

2.2.1 MtSCCD_LUSC子数据集

由表2可知,各城市同一类别的土地利用场景均包括两个时相,而两个时相的同类别场景必然是有差异的。因此,为了增加每个土地利用类别的图像数目且使图像库更具挑战性,将两个时相的数据合并构建MtSCCD_LUSC子数据集。首先,把每个城市两个时相的数据按照类别进行合并得到单时相场景数据;然后,分别将杭州、上海、武汉3个城市各类别数据按照80%和20%的比例划分为训练集与验证集;最后,将南京与合肥两个城市的数据作为两个测试集,其中,测试集A和测试集B分别为合肥和南京的土地利用数据。

表3给出了MtSCCD_LUSC数据集中训练集、验证集以及测试集的划分情况,可以看到,训练集和验证集分别包含36215和9053幅图像,测试集A和B分别包含10156和10124幅图像。同时,从图1给出的各土地利用类别的样本图像可以看出

MtSCCD_LUSC数据集类内差异较大(如交通用地、公商用地)。因此,MtSCCD_LUSC是一个大规模且具有挑战性的土地利用场景分类图像库,可用于基于深度学习的场景分类方法研究。

表3 MtSCCD_LUSC数据集划分

Table 3 Data splits in MtSCCD_LUSC dataset

类别	标签	训练集	验证集	测试集	
				A	B
居住用地	1	7684	1921	1872	2526
公商用地	2	1200	300	522	144
教育用地	3	1187	297	497	453
工业用地	4	1971	492	807	936
交通用地	5	7253	1812	2823	1668
农业用地	6	4794	1198	1609	665
水体	7	4961	1241	814	1014
绿地	8	2440	610	584	771
林地	9	3554	889	217	1608
裸地	10	1171	293	411	339
图像数目		36215	9053	10156	10124
MtSCCD_LUSC 数据集图像数目			65548		



(a) 居住用地
(a) Residential land



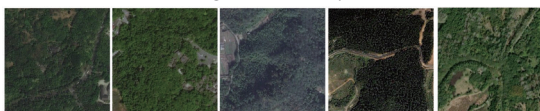
(c) 教育用地
(c) Educational land



(e) 交通用地
(e) Transportation land



(g) 水体
(g) Water body



(i) 林地
(i) Woodland



(b) 公商用地
(b) Public service and commercial land



(d) 工业用地
(d) Industrial land



(f) 农业用地
(f) Agricultural land



(h) 绿地
(h) Green space



(j) 裸地
(j) Bare land

图1 MtSCCD_LUSC数据集各类别图像实例

Fig. 1 Example images of each category in the MtSCCD_LUSC dataset

2.2.2 MtSCCD_LUCD子数据集

MtSCCD数据集中每个城市均包含两个时相的土地利用场景数据, 因此可以利用各城市的数据直接构建MtSCCD_LUCD子数据集。首先, 分别将5个城市时相1和时相2的各类别数据合并在一起, 得到各城市时相1和时相2的数据; 然后, 按照80%和20%的比例依次将杭州、上海、武汉3个城市的时相1与时相2的数据随机划分为训练集与验证集, 划分时需保证训练集与验证集中同一地面区域两个时相的场景图像同时划分到训练集或验

证集; 最后, 将南京与合肥两个城市的数据作为测试集, 其中, 测试集A和测试集B分别为合肥和南京的土地利用数据。MtSCCD_LUCD是图像级的变化检测数据集, 支持二值变化(即场景类别是否发生变化)和类别变化(即场景的变化类别)两种检测任务。表4给出了MtSCCD_LUCD数据集中训练集、验证集以及测试集的划分情况, 可以看到, 训练集和验证集分别包含18108和4526幅图像对, 测试集A和B分别包含5078和5062幅图像对, 各类别变化与未变化图像实例如图2所示。

表4 MtSCCD_LUCD数据集中训练集、验证集与测试集的划分

Table 4 Data splits of training set, validation set and testing set in MtSCCD_LUCD dataset

类别	标签	训练集		验证集		测试集			
		时相1	时相2	时相1	时相2	A		B	
						时相1	时相2	时相1	时相2
居住用地	1	3696	4073	873	963	855	1017	1187	1339
公商用地	2	513	671	138	178	248	274	65	79
教育用地	3	551	644	134	155	204	293	206	247
工业用地	4	1058	925	252	228	396	411	500	436
交通用地	5	3402	3825	860	978	1357	1466	758	910
农业用地	6	2788	1968	728	508	1016	593	489	176
水体	7	2512	2446	630	614	412	402	521	493
绿地	8	1193	1243	293	321	250	334	373	398
林地	9	1754	1766	457	466	76	141	814	794
裸地	10	641	547	161	115	264	147	149	190
图像数目		18108	18108	4526	4526	5078	5078	5062	5062
MtSCCD_LUCD 数据集图像数目		65548							



图2 MtSCCD_LUCD数据集各类别变化与未变化图像实例

Fig. 2 Changed and unchanged example images of each category in the MtSCCD_LUCD dataset

此外, 由表5给出的土地利用场景变化矩阵可知, MtSCCD_LUCD数据集包含丰富的土地利用变化, 共有14.52%的场景发生了变化。具体来说, 时相1中4.27%的居住用地发生了变化, 6.81%的交通用地发生了变化, 18.68%的绿地发生了变化。此外, 工业用地、农业用地、绿地和裸地的变化

比例较高, 分别为20.04%、36.77%、18.68%和79.26%。因此, MtSCCD_LUCD是一个大规模的土地利用场景变化检测图像库, 包含丰富的土地利用场景变化信息, 可用于基于深度学习的场景变化检测方法研究。

表5 MtSCCD_LUCD数据集土地利用场景变化矩阵
Table 5 Land use scene change matrix of MtSCCD_LUCD dataset

类别	时相2										图像数目	变化比例/%	
	1	2	3	4	5	6	7	8	9	10			
时相1	1	6329	7	30	11	81	6	2	11	1	133	6611	4.2
	2	5	940	0	1	13	0	0	2	0	3	964	2.49
	3	9	2	1076	0	4	0	1	0	0	3	1095	1.74
	4	59	6	20	1764	115	5	0	24	2	211	2206	20.04
	5	191	60	65	38	5943	11	6	45	8	10	6377	6.81
	6	273	57	59	84	626	3175	63	312	106	266	5021	36.77
	7	15	3	2	7	87	18	3812	66	4	61	4075	6.45
	8	108	25	25	21	107	2	40	1715	25	41	2109	18.68
	9	16	0	5	7	18	12	8	7	3009	19	3101	2.97
	10	387	102	57	67	185	16	23	114	12	252	1215	79.26
图像数目	7392	1202	1339	2000	7179	3245	3955	2296	3167	999	32774	14.52	

2.3 数据集的特点

MtSCCD数据集是一个面向深度学习的高分辨率遥感土地利用场景数据集, 支持场景分类与变化检测研究, 具有以下特点: (1) 图像数量大规模: MtSCCD是目前公开的规模最大的高分辨率土地利用场景分类与变化检测数据集, MtSCCD_LUSC和MtSCCD_LUCD两个子数据集均包含65548幅图像和10种土地利用类型。(2) 高度可扩展性: MtSCCD数据集是根据城市分别划分训练集、验证集以及测试集, 因此具有较高的可扩展性。后续对于新增的城市土地利用数据, 可以按照一定比例划分到训练集与验证集, 或直接作为测试集, 实现数据集的灵活扩充。(3) 符合实际应用场景: MtSCCD_LUSC和MtSCCD_LUCD两个子数据集中训练集和验证集与测试集不重合, 即模型训练和测试所用的数据来自不同区域, 因此更符合实际应用场景。此外, 这种不重合的数据集划分方式也有利于验证模型的泛化性能。(4) 场景类内差异大: MtSCCD数据集包含5个城市的土地利用数据, 而不同城市的同类别土地利用场景受成像条件等因素的影响存在较大的视觉差异。同时, MtSCCD数据集构建分类体系时, 对于相近的类别进行了合并, 进一步增加了场景的类内差异性。

例如一类、二类、三类居住用地统一划分到居住用地, 道路、铁路、桥梁统一划分到交通用地。类内差异大使得MtSCCD数据集对训练的模型更具挑战性。

3 MtSCCD数据实验与分析

本节以MtSCCD数据集为基础, 对基于深度学习的土地利用场景分类与变化检测方法进行评估。

3.1 评价指标

为了评价土地利用场景分类与变化检测方法的精度, 本文基于混淆矩阵, 采用总体精度OA (Overall Accuracy) 和Kappa系数作为评价指标。其中, 对于场景变化检测, 由于本文研究的是二值变化检测 (即土地利用类型发生变化和未发生变化两类), 因此以变化场景为正类、未变化场景为负类, 采用二分类混淆矩阵来计算总体精度和Kappa系数。

3.2 实验设置

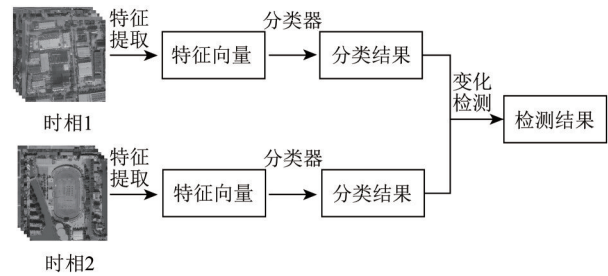
对于土地利用场景分类任务, 本文选择常用的网络, 包括AlexNet (Krizhevsky等, 2012)、VGG网络 (VGG16和VGG19) (Simonyan和Zisserman, 2015)、GoogLeNet (Szegedy等, 2015)、ResNet系列

网络 (ResNet18、ResNet50、ResNet101) (He 等, 2016)、DenseNet (Huang 等, 2017)、EfficientNet (Tan 和 Le, 2019)、SENet (Hu 等, 2018)、ViT (Vision Transformer) (Dosovitskiy 等, 2021) 和 SwinT (Swin Transformer) (Liu 等, 2021) 进行分类。其中, AlexNet 和 VGG 网络提取第一个全连接层的输出作为特征, GoogLeNet 和 ResNet 网络提取最后一个池化层的输出作为特征, 训练 SVM 分类器进行分类, 其余网络直接用自带的分类器进行分类。具体训练时, 以预训练网络为基础进行训练。其中, AlexNet、VGG、GoogLeNet、ResNet 的学习率设置为 $8E-5$, 批次大小 50, 优化器 sgd。DenseNet 和 EfficientNet 的学习率为 $1E-3$, 批次大小分别为 50 和 128, 优化器 sgd, 预训练网络分别采用 densenet201 和 efficientnetb0。SENet、ViT 和 SwinT 的学习率为 $1E-4$, 批次大小 32, 优化器 sgd (SwinT 采用 AdamW), 采用的预训练网络分别为 se_resnet50、vit_base_patch16_224_in21k 和 swin_tiny_patch4_window7_224。

对于土地利用场景变化检测任务, 本文选择两种变化检测方法, 一是常规的基于分类的“先分类后检测”方法 CDC (Change Detection after Classification), 二是基于场景相似度的变化检测方法, 包括 CNN 特征相似性度量方法 CFSM (CNN Feature Similarity Measure) 和基于相似性学习的变化检测方法 SSCD (黄宇鸿和周维勋, 2022)。图 3 给出了两种变化检测方法的基本流程, 可以看出利用场景对相似性进行变化检测是一种更为简单的方法, 能够避免基于分类的方法中的两次分类过程。

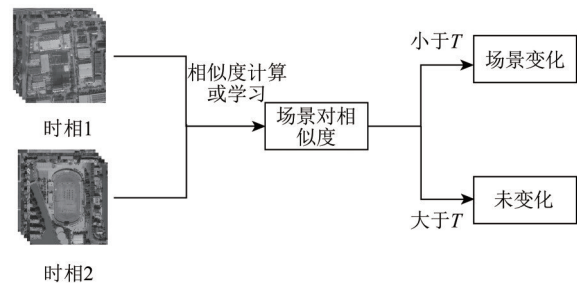
CDC 方法包括两个: 一是利用场景分类任务中训练的 AlexNet、VGG19、GoogLeNet 和 ResNet101 提取时相 1 和时相 2 土地利用场景的特征训练 SVM 进行变化检测, 或利用训练的 DenseNet、EfficientNet、SENet、ViT、SwinT 直接分类进行变化检测, 记为 CDC_1; 二是参考手工特征与视觉词袋模型 BoVW (Bag of Visual Words) 结合的思路 (Wu 等, 2016), 利用上述 4 个网络提取最后一个卷积层的特征, 并采用 BoVW 对卷积层特征进行编码得到特征向量用于训练 SVM 进行变化检测, 记为 CDC_2, 其中字典大小设置为 128。CFSM 方法同 CDC_1 方法, 不同之处在于提取特征后通过欧氏距离计算两个时相场景的相似度进行变化检测。其中, DenseNet、EfficientNet、SENet、ViT、SwinT 网络从分类层的

前一层提取特征计算相似度。SSCD 方法也是利用场景相似度进行变化检测, 但与 CFSM 方法不同, SSCD 是通过网络直接学习场景的相似度。此外, 对于 CFSM 和 SSCD 两种相似性变化检测方法, 本文采用相似度阈值搜索方法 (黄宇鸿和周维勋, 2022) 获取相似度阈值, 且相似度阈值通过验证集获取。



(a) 基于分类的变化检测基本流程

(a) Flowchart of classification-based change detection



(b) 基于相似性的变化检测基本流程

(b) Flowchart of similarity-based change detection

图 3 基于分类和相似性的变化检测基本流程

Fig. 3 Flowchart of classification-based and similarity-based change detection

3.3 MtSCCD_LUSC 数据集实验结果

表 6 给出了各网络对 MtSCCD_LUSC 的测试集 A 和 B 的分类结果, 可以看出网络层较浅的 AlexNet 在测试集 A 和 B 上的效果最差, 但简单增加网络深度并不能明显提升分类精度, 例如 ResNet18、ResNet50 和 ResNet101 分类性能基本一致, 且和 VGG16、VGG19、GoogLeNet 相比, 也并没有表现出明显优势。ResNet 之后提出的新型网络结构, 如 DenseNet 和 EfficientNet, 取得了更好的分类结果, 尤其是 DenseNet, 在两个测试集上的精度最高。此外, SENet、ViT 和 SwinT 等 3 个网络均在特征提取时考虑了注意力, 整体上分类效果比传统网络更好。若采用更优的预训练网络, 有望进一步提升分类精度。

表6 MtSCCD_LUSC数据集不同网络分类结果

Table 6 Scene classification results of different networks on MtSCCD_LUSC dataset

方法	总体精度/%	
	测试集 A	测试集 B
AlexNet	63.17	62.37
VGG16	72.41	69.77
VGG19	74.03	70.17
GoogLeNet	70.09	69.68
ResNet18	70.37	70.27
ResNet50	71.69	71.47
ResNet101	71.96	71.93
DenseNet	76.95	75.99
EfficientNet	73.20	73.52
SeNet	74.76	74.25
ViT	71.81	69.59
SwinT	76.74	72.61

注：表中加粗数值表示最好结果。

为了进一步分析MtSCCD_LUSC数据集的分类结果，选择AlexNet、VGG19、GoogLeNet、ResNet101、DenseNet、EfficientNet、SENet、SwinT绘制了在测试集A和B上分类结果的混淆矩阵（行表示真实结果，列表示分类结果，方格颜色越深表示数值越大），分别如图4和图5所示。图4、图5中AL表示农业用地、BL表示裸地、EL表示教育用地、GS表示绿地、IL表示工业用地、PSCL表示公商用地、RL表示居住用地、TL表示交通用地、WB表示水体、WL表示林地。由混淆矩阵可知，公商用地与工业用地、居住用地以及交通用地混淆严重，从而分类精度相对较低，这与土地利用场景的实际情况一致，即4种土地利用类型视觉上相似性较高，如图1所示。

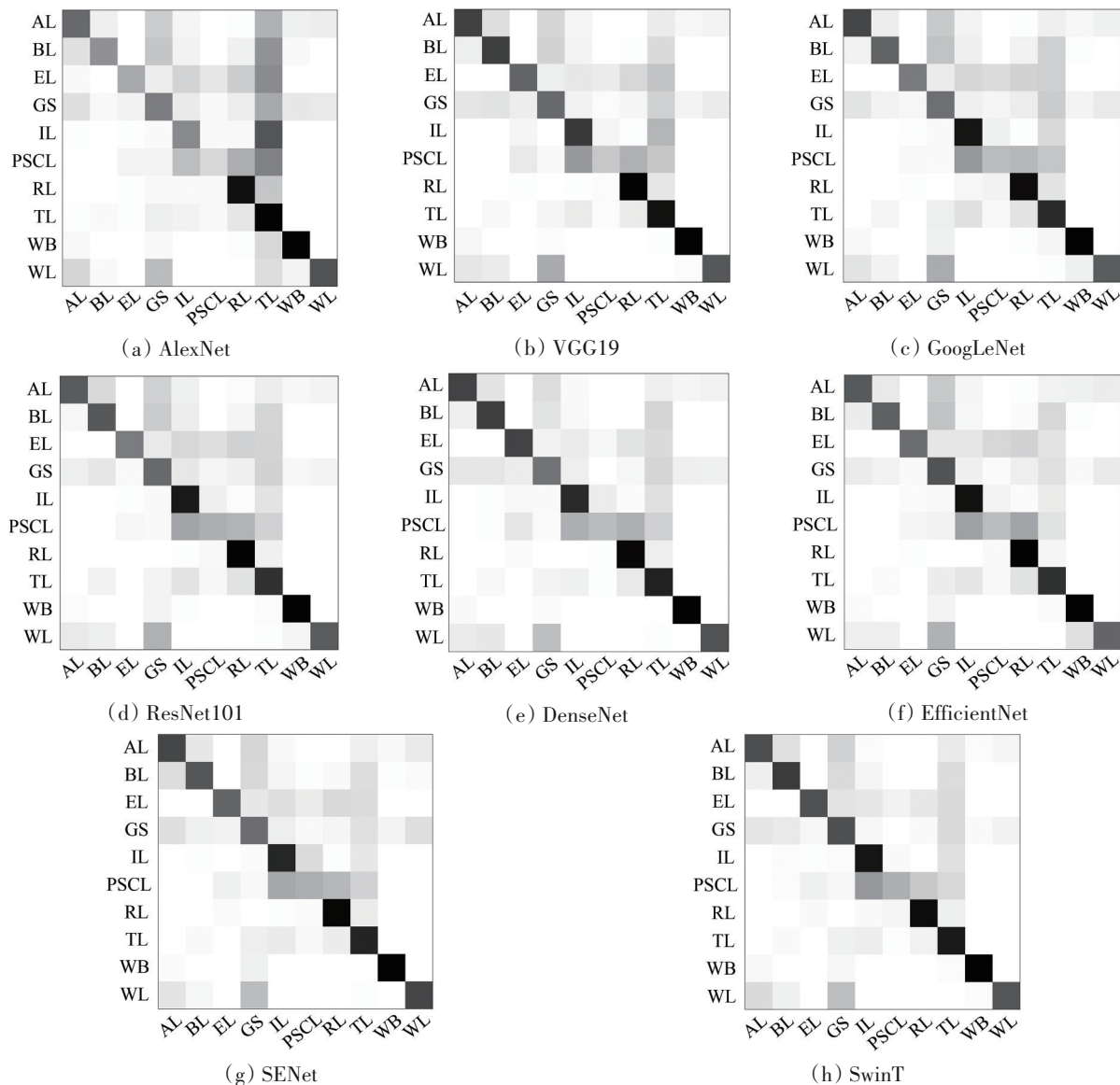


图4 MtSCCD_LUSC测试集A各网络的分类结果混淆矩阵

Fig. 4 Confusion matrixes for each network on testing set A in MtSCCD_LUSC dataset

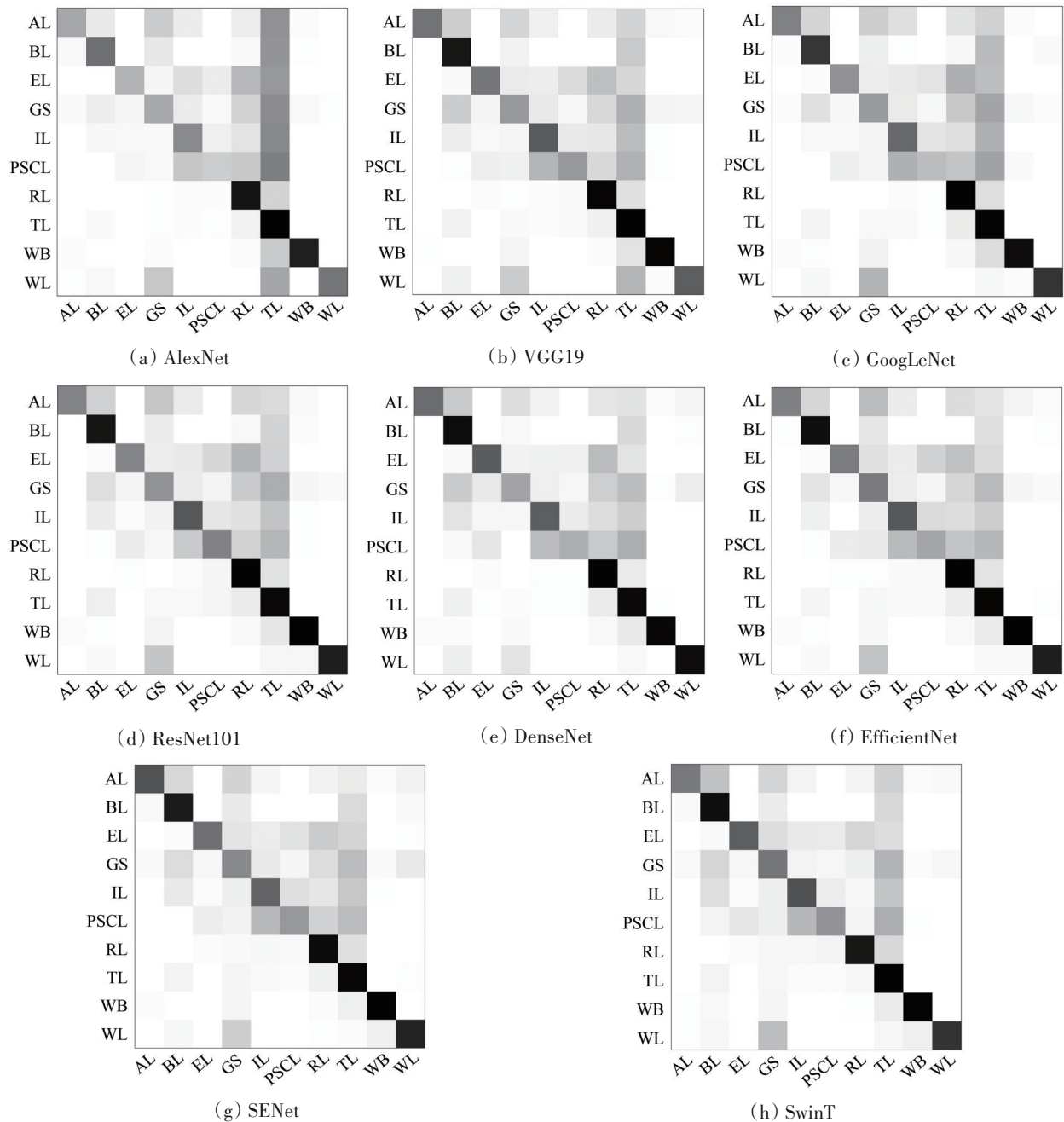


图5 MtSCCD_LUSC测试集B各网络的分类结果混淆矩阵

Fig. 5 Confusion matrixes for each network on testing set B in MtSCCD_LUSC dataset

3.4 MtSCCD_LUCD数据集实验结果

表7给出了传统分类后检测方法 (CDC_1和CDC_2) 和基于相似度的变化检测方法 (CFSM和SSCD) 对MtSCCD_LUCD数据的变化检测结果。由表7中结果可知, 对于测试集A和B, 基于相似度的变化检测方法优于传统的基于分类的变化检测方法。这是因为基于分类的变化检测方法需要对两个时相的场景各做一次分类, 从而导致最终的变化检测结果受两次分类的影响较大。与之相反, 基于相似度的变化检测利用两个时相场景的相似

度阈值判断场景是否发生变化, 避免了两次分类带来的误差累积, 从而能够得到更好的变化检测结果。此外, 对于两种相似度变化检测方法, CFSM (SwinT) 的总体精度和Kappa系数均高于SSCD方法, 尤其是显著提升了SSCD方法的Kappa系数。然而, SSCD是更具潜力的变化检测方法, 主要体现在SSCD虽然使用VGG16作为骨干网络提取特征, 但总体精度和CFSM (SwinT) 相差较小, 因此, 若以ResNet或Transformer作为骨干网络进行特征提取, SSCD的变化检测精度有望进一步提升。

表7 MtSCCD_LUCD数据集不同方法场景变化检测结果

Table 7 Change detection results of different methods on MtSCCD_LUCD dataset

方法	网络	测试集 A		测试集 B		阈值
		OA	Kappa	OA	Kappa	
CDC_1	AlexNet	0.6122	0.2494	0.6304	0.2357	—
	VGG19	0.6268	0.2508	0.6557	0.2633	—
	GoogLeNet	0.6327	0.2623	0.6367	0.2320	—
	ResNet101	0.6788	0.3209	0.6993	0.3233	—
	DenseNet	0.7637	0.7464	0.7778	0.7640	—
	EfficientNet	0.7363	0.7155	0.7513	0.7345	—
	SENet	0.7420	0.7221	0.7452	0.7275	—
	SwinT	0.7759	0.4588	0.7596	0.4116	—
	ViT	0.6004	0.2174	0.6586	0.2529	—
CDC_2	AlexNet	0.6881	0.2780	0.6812	0.3039	—
	VGG19	0.7302	0.3812	0.7495	0.3890	—
	GoogLeNet	0.6958	0.3002	0.6479	0.2688	—
	ResNet101	0.6685	0.2790	0.6329	0.2644	—
CFSM	AlexNet	0.7808	0.2313	0.8248	0.3687	0.3428
	VGG19	0.7674	0.1669	0.7966	0.2475	0.3461
	GoogLeNet	0.8062	0.3661	0.7861	0.2747	0.3567
	ResNet101	0.8620	0.4516	0.8469	0.3707	0.2847
	DenseNet	0.8163	0.8163	0.8251	0.8251	0.5072
	EfficientNet	0.8163	0.8163	0.8264	0.8264	0.4525
	SENet	0.8322	0.8266	0.8556	0.8515	0.5063
	SwinT	0.8775	0.8735	0.8820	0.8774	0.5191
ViT	0.8163	0.8163	0.8264	0.8164	0.7765	
SSCD		0.8533	0.5202	0.8196	0.4675	0.5159







注：表中加粗数值表示各评价指标的最好结果。

为了直观分析各方法的检测效果，表8给出了MtSCCD_LUCD数据集部分检测实例。可以看出，











对于同一网络，基于相似度方法的检测效果优于传统的基于分类的检测方法。

表8 MtSCCD_LUCD数据集变化检测实例

Table 8 Change detection examples of MtSCCD_LUCD dataset

数据			变化检测方法				
时相1	时相2	真值	CDC_1 (DenseNet)	CDC_1 (SwinT)	CFSM (DenseNet)	CFSM (SwinT)	SSCD
		未变化	×	√	√	×	√
农业用地	农业用地						
		未变化	×	×	√	√	√
绿地	绿地						
		变化	√	×	√	×	√
工业用地	裸地						

续表

数据			变化检测方法				SSCD
时相 1	时相 2	真值	CDC_1 (DenseNet)	CDC_1 (SwinT)	CFSM (DenseNet)	CFSM (SwinT)	
		变化	√	√	√	×	√
农业用地	裸地						
		未变化	√	×	√	√	√
裸地	裸地						
		未变化	×	×	√	√	√
教育用地	教育用地						
		变化	√	√	√	×	√
裸地	农业用地						
		变化	√	√	×	√	√
裸地	教育用地						

注: 表中“√”和“×”分别表示变化检测结果正确和错误。

4 结 论

本文利用中国 5 个城市中心区域的高分影像, 构建了目前规模最大的场景分类与变化检测数据集 MtSCCD。该数据集包括 MtSCCD_LUSC 和 MtSCCD_LUCD 两个子数据集, 两个子数据集均包含 10 种土地利用类型, 共 65548 幅图像。基于上述两个子数据集, 本文评估了多个深度学习网络的场景分类与变化检测效果, 为相关研究人员提供了重要参考。最后, 希望本文构建的 MtSCCD 数据集能够促进土地利用类型识别与监测领域的研究进展。

参考文献 (References)

Bai K, Mu X D, Chen X B, Zhu Y Q and You X A. 2022. Unsupervised remote sensing image scene classification based on semi-supervised learning. *Acta Geodaetica et Cartographica Sinica*, 51(5): 691-702 (白坤, 慕晓冬, 陈雪冰, 朱永清, 尤轩昂. 2022. 融合半监督学习的无监督遥感影像场景分类. *测绘学报*, 51(5): 691-702) [DOI: 10.11947/j.AGCS.2022.20210270]

Cheng G, Han J W and Lu X Q. 2017. Remote sensing image scene classification: benchmark and state of the art. *Proceedings of the IEEE*, 105(10): 1865-1883 [DOI: 10.1109/JPROC.2017.2675998]

Cheng G, Wang G X and Han J W. 2022. ISNet: towards improving separability for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5623811 [DOI: 10.1109/TGRS.2022.3174276]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16x16 words: transformers for image recognition at scale. arXiv:2010.11929

Feng Q L, Chen B A, Li G Q, Yao X C, Gao B B and Zhang L C. 2022. A review for sample datasets of remote sensing imagery. *National Remote Sensing Bulletin*, 26(4): 589-605 (冯权泷, 陈泊安, 李国庆, 姚晓闯, 高秉博, 张连翀. 2022. 遥感影像样本数据集研究综述. *遥感学报*, 26(4): 589-605) [DOI: 10.11834/jrs.20221162]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]
- Huang Y H and Zhou W X. 2022. Similarity method for high-resolution remote sensing scene change detection. *Bulletin of Surveying and Mapping*, (8): 48-53 (黄宇鸿, 周维勋. 2022. 高分辨率遥感影像场景变化检测的相似度方法. *测绘通报*, (8): 48-53) [DOI: 10.13474/j.cnki.11-2246.2022.0231]
- Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc.: 1097-1105
- LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436-444 [DOI: 10.1038/nature14539]
- Li H F, Dou X, Tao C, Wu Z X, Chen J, Peng J, Deng M and Zhao L. 2020. RSI-CB: a large-scale remote sensing image classification benchmark using crowdsourced data. *Sensors*, 20(6): 1594 [DOI: 10.3390/s20061594]
- Liu K, Zhou Z, Li S Y, Liu Y F, Wan X, Liu Z W, Tan H and Zhang W F. 2020. Scene classification dataset using the Tiangong-1 hyperspectral remote sensing imagery and its applications. *Journal of Remote Sensing (Chinese)*, 24(9): 1077-1087 (刘康, 周壮, 李盛阳, 刘云飞, 万雪, 刘志文, 谭洪, 张万峰. 2020. 天宫一号高光谱遥感场景分类数据集及应用. *遥感学报*, 24(9): 1077-1087) [DOI: 10.11834/jrs.20209323]
- Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin transformer: hierarchical vision transformer using shifted windows//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE: 9992-10002 [DOI: 10.1109/ICCV48922.2021.00986]
- Ma L, Liu Y, Zhang X L, Ye Y X, Yin G F and Johnson B A. 2019. Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152: 166-177 [DOI: 10.1016/j.isprsjprs.2019.04.015]
- Qian X L, Li J, Cheng G, Yao X W, Zhao S N, Chen Y B and Jiang L Y. 2018. Evaluation of the effect of feature extraction strategy on the performance of high-resolution remote sensing image scene classification. *Journal of Remote Sensing*, 22(5): 758-776 (钱晓亮, 李佳, 程焱, 姚西文, 赵素娜, 陈宜滨, 姜利英. 2018. 特征提取策略对高分辨率遥感图像场景分类性能影响的评估. *遥感学报*, 22(5): 758-776) [DOI: 10.11834/jrs.20188015]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. *arXiv*: 1409.1556 [DOI: 10.48550/arXiv.1409.1556]
- Sui H G, Feng W Q, Li W Z, Sun K M and Xu C. 2018. Review of change detection methods for multi-temporal remote sensing imagery. *Geomatics and Information Science of Wuhan University*, 43(12): 1885-1898 (眭海刚, 冯文卿, 李文卓, 孙开敏, 徐川. 2018. 多时相遥感影像变化检测方法综述. *武汉大学学报(信息科学版)*, 43(12): 1885-1898) [DOI: 10.13203/j.whu.gis20180251]
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A. 2015. Going deeper with convolutions//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 1-9 [DOI: 10.1109/cvpr.2015.7298594]
- Tan M X and Le Q V. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach: [s.n.]: 6105-6114
- Wu C, Zhang L F and Zhang L P. 2016. A scene change detection framework for multi-temporal very high resolution remote sensing images. *Signal Processing*, 124: 184-197 [DOI: 10.1016/j.sigpro.2015.09.020]
- Wu C, Zhang L P and Du B. 2017. Kernel slow feature analysis for scene change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 55(4): 2367-2384 [DOI: 10.1109/tgrs.2016.2642125]
- Wu F, Zhang H, Wang C, Li L, Li J J, Chen W R and Zhang B. 2022. SARBuD1.0: a SAR building dataset based on GF-3 FSII imageries for built-up area extraction with deep learning method. *National Remote Sensing Bulletin*, 26(4): 620-631 (吴樊, 张红, 王超, 李璐, 李娟娟, 陈卫荣, 张波. 2022. SARBuD 1.0: 面向深度学习的 GF-3 精细模式 SAR 建筑数据集. *遥感学报*, 26(4): 620-631) [DOI: 10.11834/jrs.20220296]
- Xia G S, Hu J W, Hu F, Shi B G, Bai X, Zhong Y F, Zhang L P and Lu X Q. 2017. AID: a benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7): 3965-3981 [DOI: 10.1109/tgrs.2017.2685945]
- Xia G S, Yang W, Delon J, Gousseau Y, Sun H and Maitre H. 2010. Structural high-resolution satellite image indexing//ISPRS TC VII Symposium - 100 Years ISPRS. 298-303
- Yang B S, Han X and Dong Z. 2021. Point cloud benchmark dataset WHU-TLS and WHU-MLS for deep learning. *National Remote Sensing Bulletin*, 25(1): 231-240 (杨必胜, 韩旭, 董震. 2021. 点云深度学习基准数据集. *遥感学报*, 25(1): 231-240) [DOI: 10.11834/jrs.20210542]
- Yang Y and Newsam S. 2010. Bag-of-visual-words and spatial extensions for land-use classification//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose: ACM: 270-279 [DOI: 10.1145/1869790.1869829]
- Yu W Q, Cheng G, Wang M J, Yao Y Q, Xie X X, Yao X W and Han J W. 2022. MAR20: a benchmark for military aircraft recognition in remote sensing images. *National Remote Sensing Bulletin*, 1-11 (禹文奇, 程焱, 王美君, 姚艳清, 谢星星, 姚西文, 韩军伟. 2022. MAR20: 遥感图像军用飞机目标识别数据集. *遥感学报*: 1-11) [DOI: 10.11834/jrs.20222139]
- Yuan J W, Ru L X, Wang S G and Wu C. 2022. WH-MAVS: a novel dataset and deep learning benchmark for multiple land use and land cover applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 1575-1590 [DOI: 10.1109/jstars.2022.3142898]
- Yuan J W, Wu C, Du B, Zhang L P and Wang S G. 2020. Analysis of landscape pattern on urban land use based on GF-5 hyperspectral data. *Journal of Remote Sensing (Chinese)*, 24(4): 465-478 (袁静文, 武辰, 杜博, 张良培, 王树根. 2020. 高分五号高光谱遥感影像的城市土地利用景观格局分析. *遥感学报*, 24(4): 465-478) [DOI: 10.11834/jrs.20209252]
- Zhang L P and Wu C. 2017. Advance and future development of change detection for multi-temporal remote sensing imagery. *Acta Geodaetica et Cartographica Sinica*, 46(10): 1447-1459 (张良培, 武辰. 2017. 多时相遥感影像变化检测的现状与展望. *测绘学报*, 46(10): 1447-1459) [DOI: 10.11947/j.AGCS.2017.20170340]

- Zhang L P, Zhang L F and Du B. 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2): 22-40 [DOI: 10.1109/mgrs.2016.2540798]
- Zhou W X, Newsam S, Li C M and Shao Z F. 2018. PatternNet: a benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145: 197-209 [DOI: 10.1016/j.isprsjprs.2018.01.004]
- Zhu X X, Tuia D, Mou L C, Xia G S, Zhang L P, Xu F and Fraundorfer F. 2017. Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4): 8-36 [DOI: 10.1109/mgrs.2017.2762307]
- Zou Q, Ni L H, Zhang T and Wang Q. 2015. Deep learning based feature selection for remote sensing scene classification. *IEEE Geoscience and Remote Sensing Letters*, 12(11): 2321-2325 [DOI: 10.1109/lgrs.2015.2475299]

MtSCCD: Land-use scene classification and change-detection dataset for deep learning

ZHOU Weixun^{1,2}, LIU Jinglei¹, PENG Daifeng¹, GUAN Haiyan¹, SHAO Zhenfeng³

1. School of Remote Sensing & Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China;

2. State Key Laboratory of Remote Sensing Science, Beijing Normal University, Beijing 100875, China;

3. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Abstract: Land-Use Scene Classification and change Detection (LUSCD) aim to recognize land-use types and monitor their changes by using Remote-Sensing (RS) images, which play an important role in urban planning and land-use optimization. In the era of RS big data, conventional hand-crafted feature-based methods are infeasible for LUSCD because the extracted features are not sufficiently discriminative for RS images with high complexity. As a novel data-driven paradigm for information extraction from RS images, deep learning provides a new solution for LUSCD. However, the existing publicly available datasets have limited samples and is thus unable to train a successful deep-learning model. Therefore, it has great significance in constructing an open and large-scale LUSCD benchmark.

To advance the progress of LUSCD using deep-learning methods, this paper releases a large-scale scene classification and change-detection dataset termed Multi-temporal Scene Classification and Change Detection (MtSCCD). The RGB images in MtSCCD are cropped from large-size high-resolution RS images captured from the central areas of five China cities, namely, Hangzhou, Shanghai, Wuhan, Nanjing, and Hefei. The size of the cropped images is 300×300 pixels with the spatial resolution of around 1 m. MtSCCD has 10 land use classes, which are residential land, public service and commercial land, educational land, industrial land, transportation land, agricultural land, water body, green space, woodland, and woodland. Based on the cropped land-use images in MtSCCD, this paper constructs two sub-datasets termed MtSCCD_LUSC (MtSCCD Land Use Scene Classification) and MtSCCD_LUCD (MtSCCD Land Use Change Detection) for land-use scene classification (LUSC) and land-use change detection (LUCD), respectively. MtSCCD dataset has the following characteristics. (1) It is currently the largest publicly available LUSCD dataset, and both of the two sub-datasets (i.e., MtSCCD_LUSC and MtSCCD_LUCD) have 65548 images in total. (2) The images in MtSCCD are split into training set, validation set, and testing set according to the five cities. For example, images from three of the five cities are randomly split into training and validation set, whereas the rest remain to be the testing set. Therefore, MtSCCD has high extensibility, i.e., it can be easily extended to be a larger dataset. (3) For a deep-learning model, the training set and testing set are categorized from different cities, so it is beneficial to demonstrate the model's generalization ability. (4) MtSCCD has high intra-class diversity, making it a challenging dataset.

Based on MtSCCD_LUSC and MtSCCD_LUCD, this paper evaluates several deep-learning feature-based methods for LUSC and LUCD. Specifically, AlexNet, VGG networks (i.e., VGG16 and VGG19), GoogLeNet, and ResNet networks (i.e., ResNet18, ResNet50, and ResNet101) are selected to extract deep-learning features that are then fed into SVM for LUSC. We also evaluate DenseNet, EfficientNet, SENet, ViT, and SwinT for LUSC. Two kinds of LUCD approaches including conventional classification-based methods and current similarity-based methods have been evaluated. Experimental results show that the highest overall accuracy of MtSCCD_LUSC dataset is around 76%, indicating much room for improvement. Regarding LUCD, similarity-based methods particularly similarity learning-based ones outperform classification-based methods by a significant margin, providing a promising research direction for LUCD.

This paper presents the currently largest scene classification and change-detection dataset MtSCCD based on high-resolution RS images of the central area of five China cities. MtSCCD contains two subsets MtSCCD_LUSC and MtSCCD_LUCD. Both had 10 land-use types and 65548 images in total. Based on the two sub-datasets, this paper evaluates the performance of several deep networks for scene classification and change detection, expecting to provide baseline results for related researchers. We hope that the MtSCCD dataset can promote this progress in land-use type recognition and monitoring.

Key words: land use, scene classification, change detection, dataset, information extraction, feature extraction, deep learning, convolutional neural network

Supported by National Natural Science Foundation of China (No. 42001285); Natural Science Foundation of Jiangsu Province (No. BK20200813); Open Fund of State Key Laboratory of Remote Sensing Science (No. OFSLRSS202215); Key Laboratory of Land Satellite Remote Sensing Application, Ministry of Natural Resources of the People's Republic of China (No. KLSMNR-G202202)