

面向视觉与遥感图像的关联数据集综述及展望

李京竹¹, 刘小煜², 郝凌翔², 吴国玄², 孙铮², 巩敏², 李亚²

1. 中国科学院大学 集成电路学院, 北京 101408;

2. 中国科学院空天信息创新研究院, 北京 100094

摘要: 关联是目标感知向目标认知转变中的一个重要步骤, 近年来基于计算机视觉技术的目标关联分析已逐渐受到相关领域的研究人员关注, 遥感领域关联分析则起步较晚, 其中一个重要原因是缺少面向遥感图像的关联数据集。数据集是支撑算法训练、验证以及评估的基础, 多样化的数据集能够推动算法持续优化与性能提升。然而, 当前尚未有研究对已发布的关联数据集进行全面的归纳和梳理, 尤其是在遥感领域, 对关联数据集的研究还远不够充分。针对这一问题, 本文对相关领域的文献进行全面调研, 围绕视觉领域以及遥感领域, 对已发布的图像、视频两种模态 15 个视觉关联数据集以及 4 个遥感图像关联数据集分别进行了回顾和概述, 系统性地介绍了其特点、数据规模、关联类别、三元组数量等内容。同时, 本文对已在上述数据集验证的当前主流目标关联分析算法进行了梳理, 将现有算法分为基于条件随机场的关联分析算法和基于深度学习的关联分析算法, 对每类算法的优缺点进行了分析。最后本文针对当前关联数据集和算法模型的局限性, 分别提出了未来可能的研究方向。面向关联数据集, 提出一个构建面向时序数据的遥感关联数据集的设想, 并初步研制了一套面向遥感图像的关联标注及可视化工具, 给未来关联数据集的建设方向提出了优化建议。面向关联算法, 本文提出了关联动态分层建模、动态场景图等创新方法, 以期对相关领域研究人员提供思路。综上所述, 本文为后续遥感领域关联数据集的构建及相关算法的研究工作提供参考。

关键词: 遥感, 目标关联分析, 关联数据集, 场景图生成, 综述, 数据集

中图分类号: TP79/P2

引用格式: 李京竹, 刘小煜, 郝凌翔, 吴国玄, 孙铮, 巩敏, 李亚. 2025. 面向视觉与遥感图像的关联数据集综述及展望. 遥感学报, 29(5): 1089–1102

Li J Z, Liu X Y, Hao L X, Wu G X, Sun Z, Gong M and Li Y. 2025. Overview and prospect of relationship dataset for visual and remote sensing images. National Remote Sensing Bulletin, 29(5): 1089–1102 [DOI: 10.11834/jrs.20254523]

1 引言

遥感图像智能解译是利用人工智能技术对遥感图像进行自动化处理和分析的过程, 通过这些技术可以实现对遥感图像中的地物分类 (Wang 等, 2023b, 2024a, 2024b)、目标检测 (Li 等, 2023a; Zhang 等, 2023; Xu 等, 2024)、语义分割 (Hong 等, 2023; Ma 等, 2024) 等任务的智能化处理, 尤其是大模型的出现, 使得多任务处理得以集中于统一模型, 并且能够更好地适应复杂场景 (Sun 等, 2023)。随着计算机视觉技术研究逐渐深入, 相关领域内的研究人员不再满足于只从遥感图像中获取目标的几何外观、位置和类别等

感知信息, 而是着手运用各种技术方法进一步挖掘蕴含在图像中的丰富信息, 实现目标认知 (Fu 等, 2024)。关联分析是目标感知到认知的一个重要步骤 (Wang 等, 2023a), 它基于提取的目标检测框和目标分类结果, 推理出图像中各个目标之间的空间及语义关系, 深化对图像内容的理解。

近年来相关的算法研究在视觉领域已取得了不错的进展, 但在遥感领域, 受限于关联数据集的不足, 还难以支持更深入的遥感图像认知任务, 相比自然场景图像, 空天观测平台获取的遥感图像有着更丰富的目标信息, 图像背景、目标间关系更为复杂, 关联分析难度较大, 因而构建面向遥感图像的大规模关联数据集, 支撑遥感图像解

收稿日期: 2024-11-27; 预印本: 2025-02-24

基金项目: 国家重点研发计划 (编号: 2021YFB3900504)

第一作者简介: 李京竹, 研究方向为态势分析预测。E-mail: lijingzhu21@mails.UCAS.ac.cn

通信作者简介: 刘小煜, 研究方向为关联分析预测。E-mail: liuxiaoyu@aircas.ac.cn

译认知任务就显得十分重要。

图1展示了本文检索到的公开的与目标关联相关的数据集，从近10年数据集数量的变化趋势可以看出，关注目标关联的数据集数量逐年增加，并由纯静态图像向动态视频扩展，已成为目标认知领域的热点方向，但目前尚未有涵盖多类关联数据集的总结分析工作。此外，在遥感领域，面向目标关联分析任务的数据集较少，且极少存在服务于遥感目标关联的相关工作。

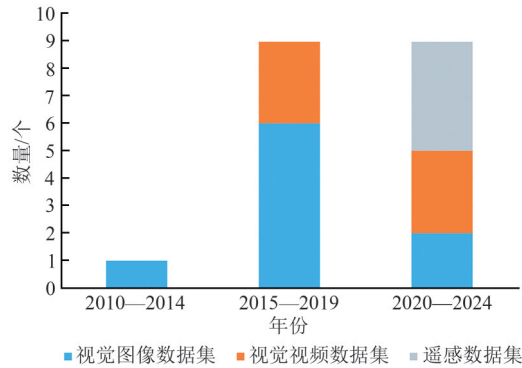


图1 关联数据集统计 (此次检索于2024年9月进行)

Fig. 1 Relationship dataset statistics (This search was conducted in September 2024)

本文希望可以为遥感领域目标关联分析做出以下3方面贡献：

(1) 本文对面向视觉和面向遥感的关联数据集进行全面、及时的综述，并总结关联数据集的局限，以期能够为读者提供关联数据集的总体情况；

(2) 对于基于上述数据集的相关算法，本文也进行了相应梳理，介绍了各类算法的优缺点，以及应用于遥感图像时可能存在的问题；

(3) 基于上述对数据集和算法发展现状的总结和局限性分析，本文提出新一代面向遥感图像的目标关联数据集设想，并提供相应算法的可行思路，对满足现实应用需求具有实际意义。

2 关联数据集

关联数据集是在目标检测、语义分割数据集基础上增加多个目标三元组后形成的数据集，旨在支撑更高层次的视觉认知，具体地，首先需要对大量原始数据进行目标检测框标注/像素级标注，目标框通常为水平边界框 HBB (Horizontal Bounding Box) 和定向边界框 OBB (Oriented Bounding Box) 两种形式，像素级标注则以 mask 为主，在此基础上构建多个形式如<目标 1, 关系, 目标 2>的三元

组，实现关联信息的标注。截止目前，国内外已有多家单位累计发布了十余个关联数据集，可支撑以场景图生成 SGG (Scene Graph Generation)、图卷积网络为主的关联分析算法验证测试。下文首先介绍视觉领域的关联数据集，之后对遥感领域关联数据集进行详细介绍。

2.1 视觉领域关联数据集

在计算机视觉领域，近年来涌现了一批数据成果，如表1所示，可通过图像、视频两种数据模态进行划分，其中 Visual Genome (Krishna 等, 2017)、VrR-VG (Liang 等, 2019)、Open Images Dataset V4 (Kuznetsova 等, 2020) 是比较具有代表性的图像关联数据集，而视频关联数据集，则以 VidOR (Shang 等, 2019)、Action Genome (Ji 等, 2020)、VirtualHAG (Qiu 等, 2023) 为代表，可支撑动态场景中的场景图生成等关联算法验证。

表1 视觉领域关联数据集

Table 1 Relationship datasets in the field of visual

数据类型	数据集	数据规模	目标关联		时间/年
			关联类别	三元组数量	
图像数据	Visual Phrases (Sadeghi 等, 2012)	2769	9	1796	2011
	Scene Graph (Johnson 等, 2015)	5000	68	109535	2015
	VRD (Lu 等, 2016)	5000	70	37993	2016
	Visual Genome (Krishna 等, 2017)	108077	40480	2347187	2017
	SpatialVOC2K (Belz 等, 2018)	2026	34	9804	2018
	SpatialSense (Yang 等, 2019)	11569	9	13229	2019
	VrR-VG (Liang 等, 2019)	58983	117	203375	2019
	Open Images Dataset V4 (Kuznetsova 等, 2020)	9278275	329	374768	2020
	PSG (Yang 等, 2022)	49000	56	279300	2022
	VidVRD (Shang 等, 2017)	1000	132	3972	2017
视频数据	CAD120++ (Zhuo 等, 2019)	551	10	—	2019
	VidOR (Shang 等, 2019)	10000	50	297352	2019
	Action Genome (Ji 等, 2020)	9848	25	1715568	2020
	HOMAGE (Rai 等, 2021)	5700	29	583481	2021
	VirtualHAG (Qiu 等, 2023)	2588	19	—	2023

2.1.1 图像关联数据集

Sadeghi 和 Farhadi (2012) 构建了 Visual Phrases 关联数据集, 该数据集是计算机视觉领域内第一个具有目标关系标注的图像数据集, 在这之后, 李飞飞团队为视觉关系识别领域的数据库建设做出了巨大贡献, 先后发布了 Scene Graph (Johnson 等, 2015)、VRD (Lu 等, 2016)、Visual Genome (Krishna 等, 2017) 数据集, 并确定了三元组的标注方式, 其中 Visual Genome 数据集关联类型达到了 40480 种, 已成为领域内的主流基准数据集之一 (图2)。



图2 静态图像的场景图表示

Fig. 2 Scene representation of static images

后续陆续有一些研究工作对数据集进行细化, Liang 等 (2019) 提出的 VrR-VG 是基于 Visual Genome 构建的场景图数据集, 它去除了 Visual Genome 数据集中冗余的视觉无关信息, 仅保留重要的视觉相关信息, 有效提升模型语义信息学习能力, 同时对 Visual Genome 数据集中的关系进行了筛选和提炼, 一定程度上缓解了数据集的长尾分布效应。Belz 等 (2018) 构建的 SpatialVOC2K 是领域内第一个多语种关联数据集, 支持英语和法语两种语言的关联场景图生成。Yang 等 (2019) 构建的 SpatialSense 数据集是一个用于空间关系识别的基准数据集, 相比之前的数据集, 它对空间关系的定义限定更为精确, 避免了模糊性和歧义, Open Images Dataset v4 (Kuznetsova 等, 2020) 数据集由 Google AI 发布, 后续又发布了 v5—v7 的迭代版本, 该数据集数据丰富, 标注质量高、可扩展性强, 目前已作为视觉关系检测的主流基准数据集之一, 同时 Open Images Dataset v4 构建了良好的开源生态, 为计算机视觉社区的研究者和开发者提供了免费的数据资源, 促进了技术的发展。PSG 数据集 (Yang 等, 2022) 是基于全景分割的全场景图生成数据集, 通过像素级标注解决了传统场景图生成任务中边界框标注不准确和背景无法标注等固有缺点, 推动了视觉模型对场景更全

面的理解和感知。

2.1.2 视频关联数据集

随着场景图关联算法在静态图像数据上取得了巨大进展, 相关领域的算法研究逐渐转向视频数据, 同时为支撑算法验证, 视频关联数据集开始被构建出来。

Shang 等 (2017) 基于 ILVSRC2016-VID 数据集集中的视频素材构建了 VidVRD 数据集, 该数据集专注于捕捉到视频中目标之间动态的、随时间变化的关系, 它的出现为视频视觉关系检测的研究提供了重要的基础, Shang 等 (2019) 后续又发布了 VidOR 数据集, 相比 VidVRD, 该数据集样本量更大, 且增加了人群密集场景、快速运动场景、光线变化剧烈场景等复杂场景, 同时标注也更为精细严格。Zhuo 等 (2019) 基于 CAD-120 人类活动视频数据集进行逐帧标注, 形成具有目标、属性、关系的时间序列, 构建 CAD-120++ 数据集, 共包含 32327 帧的 551 段视频切片, 并基于该数据集进行了算法验证。在这之后, Ji 等 (2020) 构建了 Action Genome 数据集, 该数据集是一个面向视频动作理解领域的数据集, 它将视频中的人类动作分解为时空场景图的表示, 如图3所示, 在视频中不仅标记了物体及其位置, 还详细记录了人物与物体之间的关系, 这种深度标注极大地丰富了理解视频内容的可能性, 该数据集是当前相关领域主流的基准数据集之一。

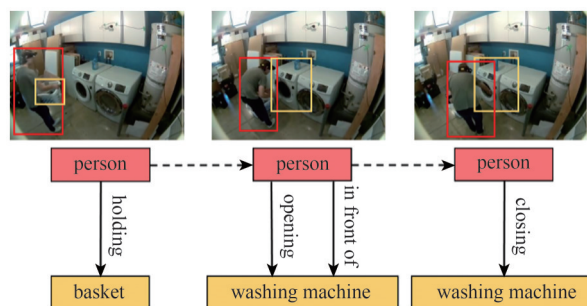


图3 时空场景图表示

Fig. 3 Spatiotemporal scene diagram representation

Rai 等 (2021) 参考 Action Genome 构建了 HOMAGE 数据集, 该数据集是具有多种模式和视角的人类动作识别基准数据集, 相比之前的 Action Genome 数据集, 其提升体现在 3 个方面, 一是场景更为聚焦, 重点关注住宅环境中的动作, 二是数据丰富性上, 在数据模态 (RGB、红外

等)、视角(自我视角、第三人称视角)都有所提升,三是在标注信息上,对关系进行了分层次标注。在这两个数据集基础上,Qiu等(2023)构建了VirtualHAG数据集,该数据集针对Action Genome和HOMAGE两个数据集出现的稀疏标注和帧间标注错误等问题,数据标注更为严格,保证了帧间的一致性,并引入了时间、空间上下文关系,以支撑基于时间关联的推理。

2.2 遥感图像关联数据集

相比自然场景图像,遥感图像呈现出显著不同的特点,单张遥感图像覆盖范围更大,地物目标更多且密集,再者,遥感图像的分析专业性高、

难度大,涉及大量背景知识,导致遥感领域的关联数据集更难标注构建,因此在缺少大量训练样本的情况下,视觉领域的算法很难直接迁移到遥感领域。但近年来随着研究人员逐渐发现该领域的样本数据空缺,也开展了遥感图像场景图生成数据集等一系列研究。目前基于不同空间分辨率的光学遥感影像,也形成了一批以RSSGD(Li等,2021)、GRTRD(Chen等,2022)、STAR(Li等,2025)数据集为代表的成果,如表2所示,数据集规模、目标数量、关联类型也呈现出逐年增加的态势,尤其是最新的基准数据集STAR,已可支撑基于遥感图像的场景图生成算法验证试验(图4),下文将对各个数据集进行详细介绍。

表 2 遥感领域关联数据集
Table 2 Relationship datasets in the field of remote sensing

数据集	图像大小	空间分辨率/m	数据规模	目标关联		时间/年
				关联类别	三元组数量	
S2SG(Lin等,2022)	224	—	—	全部为空间关系,包括cross、next to、through等13种	76	2021
GRTRD(Chen等,2022)	—	—	—	主要为空间关系,包括surround、blend、inlaid、across等12种	1200	2021
STAR(Li等,2025)	600	0.5	3200	分为拓扑关系、方向关系、距离关系3大类,包括pass through、cover、north、near等33种	18602	2022
S2SG(Lin等,2022)	512—31096	0.15—1.0	—	可分为停驻状态、空间拓扑、功能描述、移动状态、距离警告、电路布局、建筑状态以及排放状态8大类,包括parked alongside with、parallelly parked on、co-storage with、connect等58种	400795	2024

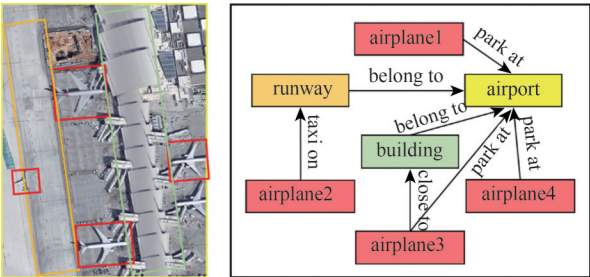


图 4 遥感图像的场景图表示
Fig. 4 Scene representation of remote sensing images

为支撑面向遥感图像的场景图生成算法验证,Li等(2025)基于RSICD遥感图像数据集构建了RSSGD数据集,RSICD数据集包含上万幅不同分辨率的遥感图像,图像大小固定为224×224像素,研究团队在该数据集基础上标注目标三元组,实现数据集构建。RSSGD数据集是遥感领域首个场景图关联数据集,它涵盖飞机、车辆、机场等39类典型遥感目标,并提供13种空间上的关联关系,包括cross、next to、through、surround等典型关

系,且分布相对均衡。RSSGD数据集中遥感图像全部采用水平检测框对目标进行标注,导致部分地物目标区域定位并不十分准确,同时涉及的关联类型和实例数量相对有限,缺少语义关系,但该数据集的出现,仍然促进了遥感图像关联算法的研究,具有重要意义。

在这之后,Lin等(2022)构建了S2SG数据集,与常规的关联数据集基于目标检测框构建三元组不同,S2SG数据集基于目标的像素级标注进行关联标注,在图像掩膜上构建关联三元组。S2SG数据集涵盖建筑物、道路、植被、水体等6类目标,surround、blend、inlaid、across等12种关系,且每种关系都标注了100个实例,合计1200个目标关联实例,为保证数据集的分布均衡性,研究团队只对具有明显关系的目标对进行标注,舍去了部分难例。S2SG数据集为相关领域研究起到了一定推动作用,但它更多关注面向语义分割任务的常见地物遥感图像,这导致数据集中

目标类型数量较少, 数据规模不大, 在一定程度上限制了它成为基准数据集。

2022 年, 研究人员构建了 GRTRD 数据集 (Chen 等, 2022), 它以国家地理状况监测所使用的高分影像作为数据源, 图像大小全部为 600×600 像素, 分辨率为 0.5 m, 共计 3200 张图像, 覆盖建筑区、林地、道路、草地、体育场等 12 类地物目标, 拓扑关系、方向关系、距离关系 3 大类共 33 种关联关系, 合计 18602 关系实例。相比之前的遥感关联数据集, GRTRD 数据集无论是在数据规模、质量, 还是在关联类型和实例数量上都有着较大的提升, 且详细介绍了其数据组织格式和标注方法, 都为后续相关领域的研究者提供了重要参考。但数据集只关注了一部分较大的地物目标, 在遥感图像中并未标注船舶、车辆等目标, 因而单张图像中的关联关系挖掘还不充分。

Li 等 (2025) 构建了 STAR 数据集, 该数据集是领域内首个面向大尺寸、高分辨率卫星影像的大规模场景图生成数据集, 相比之前的场景图生成数据集, STAR 数据集囊括了更多大尺寸的卫星影像, 最大达到了 27860×31096 像素, 除此之外, 它涵盖了机场、港口、核电站、火电站、水坝、体育场等 11 类与人类活动密切相关的复杂地理空间场景, 地理目标更为丰富, 在关联信息标注上, 具有 8 大类 58 个细粒度类别的关系标签, 且目标-关系三元组实例数量超过 40 万, STAR 数据集的出现为大尺度复杂场景下的场景图生成算法研究提供了重要的数据支撑, 可作为基准数据集推动智能解译技术向更高级的图像认知发展。

虽然当前关联分析已逐步从视觉领域推广至遥感领域, 并构建了大规模的遥感关联数据集, 但尚未出现面向遥感领域的视频关联数据集, 这也为研究者们提供了一个可行的方向。

2.3 当前数据集的局限性

由此可见, 计算机视觉领域已形成一批较为完备的关联数据集, 覆盖图像、视频两种模态, 可支持静态/动态等多种类型复杂场景下的场景图生成算法的训练验证, 相比而言, 遥感领域关联数据集的研究目前还不够充分, 语义关联类型相对有限, 且缺少多个大规模公开基准数据集, 但近年来越来越受到研究者重视, 持续产出新的成果。总的来说, 仍存在以下问题:

(1) 标注格式不统一。不同研究团队在构建数据集时使用各自的标注格式, 检测框的定义 (水平框、旋转框)、类别标签以及目标关系标注文件的组织形式, 都缺乏统一的规范, 导致数据的应用共享相对困难, 这在一定程度上限制了遥感图像关联数据集的开源生态建设;

(2) 遥感数据中的关联信息挖掘还不充分。目前的数据集成果中仍以标注卫星影像中空间层面的关联为主, 遥感时序数据利用不足, 时空关联 (例如: 车辆移动中与周围目标关系变化)、基于专家先验的较高层次语义关联 (例如: 小型拖船拖拽大船进出泊位) 相对有限, 无法支撑动态、复杂场景下的场景图生成;

(3) 长尾效应明显。由于场景图数据集的本身特性, 其类别分布并不均衡, 当数据集规模扩大后, 关联关系的分布呈现显著的长尾特性, 因而基于这类数据集训练的模型倾向于识别那些频繁出现的关系, 而对不常见的重要关系识别能力较差, 模型泛化能力不足, 难以满足实际需求;

(4) 缺乏复合型目标以及大尺度场景下的遥感图像。目前的数据成果通常只对单一目标进行了细致标注, 而忽略了一些具有重要意义的复合目标 (例如: 车队、建筑群、街区等), 导致模型无法学习到图像的全局信息, 同时现有数据成果也多以卫星影像切片为主, 无法支撑大范围复杂场景下的智能解译。

3 数据集相关算法分析

上述关联数据集已广泛应用于关联关系识别和场景图生成 (Chang 等, 2023) 任务, 为以深度学习为基础的相关算法发展奠定了数据基础, 本节对以上数据集作为训练及测试样本的相关算法进行归纳和分析, 分别介绍了基于条件随机场的方法和基于深度学习的方法, 以及关联分析算法常用的评估指标, 最后对目标关联分析算法的局限性进行总结。

3.1 基于条件随机场的目标关联分析算法

在关联分析任务中, 关联类型和目标类型间往往存在一种映射规律, 变现出较强的统计相关性, 有效利用这些规律可以帮助算法更准确、便捷地识别关联类型。条件随机场 CRF (Conditional Random Field) (Lafferty 等, 2001) 是一种典型的

统计学方法,通过学习样本和类别间的统计学特征,来估计样本属于某一类别的概率,常被用于图推理相关的任务中,如图像分割(Liang等,2016;Zheng等,2015)、图像检索(Johnson等,2015)、命名实体识别(Lample等,2016)。

基于条件随机场的关联分析算法包括两个步骤:目标检测和关联识别,目标检测获取图像中目标的区域视觉特征,关联识别基于检测出的视觉特征判定目标间的关联类别(Chang等,2023)。一些算法从目标检测和关联识别角度出发,均能取得了较好的模型效果。如DR-Net算法(Dai等,2017)通过迭代展开的方式,将统计推理过程转化成一个前馈神经网络结构。Xu等(2021)提出JM-SGG模型,通过建立统一的CRF模拟场景图中对象和关联的联合分布,充分利用上下文关联信息,提升了关联识别的准确性。

基于条件随机场的目标关联分析算法能够有效地对关联关系中的统计相关性进行建模,在深层语义关联分析时,往往不能取得良好的效果。

3.2 基于深度学习的目标关联分析算法

基于深度学习的目标关联分析的算法主要分为3类:基于视觉翻译嵌入的目标关联分析算法、基于上下文信息的目标关联分析算法、基于图网络的目标关联分析算法,下面分别对3类算法进行介绍。

3.2.1 基于视觉翻译嵌入的目标关联分析算法

与知识图谱类似,目标关联任务中也使用三元组的方式保存目标及关联数据,基于翻译嵌入的模型将实体和关系映射到低维向量空间中,学习实体和关系间的联系,因此也常用于目标关联分析任务。Zhang等(2017)提出场景图领域第一个端到端的关联识别网络VTransE,在低维空间中映射对象和谓词的特征来建模视觉关系,将关系三元组解释为向量平移,并构建一种新的特征提取层,提取3种类型的对象特征:类概率、位置和RoI视觉特征,实现对象和谓词间的知识转移。但VTransE模型只关注局部的主宾语特征,忽略了上下文信息,针对这一问题,MATransE(Gkanatsios等,2019)提出一种深度监督的双分支架构,在两个独立的分支中直接使用谓词特征和主语—宾语特征,并使用空间语言注意机制来

驱动视觉特征,在实际应用中取得了优异的效果。对于训练样本少而实际存在的关系,以上方法就不能很好的识别了。在VTransE的基础上,Hung等(2021)提出UVTransE方法,UVTransE使用主语和宾语并集的边界框捕获上下文信息,再删去其中与对象相关的信息,获得仅包含谓词的信息嵌入,在面对罕见关系时取得了更好的性能。

基于视觉翻译嵌入的目标关联分析算法使用直观的建模方法,增加了模型的可解释性,但其更多关注局部信息,对上下文信息考虑不充分,限制了算法性能提升。

3.2.2 基于上下文信息的目标关联分析算法

循环神经网络RNN(Recurrent Neural Network)/长短期记忆网络LSTM(Long Short-Term Memory)(Sherstinsky,2020)等方法在目标上下文信息获取方面具有天然优势,因此常被用于目标关联关系分析。IMP(Xu等,2017)提出一种原对偶图,不再孤立推断场景图每个组件,而是通过沿场景图的拓扑结构传递上下文消息来迭代地完善其预测。Zellers等(2018)认为局部预测中的强独立性假设限制了全局预测的质量,因此提出MotifNet模型,通过LSTM对全局上下文信息进行编码。然而上述方法往往只关注场景中的上下文信息,忽略了谓词之间的内部关联,Chen等(2019)提出谓词关联网络PANet,受限提取实例级和场景级上下文信息用于对象分析,随后使用注意力机制对谓词关联进行建模,取得了很好的效果。IS-GGT方法(Kundu和Aakur,2023)尝试将图形生成和关系建模分离,首先对检测到的实体间的底层交互图进行采样,再对该采样的语义结果进行推理,这种方法能够减少计算开销,同时保证场景图生成质量。Cong等(2023)提出了一种新的端到端场景图生成框架RelTR,利用实体解码器的实体检测结果来改善主体和客体的定位和分类,提升关系预测性能。

以上方法都是面向视觉场景,遥感图像在关联分析时信息跨度远大于传统视觉图像,因此现有方法很难在遥感关联分析任务中取得好的结果。SRSG模型(Lin等,2022)针对遥感图像尺度大、覆盖区域广的特点,将语义分割结果作为输入,更好地学习目标的空间位置和形态信息,在基于遥感图像的目标关联分析任务中取得了良好的效

果。Chen 等 (2022) 从目标和空间关系两方面提取特征, 并通过消息传递机制相互补充特征, 实现关系提取。Li 等 (2025) 提出一个上下文感知级联认知框架 HOD-Net, 用于处理远程上下文信息, 提升了遥感目标关联分析性能。

基于上下文信息的目标关联分析算法具有更好的关联上下文建模能力, 同时, 充分考虑上下文信息的前提使得此方法更适合于遥感领域大尺度图像数据。

3.2.3 基于图网络的目标关联分析算法

目标间关联可以被视为一个复杂的图结构, 图神经网络 (Wu 等, 2021) 能够有效学习到相邻节点间的局部信息, 因此被广泛用于关系推理 (Huang 等, 2023) 和视觉理解 (Li 等, 2023b; Shoori 等, 2023) 等任务。F-Net (Li 等, 2018) 是一种基于子图的场景图生成架构, 将图像聚类为若干子图, 子图内的对象对共享学习到的特征, 再将子图作为节点获得分解的连接图, 在其上传递消息以获得更好的表示, 通过使用子图替换原始图像的方式, 提升了场景图生成效率。Graph R-CNN (Yang 等, 2018) 引入关系提议网络, 有效地计算对象对间的相关性分数, 修剪不太可能的关联关系, 此外通过预测每个节点的边缘注意力, 调节不可靠的边缘信息流, 进而生成稀疏候选图结构。ARN (Qi 等, 2019) 模型引入语义转换模块, 将词嵌入和视觉特征映射到同一空间来生成语义嵌入, 而关系推理模块对实体和关系进行分类生成场景图, 特别地, ARN 还提出一个图自注意力模块, 旨在将所有关系嵌入一个自适应图中进行表示, 该模块有助于生成更准确的场景图。RAAL-SGG (Liu 等, 2022) 首次将区域感知注意力建模用于关系识别, 将注意力模块合并到每个图节点中, 关联检测到的对象与细粒度区

域, 从而对更细粒度的对象区域推理。Liu 等 (2023) 提出一种新的神经信念传播 (NBP) 范式, 定义一种新的评分函数, 能够获得更好的偏差一方差权衡, 在用于非常见关系时取得了更好的性能。

关联的图结构使得基于图网络的关联分析算法能够获得相应的注意力信息, 因此更适用于多层复杂关联分析及罕见关联识别任务。

3.3 评价指标

关联分析任务中, 主要使用召回率 Recall@K (R@K) 和平均召回率 mean Recall@K (mR@K) 作为评价指标。表 3 统计了上文提到的各类关联分析算法在两类评估指标上的性能。

召回率 Recall 是评估关联分析算法性能的常用指标, 该指标用来衡量目标识别真实存在的关联的能力。R@K 表示在前 K 个置信度最高的关联中, 识别正确的关联占有所有真实关联的比例, 因此该指标越高说明模型性能越好。其计算方式如下:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

式中, TP 表示真实关联中识别正确的关联数量, FP 表示真实关联中未能识别正确的关联数量。

由于关联数据集长尾效应明显, 召回率指标偏好于头部关联类别, 无法证明对尾部关联类别的识别性能。平均召回率很好地弥补了这一缺陷, 通过给予不同样本数的类别同等的关注, 更公正的评估算法性能。该指标针对每个关联类别独立计算召回率, 再对所有类别的召回率取平均值。其计算方式如下:

$$\text{m Recall} = \frac{1}{n} \sum_{i=1}^n \text{Recall}$$

式中, n 为关联类别数量, Recall 为该类别样本的召回率。

表 3 关联分析算法性能对比

Table 3 Performance comparison of relationship analysis method

算法	Predicate Det.		Phrase Det.		Relationship Det.	
	R@{20/50/100}	mR@{20/50/100}	R@{20/50/100}	mR@{20/50/100}	R@{20/50/100}	mR@{20/50/100}
条件随机场						
DR-Net (Dai 等, 2017)	-/80.78/81.90	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
JM-SGG (Xu 等, 2021)	-/14.9/18.6	-/-/-	5.1/6.2	-/-/-	2.9/3.6	-/-/-

续表

算法		Predicate Det.		Phrase Det.		Relationship Det.	
		R@{20/50/100}	mR@{20/50/100}	R@{20/50/100}	mR@{20/50/100}	R@{20/50/100}	mR@{20/50/100}
视觉 翻译 嵌入	VTransE (Zhang 等, 2017)	-/-/-	-/-/-	-2.65/3.51	-/-/-	1.71/2.14	-/-/-
	MATransE (Gkanatsios 等, 2019)	-89.79/96.26	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
	UVTransE (Hung 等, 2021)	-55.46/26.49	-/-/-	-30.01/36.18	-/-/-	-25.66/29.71	-/-/-
上下 文 信息	IMP (Xu 等, 2017)	-41.8/55.5	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
	MotifNet (Zellers 等, 2018)	58.5/65.2/67.1	-/-/-	32.9/35.8/36.5	-/-/-	21.4/27.2/30.3	-/-/-
	PANet (Chen 等, 2019)	59.7/66.0/67.9	-/-/-	37.4/40.9/41.8	-/-/-	21.5/26.9/29.9	-/-/-
	IS-GGT (Kundu 和 Aakur, 2023)	-/-/-	-26.4/31.9	-/-/-	-15.8/18.9	-/-/-	-9.1/11.3
	RelTR (Cong 等, 2023)	63.1/64.2/-	20.0/21.2/-	29.0/36.6/-	7.7/11.4/-	21.2/27.5/-	6.8/10.8/-
	SRSG (Lin 等, 2022)	0.1602/0.2561/-	-/-/-	-/-/-	-/-/-	-/-/-	-/-/-
	Chen 等 (Chen 等, 2022)	-26.67/32.92	-/-/-	-19.59/23.28	-/-/-	-/-/-	-/-/-
	HOD-Net (Li 等, 2025)	64.27/65.67 (MR@1500/2000)	38.73/39.70 (mMR@1500/2000)	46.28/47.55 (MR@1500/2000)	25.32/26.12 (mMR@1500/2000)	30.33/31.36 (MR@1500/2000)	14.09/14.78 (mMR@1500/2000)
	F-Net (Li 等, 2018)	-26.03/30.77	-/-/-	-18.32/21.20	-/-/-	-/-/-	-/-/-
图 网 络	Graph R-CNN (Yang 等, 2018)	-54.2/59.1	-/-/-	-29.6/31.6	-/-/-	-28.5/35.9	-/-/-
	ARN (Qi 等, 2019)	-56.5/61.3	-/-/-	-38.2/40.4	-/-/-	-/-/-	-/-/-
	RAAL-SGG (Liu 等, 2022)	59.1/66.2/68.4	-/-/-	33.5/36.7/37.6	-/-/-	21.7/27.3/29.9	-/-/-
	NBP (Liu 等, 2023)	-/-/-	-28.5/30.6	-/-/-	15.1/16.5	-/-/-	12.9/14.7

3.4 当前算法的局限性

随着深度学习的发展，目标感知技术已基本成熟，因而目标认知越来越受到重视，近年来，关联识别算法无论是类型还是数量都有了很大提升。目标关联分析算法大多将目标作为实体，将关联作为实体间的关系，学习实体和关系间的联系。虽然深度学习方法已经在关联识别领域取得了良好的效果，但总的来说仍存在一些局限，主要体现在以下几个方面：

(1) 简单依赖规则的算法在处理复杂的关系时往往表现不佳。现有在进行关系推断时，大多依赖于手动定义的简单规则或浅层语义关系，缺

乏对于对象之间复杂交互关系的建模能力。这导致在多维场景下，特别是涉及多个对象间复杂依赖关系的场景时，模型无法准确捕捉这些关系，进而影响整体性能。

(2) 基于静态图的算法难以适应动态多变场景。现有场景图生成算法大多使用固定的图结构，例如全连接图或链式结构，来表示对象之间的关系。这种静态图结构虽然可以在简单的场景中发挥作用，但在动态变化的复杂场景中表现不足。场景中对象的关系并不是固定的，可能随着时间、位置或事件发生动态变化，而固定的图结构无法灵活应对这些变化，导致算法在多变场景中的推断能力有限。

(3) 视觉数据集训练得到的算法无法直接迁移到遥感数据集中。现有的关联算法大多针对视觉数据集, 其平视视角的数据与遥感俯视视角数据在训练时存在巨大差异, 此外, 两类数据包含的目标类型、需要学习的特征不同, 因此现有基于视觉数据的算法直接迁移到遥感数据时往往无法取得好的效果。

4 探索性方向

基于面向遥感图像的关联关系的研究中的诸多问题, 本文提出了一系列针对未来发展的改进方案与展望, 这些方案旨在提升数据集的标注质量、工具的易用性以及算法的可视化表现, 推动遥感图像关联技术向更加智能化、高效化方向发展。

4.1 智能化标注与样本关系的可视化实现

在未来, 随着数据集规模的不断扩大和应用领域的多样化, 数据标注工具需要具备更高的灵活性、统一性与直观性, 才能应对复杂场景中对象和关系的标注需求。理想的标注工具应能够支持多种类型的样本框表示与多层次关联关系的标注, 同时具备自动化处理能力以提高效率, 还要能够通过直观的可视化方式展现复杂关系, 从而辅助标注人员进行有效的校验和优化。

本文探索性地开发了一套基于灵活性、统一性与直观性的图像标注与关系标注系统, 提出了一个基于对象—关系的统一标注框架, 定义了标

准的样本框表示方法和关系表示形式。样本框统一采用标准的四点坐标表示, 并包含类别标签、置信度等元数据。而样本之间的关系采用基于图论的形式表示, 每个对象被视为图中的节点, 关系为节点间的有向或无向边。这样, 标注结果具备了高度的结构化, 能够在不同领域的应用中轻松转换与使用。通过这样的框架, 不同数据集的标注格式可以互操作, 且系统具备自动生成和检查一致性的能力, 避免了因手工操作带来的误差和不一致性问题。

为了降低标注难度, 引入了自动化和交互式标注机制。在标注对象时, 系统结合了预训练模型与用户反馈的交互式方式, 自动生成初步的样本框, 并允许用户基于模型输出进行调整。自动标注与手动精细化标注相结合的方式, 极大地提升了大规模数据集的标注效率。此外, 我们还设计了一种动态标注机制, 允许用户在标注完成后, 随时通过系统界面对样本框或样本间关系进行修改。这种动态标注的灵活性打破了传统固化标注的局限, 使得数据集能够适应不断变化的需求。

我们在系统中集成了直观的样本框与关系的可视化处理模块, 基于图形化展示将抽象的样本间关系直观地呈现出来。通过不同的颜色、形状和线条, 系统能够清晰地展示样本之间的从属关系、邻近关系或其他复杂关系。与传统的数值化输出不同, 这种可视化方式能够让标注人员快速发现异常关系或标注错误, 从而大大提高了标注验证与调试的效率 (图5)。

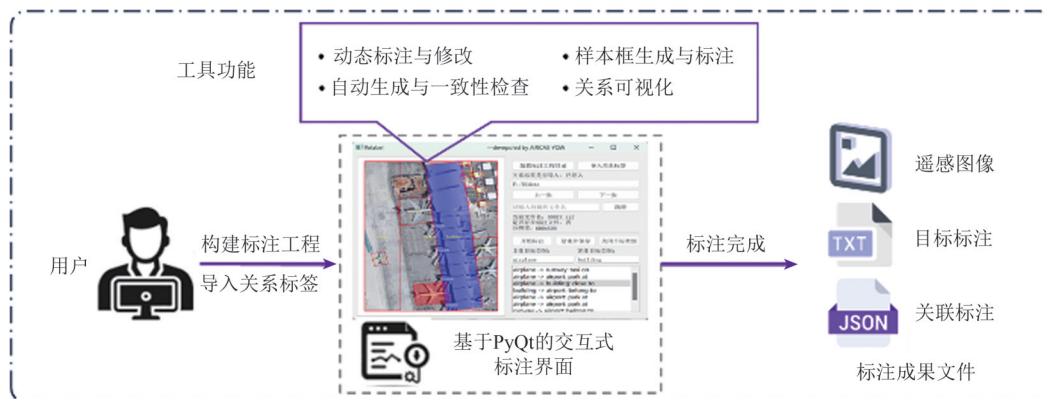


图5 关联标注工具流程图

Fig. 5 Flow chart of associated annotation tool

4.2 面向时序数据的遥感关联数据集

为了提升遥感图像数据集在关联关系分析上

的表现, 未来应创建一个“面向时序数据的遥感关联数据集”, 该数据集将着重于对时序数据的利

用,通过整合不同时刻的遥感影像,捕捉目标在不同时间点的动态变化,并对这些变化进行详细的时序关联标注。这将使其不仅能够反映空间上的关系,还能够描述时间维度上的动态交互。此外,该数据集还应进一步加强高层次语义关联的标注,涵盖的不仅仅是基础的空间关系,还应扩展到功能性关联(如目标之间的相互作用)和时间关联(如目标随时间的变化与其他对象的关系),以更全面、准确地表征复杂场景中的多维度关系。

其次,为了进一步提升遥感关联数据集的标注精度与解析能力,该数据集构建可采用层级式标注方法,在单一目标标注的基础上,对复合型目标进行更高层次的标注(例如:在标注单个车辆的同时,还对车队进行整体标注),捕捉个体与整体的空间和功能关联,提供更多层次的信息。这种层级化的标注方式使得模型能够同时学习个体和整体的关系结构,有助于提升对复合型目标的解译能力。同时针对复杂目标群体,该数据集还可引入语义聚合标注策略,不仅对每个个体进行标注,还对其在复合群体中的功能或角色进行描述(例如:在标注车队时,不仅要标注每辆车的空间位置,还应区分车队中的不同功能性角色,如指挥车、补给车等)。这种细化的标注方式能够捕捉到复杂群体行为中的语义关联,有助于增强模型对多目标复杂关系的理解与推理能力。

最后,该数据集应进一步向大尺度、多分辨率的构建方向发展。通过整合不同分辨率的影像数据,使得模型可以从全局和局部两个层次进行解译,既能够把握整体场景的宏观布局,又能在细节层面精确解析个体目标。通过卫星影像拼接或多源数据融合技术,在不损失细节的前提下生成大范围的影像数据,从而支持模型对复杂场景进行全面而细致的解译。

4.3 动态分层关系推理算法

针对于当前关联关系算法存在的不足,未来的算法应该向动态分层的方向发展,首先,为了提升算法对复杂关系的处理能力,在未来可以采用分层关系建模的方法,将样本间的关系分为多层次进行处理。首先对样本之间的空间关系进行建模,然后进一步推断其功能依赖、因果关系等高层次关系。通过将复杂关系逐层解耦,模型可

以更有效地处理每个层次的关系,从而提升对复杂语境中关系的整体理解。

其次,关联关系生成模型可以引入动态图结构。动态图不仅能够随场景的变化自动调整节点和边的连接方式,还可以通过实时更新关系来更准确地反映对象间的关联性。这种图结构能够在复杂场景中表现得更加灵活,使得算法在动态环境下具备更强的适应能力。同时结合时序数据与动态图结构,未来可以通过追踪对象在不同时刻的状态变化,来推断其间的关系演变。这不仅有助于提升模型的时序推理能力,还能够更好地应对动态场景下的关系解译任务。

最后,为了减轻长尾效应的影响,未来可以通过引入强化学习或自适应采样策略来应对长尾效应。针对低频类别,采用动态加权的方式增强其在模型中的重要性,确保模型能够在训练时更充分地学习这些稀有类别的特征。同时,通过结合迁移学习和元学习,将高频类别中学到的特征迁移到低频类别上,提升稀有类别的识别能力,帮助模型更快地适应新的、未见过的类别,增强在低频数据上的表现。

5 结 语

关联是目标感知向目标认知转变中至关重要的一个步骤,本文整理了面向视觉场景的关联数据集、面向遥感场景的关联数据集、基于关联数据集提出的相关算法的研究现状,形成该领域最新进展综述。在此基础上,通过分析当前数据集、算法上存在的局限性,提出面向时序数据的遥感关联数据集的设想,以及算法提升方面可能存在的探索性思路,并探索性地构建了一套完整的关联标注与可视化工具,为研究人员后续研究提供便利。后续工作将在面向遥感的关联数据集构建、目标关联分析算法方面进行针对性突破,同时探索时序数据下关联的演化,实现对目标的动态认知。

参考文献(References)

- Belz A, Muscat A, Anguill P, Sow M, Vincent G and Zinessabah Y. 2018. SpatialVOC2K: a multilingual dataset of images with annotations and features for spatial relations between objects//Proceedings of the 11th International Conference on Natural Language

- Generation. Tilburg: ACL: 140-145 [DOI: 10.18653/V1/W18-6516]
- Chang X J, Ren P Z, Xu P F, Li Z H, Chen X J and Hauptmann A. 2023. A comprehensive survey of scene graphs: generation and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 1-26 [DOI: 10.1109/TPAMI.2021.3137605]
- Chen J, Zhou X, Zhang Y, Sun G, Deng M and Li H F. 2022. Message-passing-driven triplet representation for geo-object relational inference in HRSI. *IEEE Geoscience and Remote Sensing Letters*, 19: 8003205 [DOI: 10.1109/LGRS.2020.3038569]
- Chen Y N, Wang Y J, Zhang Y and Guo Y W. 2019. PANet: a context based predicate association network for scene graph generation//2019 IEEE International Conference on Multimedia and Expo (ICME). Shanghai: IEEE: 508-513 [DOI: 10.1109/ICME.2019.00094]
- Cong Y R, Yang M Y and Rosenhahn B. 2023. RelTR: relation transformer for scene graph generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9): 11169-11183 [DOI: 10.1109/TPAMI.2023.3268066]
- Dai B, Zhang Y Q and Lin D H. 2017. Detecting visual relationships with deep relational networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 3298-3308 [DOI: 10.1109/CVPR.2017.352]
- Fu K, Lu W X, Liu X Y, Deng C B, Yu H F and Sun X. 2024. A comprehensive survey and assumption of remote sensing foundation modal. *National Remote Sensing Bulletin*, 28(7): 1667-1680 (付琨, 卢宛萱, 刘小煜, 邓楚博, 于泓峰, 孙显. 2024. 遥感基础模型发展综述与未来设想. *遥感学报*, 28(7): 1667-1680) [DOI: 10.11834/jrs.20233313]
- Gkanatsios N, Pitsikalis V, Koutras P, Zlatintsi A and Maragos P. 2019. Deeply supervised multimodal attentional translation embeddings for visual relationship detection//2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE: 1840-1844 [DOI: 10.1109/ICIP.2019.8803106]
- Hong D F, Zhang B, Li H, Li Y X, Yao J, Li C Y, Werner M, Chanussot J, Zipf A and Zhu X X. 2023. Cross-city matters: a multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks. *Remote Sensing of Environment*, 299: 113856 [DOI: 10.1016/j.rse.2023.113856]
- Huang B Z, Ju J Y, Li Z H and Wang Y G. 2023. Reconstructing groups of people with hypergraph relational reasoning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE: 14827-14837 [DOI: 10.1109/ICCV51070.2023.01366]
- Hung Z S, Mallya A and Lazebnik S. 2021. Contextual translation embedding for visual relationship detection and scene graph generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11): 3820-3832 [DOI: 10.1109/TPAMI.2020.2992222]
- Ji J W, Krishna R, Li F F and Niebles J C. 2020. Action genome: actions as compositions of spatio-temporal scene graphs//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 10233-10244 [DOI: 10.1109/CVPR42600.2020.01025]
- Johnson J, Krishna R, Stark M, Li L J, Shamma D A, Bernstein M S and Li F F. 2015. Image retrieval using scene graphs//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 3668-3678 [DOI: 10.1109/CVPR.2015.7298990]
- Krishna R, Zhu Y K, Groth O, Johnson J, Hata K, Kravitz J, Chen S, Kalantidis Y, Li L J, Shamma D A, Bernstein M S and Li F F. 2017. Visual genome: connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123(1): 32-73 [DOI: 10.1007/s11263-016-0981-7]
- Kundu S and Aakur S N. 2023. IS-GGT: iterative scene graph generation with generative transformers//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE: 6292-6301 [DOI: 10.1109/CVPR52729.2023.00609]
- Kuznetsova A, Rom H, Alldrin N, Uijlings J, Krasin I, Pont-Tuset J, Kamali S, Popov S, Mallocci M, Kolesnikov A, Duerig T and Ferrari V. 2020. The open images dataset V4: unified image classification, object detection, and visual relationship detection at scale. *International Journal of Computer Vision*, 128(7): 1956-1981 [DOI: 10.1007/S11263-020-01316-Z]
- Lafferty J, McCallum A, Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]//Icml. 2001, 1(2): 3.
- Lample G, Ballesteros M, Subramanian S, Kawakami K and Dyer C. 2016. Neural architectures for named entity recognition//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: ACL: 260-270 [DOI: 10.18653/v1/N16-1030]
- Li F M, Zhang L, Bandukda M, Stangl A, Shinohara K, Findlater L and Carrington P. 2023b. Understanding visual arts experiences of blind people//Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. Hamburg: ACM: 1-21 [DOI: 10.1145/3544548.3580941]
- Li P, Zhang D Z, Wulamu A, Liu X and Chen P. 2021. Semantic relation model and dataset for remote sensing scene understanding. *ISPRS International Journal of Geo-Information*, 10(7): 488 [DOI: 10.3390/ijgi10070488]
- Li Y K, Ouyang W L, Zhou B L, Shi J P, Zhang C and Wang X G. 2018. Factorizable net: an efficient subgraph-based framework for scene graph generation//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer: 346-363 [DOI: 10.

- 1007/978-3-030-01246-5_21]
- Li Y S, Wang L L, Wang T Z, Yang X, Luo J W, Wang Q, Deng Y M, Wang W B, Sun X, Li H F, Dang B, Zhang Y J, Yu Y and Yan J C. 2025. STAR: a first-ever dataset and a large-scale benchmark for scene graph generation in large-size satellite imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(3): 1832-1849 [DOI: 10.1109/TPAMI.2024.3508072]
- Li Y X, Hou Q B, Zheng Z H, Cheng M M, Yang J and Li X. 2023a. Large selective kernel network for remote sensing object detection//*Proceedings of 2013 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE: 16748-16759 [DOI: 10.1109/ICCV51070.2023.01540]
- Liang X D, Shen X H, Feng J S, Lin L and Yan S C. 2016. Semantic object parsing with graph LSTM//*14th European Conference on Computer Vision*. Amsterdam: Springer: 125-143 [DOI: 10.1007/978-3-319-46448-0_8]
- Liang Y Z, Bai Y L, Zhang W, Qian X M, Zhu L and Mei T. 2019. VrR-VG: refocusing visually-relevant relationships//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE: 10402-10411 [DOI: 10.1109/ICCV.2019.01050]
- Lin Z Y, Zhu F, Kong Y Z, Wang Q and Wang J Y. 2022. SRSG and S2SG: a model and a dataset for scene graph generation of remote sensing images from segmentation results. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 4707411 [DOI: 10.1109/TGRS.2022.3185678]
- Liu A A, Tian H S, Xu N, Nie W Z, Zhang Y D and Kankanhalli M. 2022. Toward region-aware attention learning for scene graph generation. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12): 7655-7666 [DOI: 10.1109/TNNLS.2021.3086066]
- Liu D Q, Bober M and Kittler J. 2023. Neural belief propagation for scene graph generation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 10161-10172 [DOI: 10.1109/TPAMI.2023.3243306]
- Lu C W, Krishna R, Bernstein M and Li F F. 2016. Visual relationship detection with language priors//*14th European Conference on Computer Vision*. Amsterdam: Springer: 852-869 [DOI: 10.1007/978-3-319-46448-0_51]
- Ma X P, Zhang X K and Pun M O. 2024. RS³amba: visual state space model for remote sensing image semantic segmentation. *IEEE Geoscience and Remote Sensing Letters*, 21: 6011405 [DOI: 10.1109/LGRS.2024.3414293]
- Qi M S, Li W J, Yang Z Y, Wang Y H and Luo J B. 2019. Attentive relational networks for mapping images to scene graphs//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE: 3952-3961 [DOI: 10.1109/CVPR.2019.00408]
- Qiu Y, Nagasaki Y, Hara K, Kataoka H, Suzuki R, Iwata K and Satoh Y. 2023. VirtualHome action genome: a simulated spatio-temporal scene graph dataset with consistent relationship labels//*Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa: IEEE: 3340-3349 [DOI: 10.1109/WACV56688.2023.00335]
- Rai N, Chen H F, Ji J W, Desai R, Kozuka K, Ishizaka S, Adeli E and Niebles J C. 2021. Home action genome: cooperative compositional action understanding//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE: 11179-11188 [DOI: 10.1109/CVPR46437.2021.01103]
- Sadeghi M A and Farhadi A. 2012. Recognition using visual phrases//*24th IEEE Conference on Computer Vision and Pattern Recognition*. Colorado Springs: IEEE: 1745-1752 [DOI: 10.1109/CVPR.2011.5995711]
- Shang X D, Di D L, Xiao J B, Cao Y, Yang X and Chua T S. 2019. Annotating objects and relations in user-generated videos//*Proceedings of the 2019 on International Conference on Multimedia Retrieval*. Ottawa: ACM: 279-287 [DOI: 10.1145/3323873.3325056]
- Shang X D, Ren T W, Guo J F, Zhang H W and Chua T S. 2017. Video visual relation detection//*Proceedings of the 25th ACM International Conference on Multimedia*. Mountain: ACM: 1300-1308 [DOI: 10.1145/3123266.3123380]
- Sherstinsky A. 2020. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Non-linear Phenomena*, 404: 132306 [DOI: 10.1016/j.physd.2019.132306]
- Shoouri S, Yang M Y, Fan Z C and Kim H S. 2023. Efficient computation sharing for multi-task visual scene understanding//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE: 17084-17095 [DOI: 10.1109/ICCV51070.2023.01571]
- Sun X, Wang P J, Lu W X, Zhu Z C, Lu X N, He Q B, Li J X, Rong X E, Yang Z J, Chang H, He Q L, Yang G, Wang R P, Lu J W and Fu K. 2023. RingMo: a remote sensing foundation model with masked image modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5612822 [DOI: 10.1109/TGRS.2022.3194732]
- Wang B, Deng C B, Liu R. 2023a. Generation Mechanism of Battlefield "Potential" Oriented to Cognition. *Journal of Command and Control*, 9(05):606-611 (王博,邓楚博,刘然,等.面向认知的战场“势”生成机理.指挥与控制学报,2023,9(05):606-611)
- Wang J J, Li W, Wang Y J, Tao R and Du Q. 2024a. Representation-enhanced status replay network for multisource remote-sensing image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11): 15346-15358 [DOI: 10.1109/TNNLS.2023.3286422]
- Wang J J, Li W, Zhang M M, Tao R and Chanussot J. 2023b. Remote-sensing scene classification via multistage self-guided separation network. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5615312 [DOI: 10.1109/TGRS.2023.3295797]

- Wang Y, Yang L, Liu X Z, and Yan P F. 2024b. An improved semantic segmentation algorithm for high-resolution remote sensing images based on DeepLabv3+. *Scientific Reports*, 14(1): 9716 [DOI: 10.1038/s41598-024-60375-1]
- Wu Z H, Pan S R, Chen F W, Long G D, Zhang C Q and Yu P S. 2021. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1): 4-24 [DOI: 10.1109/TNNLS.2020.2978386]
- Xu D F, Zhu Y K, Choy C B and Li F F. 2017. Scene graph generation by iterative message passing//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 3097-3106 [DOI: 10.1109/CVPR.2017.330]
- Xu D Q and Wu Y Q. 2024. Progress of research on deep learning algorithms for object detection in optical remote sensing images. *National Remote Sensing Bulletin*, 28(12): 3045-3073 (徐丹青, 吴—全. 2024. 光学遥感图像目标检测的深度学习算法研究进展. *遥感学报*, 28(12): 3045-3073) [DOI: 10.11834/jrs.20243166]
- Xu M H, Qu M, Ni B B and Tang J. 2021. Joint modeling of visual objects and relations for scene graph generation//*Proceedings of the 35th International Conference on Neural Information Processing Systems*. [s.l.]: Curran Associates Inc.: 589
- Yang J K, Ang Y Z, Guo Z J, Zhou K Y, Zhang W and Liu Z W. 2022. Panoptic scene graph generation//*17th European Conference on Computer Vision*. Tel Aviv: Springer: 178-196 [DOI: 10.1007/978-3-031-19812-0_11]
- Yang J W, Lu J S, Lee S, Batra D and Parikh D. 2018. Graph R-CNN for scene graph generation//*Proceedings of the 15th European Conference on Computer Vision*. Munich: Springer: 690-706 [DOI: 10.1007/978-3-030-01246-5_41]
- Yang K Y, Russakovsky O and Deng J. 2019. SpatialSense: an adversarially crowdsourced benchmark for spatial relation recognition//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE: 2051-2060 [DOI: 10.1109/ICCV.2019.00214]
- Zellers R, Yatskar M, Thomson S and Choi Y. 2018. Neural motifs: scene graph parsing with global context//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE: 5831-5840 [DOI: 10.1109/CVPR.2018.00611]
- Zhang C, Su J R, Ju Y K, Lam K M, and Wang Q. 2023. Efficient inductive vision transformer for oriented object detection in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5616320 [DOI: 10.1109/TGRS.2023.3292418]
- Zhang H W, Kyaw Z, Chang S F and Chua T S. 2017. Visual translation embedding network for visual relation detection//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE: 3107-3115 [DOI: 10.1109/CVPR.2017.331]
- Zheng S, Jayasumana S, Romera-Paredes B, Vineet V, Su Z Z, Du D L, Huang C and Torr P H S. 2015. Conditional random fields as recurrent neural networks//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE: 1529-1537 [DOI: 10.1109/ICCV.2015.179]
- Zhuo T, Cheng Z Y, Zhang P, Wong Y and Kankanhalli M. 2019. Explainable video action reasoning via prior knowledge and state transitions//*Proceedings of the 27th ACM International Conference on Multimedia*. Nice: ACM: 521-529 [DOI: 10.1145/3343031.3351040]

Overview and prospect of relationship dataset for visual and remote sensing images

LI Jingzhu¹, LIU Xiaoyu², HAO Lingxiang², WU Guoxuan², SUN Zheng², GONG Min², LI Ya²

1.School of Integrated Circuits, University of the Chinese Academy of Sciences, Beijing 101408, China;

2.Aerospace Information Innovation Institute of the Chinese Academy of Sciences, Beijing 100094, China

Abstract: Target relationship analysis aims to gain a deep understanding of the target based on the detection task. In recent years, target relationship analysis based on computer vision has attracted the attention of researchers in related fields, whereas the relationship analysis algorithm in remote sensing emerged later. A primary concern is the lack of relationship datasets for the remote sensing field. Therefore, this article aims to fill the gap in the review of association datasets, conduct a detailed review and analysis of relationship datasets and target relationship analysis algorithms in the fields of computer vision and remote sensing, and provide new ideas for the research and development of target relationship analysis in the remote sensing field.

In terms of relationship datasets, this article reviews and outlines 15 published image and video visual association datasets and four remote sensing image association datasets in the fields of computer vision and remote sensing. It systematically introduces their characteristics, data scale, association categories, and number of triples, as well as summarizes the limitations of existing remote sensing relationship datasets such as inconsistent annotation formats and obvious long-tailed small groups. In terms of target relationship analysis

algorithms, this article divides the existing algorithms into relationship analysis algorithms based on conditional random fields and relationship analysis algorithms based on deep learning. Moreover, it summarizes the advantages and disadvantages of each type of algorithm and compares the performance of the most advanced algorithms based on the evaluation metric commonly used in target relationship analysis tasks. Finally, the factors that improve the performance of current remote sensing target relationship analysis algorithms are summarized, such as complex semantic relationships and dynamic time-varying scenes.

In view of the limitations of current relationship datasets and target relationship analysis algorithms, this article proposes possible future research directions. Regarding association datasets, this article proposes a concept of building a remote sensing relationship dataset for time-series data and annotation specifications for remote sensing target relationships. Moreover, it preliminarily develops a set of association annotation tools for remote sensing images, which provides optimization suggestions for the construction direction of future relationship datasets. Regarding target relationship analysis algorithms, this article proposes innovative methods such as association dynamic hierarchical modeling and dynamic scene graphs, to provide ideas for researchers in related fields.

Target relationship analysis is an important step to move from target perception to cognition. However, research on target relationship analysis in the field of remote sensing is still in its infancy, and there is a lack of discussion of remote sensing target relationship datasets. To address this problem, this article systematically organizes the existing relationship datasets in the current computer vision and remote sensing fields and summarizes the limitations of remote sensing relationship datasets. This article also summarizes the research status of target association algorithms and demonstrates the advantages and disadvantages of existing algorithms according to the algorithm type. Finally, this article proposes a series of possible research ideas for the development of target relationship analysis in the field of remote sensing. In terms of sample accumulation, this article proposes the idea of constructing a remote sensing time-series relationship dataset. To unify the annotation format, a set of sample annotation and visualization tools is developed. In terms of algorithm innovation, this article proposes the idea of a dynamic hierarchical relationship reasoning method, which provides new ideas for the future development of target relationship analysis algorithms.

Key words: remote sensing, target relationship analysis, relationship datasets, scene graph generation, review, dataset

Supported by National Key Research and Development Program of China (No. 2021YFB3900504)