

遥感影像深度学习配准方法综述

李星华¹, 艾文浩¹, 冯蕊涛^{1,2}, 罗少杰³

1. 武汉大学 遥感信息工程学院, 武汉 430079;

2. 陕西师范大学 地理科学与旅游学院, 西安 710119;

3. 国网浙江省电力有限公司杭州供电公司, 杭州 310020

摘要: 遥感影像配准是指通过几何变换使两景或多景影像空间位置对齐的过程, 是影像融合、变化检测、农业监测等应用的重要预处理步骤。近年来, 深度学习引起了人们的广泛关注, 并在遥感影像配准中成功应用。本文在简要介绍传统遥感影像配准方法的基础上, 重点分析了深度学习在基于区域的配准方法、基于特征的配准方法两方面取得的重要进展, 分享了用于遥感影像配准的公开数据集, 并总结了深度学习在遥感影像配准中的机遇与挑战。

关键词: 深度学习, 影像配准, 基于区域, 基于特征, 配准数据集

中图分类号: P2

引用格式: 李星华, 艾文浩, 冯蕊涛, 罗少杰. 2023. 遥感影像深度学习配准方法综述. 遥感学报, 27(2): 267-284

Li X H, Ai W H, Feng R T and Luo S J. 2023. Survey of remote sensing image registration based on deep learning. National Remote Sensing Bulletin, 27(2): 267-284 [DOI: 10.11834/jrs.20235012]

1 引言

作为重要的空间信息载体, 遥感影像在灾害防控、环境监测、资源勘探等众多领域都起着关键作用。许多遥感应用, 如变化检测 (Girard 等, 2019)、影像融合 (Seo 等, 2020)、影像拼接 (Li 等, 2019) 等, 需要多景遥感影像进行协同分析, 都离不开同名像元的空间对齐。遥感影像配准是指将具有重叠区域的影像统一到相同坐标系中, 使影像同名点在空间位置上达到最佳对齐。一方面, 受到传感器类型、成像角度、成像时间等因素的影响, 仅仅依靠影像地理参考信息无法实现影像高精度配准; 另一方面, 影像配准是定量处理与分析地表信息的基础, 但目前缺少具有泛适用性的配准算法。因此, 研究高精度、高效率、高鲁棒性的自动配准算法具有十分重要的意义。

迄今为止, 已发展出大量遥感影像配准算法。根据配准流程不同, 大多数配准算法可以划分为两个类别: 基于区域的配准方法与基于特征的配

准方法 (Zeng 等, 2021; Feng 等, 2021)。前者 (Chen 等, 2003; Suri 和 Reinartz, 2010; Liang 等, 2014) 基于预先建立的相似性测度与转换模型, 根据影像灰度或相位等浅层信息计算影像的相似性, 再通过某种优化策略调整转换模型参数, 使相似性测度达到理想阈值, 最终实现影像高精度对齐。后者 (王志衡和吴福朝, 2009; 梁艳 等, 2014; 马方龙, 2018) 则首先提取、匹配影像点、线、面等几何特征, 利用几何特征的空间对应关系解算影像转换模型参数。

综合来看, 基于区域的配准方法易于理解, 不需要将影像浅层信息映射到高维特征, 通过相似性测度建立浅层信息与转换模型的直接映射关系, 不存在误差累计的问题, 且保留了原始影像细节信息, 可以实现亚像元级高精度对齐。然而, 基于区域的配准方法存在两个明显问题: (1) 配准精度严重依赖于相似性测度 (冯蕊涛, 2020; Jiang 等, 2021a)。互相关系数 CC (Cross Correlation) 与互信息 MI (Mutual Information) (Johnson 等,

收稿日期: 2021-05-31; 预印本: 2021-09-30

基金项目: 国网科技项目 (编号: 5700-202019162A-0-0-00)

第一作者简介: 李星华, 研究方向为多时相遥感。E-mail: lixinghua5540@whu.edu.cn

通信作者简介: 冯蕊涛, 研究方向为遥感影像几何处理。E-mail: ruitaofeng@whu.edu.cn

2001) 是两种较为常用的相似性测度 (Ma 等, 2021), 但基于统计分析的评价指标无法顾及像元之间的空间位置关系, 当影像间存在较大的几何形变或无明显特征时, 此类方法难以实现高精度配准。(2) 配准效率低, 实时性差, 内存开支大。一方面, 影像浅层信息冗余问题严重; 另一方面, 相较于自然影像、医学影像等, 遥感影像空间范围更大, 所以基于区域的遥感影像配准方法往往存在实时性差、内存开支大等问题。此外, 转换模型初值与优化策略也严重影响此类方法的配准效率。基于特征的配准方法则利用影像中稀疏、显著的空间几何特征, 对原始灰度影像进行抽象表达, 有效抑制了信息冗余问题, 能够实现高效配准。整体来看, 根据几何特征的映射关系解算转换模型, 适用于大几何形变影像的配准任务, 但其也存在误差累积、配准精度较差等问题。(1) 误差累积。基于特征的配准方法是多阶段配准模型, 前一阶段的误差直接影响后续阶段的精度。(2) 配准精度较差。一方面, 此类方法难以避免误差累积问题; 另一方面, 遥感影像不同区域的几何形变可能并不一致, 基于仿射变换等全局或局部转换模型, 难以准确描述所有像元的空间对应关系。例如, 当影像覆盖区域为山区等地形变化剧烈的区域, 基于特征的配准方法难以实现高精度配准 (冯蕊涛等, 2021)。为了解决上述两类方法存在的问题, 有学者 (Gong 等, 2014; Ye 和 Shan, 2014; 赵辽英等, 2015; Feng 等, 2019) 提出将二者集成在一起, 基于“由粗到精”的配准策略, 进一步提升配准精度。但是基于混合模型的配准方法也存在算法复杂度高、计算效率低等问题 (冯蕊涛, 2020)。

概括来说, 传统配准方法存在两个主要问题: (1) 算法适用性差。(2) 无法充分利用影像的深层语义信息。近些年来, 深度学习在变化检测 (Li 等, 2021; 袁洲等, 2021)、地物分类 (Niu 等, 2020) 等遥感应用领域取得了众多成果, 其在提取影像深层特征上的优越表现再一次引起了人们的广泛关注。一般认为, 基于数据驱动的学习能够通过多层次的非线性运算, 充分提取输入数据的突出、具有代表性的抽象特征 (Schmidhuber, 2015; LeCun 等, 2015; Lee 和 Oh, 2021), 具有良好的适用性。理论上, 深度学习的深层特征提取与表达能力可以弥补浅层特征不稳定的缺陷,

提升配准算法的鲁棒性 (Ye 等, 2018)。

近 5 年来, 已有许多研究尝试将深度学习应用于遥感、自然、医学影像配准等领域。根据其应用形式, 总体上可以将其划分为两个类别: (1) 利用深度学习构建相似性测度, 并通过强化学习的方式优化转换模型; 或直接利用端到端网络解算影像位移场。(2) 将深度学习嵌入到传统基于特征的配准框架内, 使之替代传统特征提取、描述与匹配方法。与传统配准方法相比, 通过对大量样本的学习, 基于深度学习的配准方法能够提取影像的深层语义信息、学习影像间复杂的空间映射关系, 并实现遥感影像的高效率、高精度配准。

本文在简要回顾传统遥感影像配准方法的基础上, 对两类传统方法的优缺点做了详尽的总结。在此基础上, 本文重点论述了深度学习在基于区域的配准方法与基于特征的配准方法的应用, 并详细分析了其优缺点。此外, 本文分享了多个开源影像配准数据集。最后, 本文对深度学习在遥感影像配准领域的机遇与挑战进行了展望。

2 遥感影像配准中常用网络结构

相较于早期的浅层神经网络, 深层神经网络具有更多的隐藏层。一般认为, 网络隐藏层越多, 其对复杂函数的拟合效果越好, 即深层神经网络能够学习到数据更为本质的特征。随着深度学习技术的不断发展, 越来越多神经网络被应用于遥感影像处理领域 (Ma 等, 2019a)。其中, 卷积神经网络 CNN (Convolutional Neural Networks) (LeCun 等, 1998)、全卷积神经网络 FCN (Fully Convolutional Networks) (Shelhamer 等, 2017)、孪生神经网络 (Siamese Networks) (Chopra 等, 2005) 是遥感影像配准中较为常用的网络结构。

CNN 由卷积层、池化层与全连接层构成。卷积层负责提取影像局部特征, 池化层通过对提取特征进行下采样, 以实现减少数据量, 抑制模型过拟合的目的。CNN 通过多层次的“卷积—池化”操作提取影像深层语义特征, 再由网络末端的全连接层对提取到的局部特征进行整合并输出分类结果。由于卷积操作得到的特征仅仅与影像局部区域有关, 所以 CNN 提取到的特征通常具有强鲁棒性。

FCN 在 CNN 的基础上, 进行了如下改动: (1) 将全连接层替换为卷积层, 使网络的输入数据尺寸

不受约束；(2) 引入转置卷积层，可以对压缩后的特征图进行上采样，增大数据尺寸，以实现对所有像元的分类。在影像配准中，FCN 常用于特征提取及位移场估计等领域。

Siamese Networks 是一种双分支神经网络框架，由两个结构一致且权值共享的网络构成，通常用来衡量两个输入数据之间的相似程度。与其类似的还有伪孪生网络 (Pseudo-Siamese Networks)。伪孪生网络由两个不同结构的网络拼接而成，这使得其具有更广泛的适用性，但也存在需要大量训练样本的问题。

除上述网络结构外，也有学者将深度信念网络 DBN (Deep Belief Networks) (Hinton 和 Salakhudinov, 2006)、生成对抗网络 GAN (Generative Adversarial

Nets) (Goodfellow 等, 2014) 等网络应用于遥感影像配准领域，并取得了十分优秀的成果。

3 基于区域的配准方法

传统基于区域的配准方法一般是将配准问题视为一个优化问题：利用影像灰度、相位信息，根据预先建立的相似性测度，调整优化转换模型参数，以实现影像配准。如何构建稳健的相似性测度是此类研究的关键问题之一 (Simonovsky 等, 2016; Kuppala 等, 2020)。根据是否显式计算影像相似性，基于深度学习的区域配准方法可以进一步划分出两个研究方向：(1) 基于深度学习的相似性估计；(2) 基于深度学习的转换模型估计。

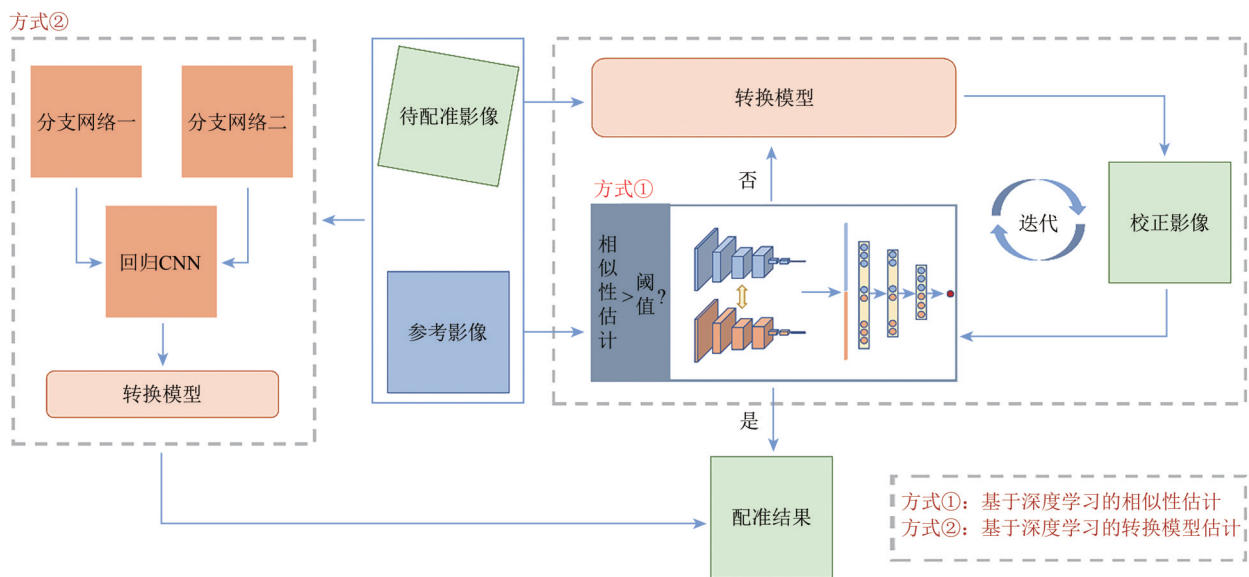


图1 基于深度学习的区域配准方法

Fig. 1 Area-based registration methods based on deep learning

基于深度学习的相似性估计。传统方法中较常用的相似性测度如CC、MI等，存在对灰度变化敏感、无法考虑影像结构要素、与配准精度并非绝对线性相关等问题 (Feng, 2020; Jiang 等, 2021a)。而深度学习具有完全数据驱动、能提取影像深层语义特征等明显优势，所以利用深度学习估计影像相似性以代替人工设计的相似性测度，是此类方法的重要研究方向之一。Simonovsky 等 (2016) 利用少量标签数据对CNN进行监督训练，以解决多模态3D影像配准问题。实验证明，对于几何变化情况较为简单的数据，仅需要利用少量

训练样本即可完成训练。Cheng 等 (2018) 将相似性估计视为两阶段任务：特征描述与相似性估计，并提出利用多层降噪自编码器对DNN (Deep Neural Networks) 进行预训练，以解决训练样本不足的问题。此外，Haskins 等 (2019) 也将CNN应用于多模态相似性估计领域。实验证明，相对于传统方法，基于学习的相似性指标具有更好的鲁棒性，能广泛适用于不同数据。

基于深度学习的相似性估计并未改变传统基于区域的配准方法流程，而是将深层神经网络嵌入其配准框架内，利用深度学习的完全数据驱动

与深层特征提取优势,以达到增强算法稳健性与适用性的目的。值得注意的是,在计算影像相似性的前提下,也有学者尝试利用强化学习的方式

对传统方法的优化策略进行改进(Liao等,2016;姚明青和胡靖,2020),进一步提升了配准精度与算法稳健性。

表1 基于深度学习的相似性估计

Table 1 Similarity estimation based on deep learning

参考文献	网络	训练方式	影像数据	针对问题	主要贡献/结论
Simonovsky等,2016	双通道CNN	监督	多模态医学影像	相似性估计	适用于不同模态影像;效果优于基于互信息的配准方法
Cheng等,2018	DNN	监督	多模态医学影像	相似性估计	精度优于基于NCC与局部互相关指数等传统方法
Haskins等,2019	双通道CNN	监督	3D MR-TRUS	相似性估计	基于CNN的相似性指标具有泛适用性;可应用于传统基于区域的配准框架;对转换模型初始值具有强鲁棒性
Liao等,2016	CNN	监督	多模态医学影像	相似性估计 优化策略	精度高,当影像差异很小时,甚至优于人工手动调整方法;具有泛适用性
姚明青和胡靖,2020	CNN LSTM	监督	多模态医学影像	相似性估计 优化策略	由策略网络与价值网络两部分组成模型;能够配准大几何形变影像

基于深度学习的转换模型估计。此类方法将转换模型计算视为回归问题(Jiang等,2021a),基于端到端网络直接建立影像与单应性矩阵参数或位移场模型的映射关系。一方面,此类方法无需显式计算影像相似性,避免了传统方法中的迭代优化流程,提升了配准效率与鲁棒性;另一方面,当影像存在严重局部几何畸变时,仿射变换与多项式模型无法准确配准每个像元,而位移场模型可以直接刻画错位影像对应像元之间的空间对应关系,理论上可以实现更高的配准精度。

针对传统特征提取方法在遥感影像中定位精度较差的问题,GMatchNet(Lee等,2021)利用CNN建立输入影像对与仿射变换参数的直接映射关系,并对浮动影像块进行几何校正以提高特征点定位精度。Zampieri等(2018)与Girard等(2019)利用级联FCN输出影像位移场,实现光学影像与地图数据的高精度对齐。考虑到医学影像配准中标签样本难以获取的问题,DLIR(Deep Learning Image Registrtion)(de Vos等,2019)提出了一种非监督训练框架,利用CNN估计医学影像仿射变换参数,在配准精度上与传统方法接近的前提下,有效提升了配准效率。此外,也有学者通过非监督或监督方式训练GAN与CNN,直接输出影像位移场以实现影像配准(Yang等,2021;Luo等,2021;Ni等,2021)。

密集光流估计是一类特殊的位移场估计方法。传统基于密集光流的配准方法是依据影像灰度信息,引入某种约束条件,建立不同时相影像的能量损失函数,并根据该函数解算影像位移场。与仿射变换、多项式模型等全局模型相比,密集

光流估计能够处理复杂的空间映射关系。但由于此类方法计算难度大、计算效率低、前提条件苛刻,所以多用于医学影像及短时间跨度的自然影像配准,在遥感影像配准领域应用较少。近些年来,越来越多高空间分辨率遥感影像被投入使用,影像局部几何畸变问题也越来越受到重视,基于深度学习的密集光流估计为上述问题提供了新的研究思路。

传统光流估计通常基于两个基本假设(冯蕊涛等,2021):(1)亮度恒定:不同时相影像对应目标的亮度值保持不变;(2)小运动:不同时相影像的空间错位较小。但在实际配准应用中,往往难以满足上述假设。针对第二条假设,SPyNet(Spatial Pyramid Network)(Ranjan和Black,2017)将影像金字塔与CNN相结合以估计影像光流。其核心思想是从高至低估计影像金字塔光流,并通过多次坐标变换,实现“由粗到精”的光流估计。LiteFlowNet(Hui等,2018)也是基于“由粗到精”的策略,通过级联网络实现密集光流估计。LiteFlowNet2(Hui等,2021)与LiteFlowNet3(Hui和Loy,2020)则在LiteFlowNet的基础上提高了光流估计精度与效率。LiteFlowNet2对计算成本与精度提升效果进行分析、取舍,减少了原影像金字塔的层数,且通过在NetE中引入一个简单的推理网络,有效提升了计算效率与估计精度。LiteFlowNet3引入局部光流一致性约束,进一步提高了光流估计精度。与上述监督学习方法不同,DDFlow(Liu等,2019a)提出了一种称作无标签数据蒸馏的非监督训练策略。相较于其他非监督

方法, DDFlow 在多个测试数据集上均有更好的光流估计精度。与 DDFlow 类似, 但 SelFlow (Liu 等, 2019b) 提出了不同的训练样本生成方式, 进一步提高了光流估计精度。

表2 基于深度学习的转换模型估计

Table 2 Transformation modal estimation based on deep learning

参考文献	网络	训练方式	影像数据	针对问题	主要贡献 / 结论
Lee 和 Oh, 2021	CNN	监督	光学遥感影像	转换模型估计	利用 GMatchNet 估计转换模型提高特征点定位精度
Zampieri 等, 2018	FCN	监督	多模态遥感影像	位移场估计	利用 FCN 提取影像尺度不变特征, 通过级联网络估计全局位移场; 代码已开源: https://www.lri.fr/~gcharpia/alignment/
Girard 等, 2019	FCN	监督	地图—航拍影像	位移场估计 建筑检测	利用级联 FCN 校正和更新现有地图, 未增加额外计算成本; 基于“由粗到精”的策略提高配准精度; 代码已开源: https://github.com/Lydom/mapalignment
DLIR (de Vos 等, 2019)	CNN	强化	多模态医学影像	转换模型估计	通过级联 CNN 实现高精度配准, 大幅度提升传统方法的配准效率
MANet (Yang 等, 2021)	CNN	强化	4D-CT 影像	位移场估计	通过级联 CNN 估计位移场, 联合影像相似性测度与形变场正则项作为损失函数, 提升配准精度与效率
Luo 等, 2021	GAN	强化	胸部 X-ray 影像	位移场估计	首次将 UNet++ 应用于影像配准, 结合 ResNet 提高配准精度; 提出了一种基于损失约束的对抗训练方法, 增加训练的稳定性
DeepMapi (Ni 等, 2021)	CNN	监督	脑影像	位移场估计	提出一种自反馈策略, 以解决训练效率与准确率之间的矛盾; 采用多层配准策略, 不仅能够处理大几何形变, 对于局部小几何形变也有很好的效果

表3 基于深度学习的光流估计

Table 3 Optical flow estimation based on deep learning

参考文献	网络	训练方式	训练数据	测试数据	主要贡献 / 结论
SPyNet (Ranjan 和 Black, 2017)	CNN	监督	Flying Chairs	MPI-Sintel KITTI	轻量级网络; 针对传统方法无法应用于大错位影像的问题, 联合影像金字塔与深度学习计算影像光流; 代码已开源: https://github.com/anuragranj/spynet
LiteFlowNet (Hui 等, 2018)	CNN	监督	Things3D	Flying Chairs MPI-Sintel KITTI 等	轻量级网络; 引入一种亚像素修正层, 以解决大位移光流估计问题并保留光流细节; 为每一级光流结果构造正则化层, 保证其平滑性; 代码已开源: https://github.com/twhui/LiteFlowNet
LiteFlowNet2 (Hui 等, 2021)	CNN	监督	Things3D	Flying Chairs MPI-Sintel KITTI 等	相较于 LiteFlowNet, 减小了金字塔层数与网络深度, 引入伪光流估计网络; 精度与效率都明显优于 LiteFlowNet; 代码已开源: https://github.com/twhui/LiteFlowNet2
LiteFlowNet3 (Hui 和 Loy, 2020)	CNN	监督	Flying Chairs	Flying Chairs MPI-Sintel KITTI 等	在光流解码之前, 自适应调整每个代价向量以解决光流异常值问题; 精度与效率均优于 LiteFlowNet2; 代码已开源: https://github.com/twhui/LiteFlowNet3
DDFlow (Liu 等, 2019a)	CNN	非监督	—	Flying Chairs MPI-Sintel KITTI 等	提出了一种基于数据蒸馏的训练方法, 实现网络的非监督训练
SelFlow (Liu 等, 2019b)	CNN	非监督	—	KITTI 2012 KITTI 2015 MPI-Sintel	对 DDFlow 的数据蒸馏方式进行了调整; 通过添加随机噪声的方式生成遮挡影像, 提高了对于影像内部遮挡情况下的估计精度; 代码已开源: https://github.com/ppliuboy/SelFlow

综合来看, 基于深度学习的区域配准方法在医学影像、自然影像配准领域应用较多, 在遥感影像配准领域应用则较少。其原因包括但不限于: (1) 相较于自然影像与医学影像, 遥感影像具有更大的空间范围, 无论是基于深度学习的相似性

估计还是基于深度学习的转换模型估计, 其配准效率难以满足实际应用需求。(2) 遥感影像存在复杂的几何畸变, 在异源遥感影像配准中, 不同模态遥感影像又存在显著的非线性灰度映射问题。虽然深度学习在一定程度上提高了此类方法的鲁

棒性与适用性,但仍需更多实验来证明其在遥感影像上的稳健性。(3)标注数据集构建难度大。

但是基于深度学习的区域配准方法仍然具有十分重要的研究价值。(1)相较于传统方法,已有众多研究证明了基于深度学习的区域配准方法在配准效率、精度、鲁棒性上,均能得到显著提升。(2)由于地形起伏等原因导致影像存在复杂几何畸变时,基于转换模型的配准方法难以准确描述影像局部的空间对应关系,而基于深度学习的位移场估计具有更高的配准精度,在一定程度上缓解了上述问题。(3)相对于医学、自然影像,遥感影像场景往往更加复杂且不易解译。手工设计相似性测度难以准确描述不同影像配准效果,而基于数据驱动的深度学习方法能够适用于不同数据。虽然目前来看,基于深度学习的区域配准方法难以应用于大范围遥感影像配准,但由于此类方法在小尺寸影像配准上有着良好研究基础,

所以针对此类方法的研究仍然具有重要意义。

4 基于特征的配准方法

基于特征的配准方式是遥感影像配准领域中最常见的一类配准方法。此类方法的配准流程通常分为4个阶段:特征提取、特征匹配、转换模型求解、坐标转换与重采样(余先川等,2013;Feng等,2021)。

目前来看,深度学习应用于基于特征的配准方法,一般是基于“嵌入式模块”的策略:将神经网络嵌入到基于特征的配准方法框架中,并代替某一阶段的传统方法。现有研究大致可分为以下3类:基于深度学习的特征提取与描述、基于深度学习的特征匹配、基于深度学习的误匹配点剔除。考虑到传统方法通常将误匹配点剔除归于特征匹配阶段,且目前相关研究较少,所以本文也将其划分入基于深度学习的特征匹配一节中进行介绍。

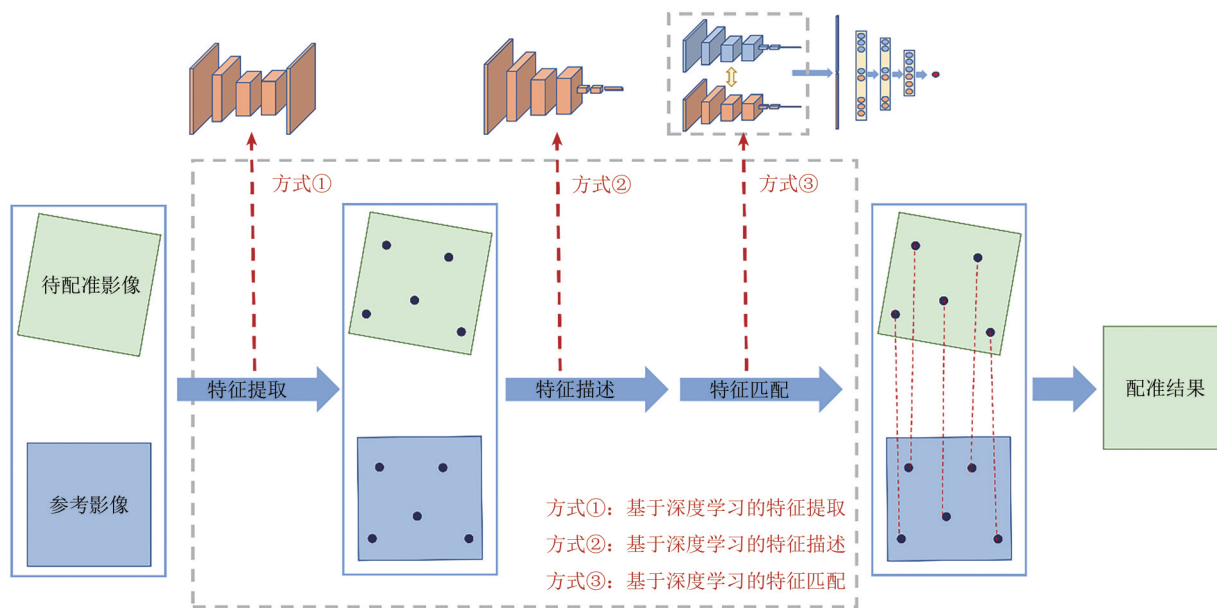


图2 基于深度学习的特征配准方法

Fig. 2 Feature-based registration methods based on deep learning

4.1 基于深度学习的特征提取与描述

基于特征的配准方法主要利用影像的几何特征实现影像配准,包括面特征(Rahaghi等,2019;Lu等,2021)、线特征(李力等,2020a)与点特征(Lee等,2021)。与二维面特征、一维线特征相比,零维点特征更易于提取、定位与描述,是更为常见的一类特征(冯蕊涛,2020)。特

征点提取能力、定位精度及特征描述是特征提取的3个关键问题。能否提取到具有代表性、稳定且分布均匀的特征点是特征提取算法最直观的评判标准,特征定位精度则直接影响配准精度,而特征描述是否鲁棒、是否具有代表性则关乎到特征匹配的质量与效率。传统特征提取方法如SUSAN (Smallest Univalued Segment Assimilating

Nucleus) 角点提取算子 (Smith 和 Brady, 1997)、SIFT (Scale Invariant Feature Transform) (Lowe, 2004)、RIFT (Radiation-variation Insensitive Feature Transform) (Li 等, 2020a) 等, 大多依赖影像浅层信息进行特征点提取与描述 (Fan 等, 2021)。由于遥感影像具有尺度变化大、覆盖地物类型复杂、受季候与光照条件影响大等特点, 存在影像灰度不连续的问题, 传统方法难以保证特征点与特征描述符的精度及稳定性 (Wang 等, 2018; Lee 等, 2021)。深度学习具有提取影像深层语义信息的能力, 相比于灰度、梯度等浅层信息, 深层语义信息冗余少, 能准确描述影像特征, 且更符合人类对于影像的理解。众多研究表明, 基于深度学习的特征提取方法可以提取稳定特征点以及赋予特征点鲁棒且具有代表性的特征描述符。

特征提取。利用神经网络提取特征一般基于以下两个策略: (1) 基于 Score Map 的特征点提取。此类方法通常利用 CNN 输出 Score Map, 再提取 Score Map 的局部响应值最大点作为特征点。TILDE (Temporally Invariant Learned DEtector) (Verdie 等, 2015)、SuperPoint (DeTone 等, 2018)、KeyNet (Barroso-Laguna 等, 2019) 是此类方法中的典型算法。TILDE 将包含稳定特征点的影像块作为训练样本, 对回归器进行监督训练, 根据输出的特征响应强度图, 提取影像特征。TILDE 在构建数据集时, 充分考虑了光照、天气等因素的影响, 并通过人工筛选保障了训练样本的可靠性, 所以此方法对于季候、光谱等变化具有强鲁棒性。但 TILDE

也存在对几何畸变考虑不充分、难以构建大规模数据集等问题。针对 TILDE 中存在的问题, SuperPoint、Quad-networks (Savinov 等, 2017) 与 Zhang 等 (2018) 提出通过合成数据集的方式对网络进行训练, 利用随机生成的转换模型对已有样本进行变换, 再基于样本间的空间对应关系构造损失函数, 提升了网络对于几何畸变的鲁棒性。在上述方法的基础上, KeyNet 联合传统 Harris 角点特征与 CNN 深层特征, 生成并融合不同尺度影像的 Score Map, 最终实现特征提取。KeyNet 充分考虑了尺度信息, 且联合了影像浅层与深层特征, 进一步提升了特征提取的鲁棒性。(2) 基于模糊匹配的特征点提取。此类方法将特征提取与匹配作为一个整体, 模拟传统相似性估计方法, 提出了一种基于相关层 (correlation-layer) 的匹配策略 (Rocco 等, 2017; Kim 等, 2019; Zeng 等, 2020): 首先利用孪生神经网络生成两景影像的特征图, 然后通过相关层计算特征图各点的相似性, 最后通过非极大值抑制的方式获取匹配点对。但该方法计算量大且不适用于大尺寸影像。Ma 等 (2019b) 基于由粗到精的策略, 逐特征层定位影像特征, 在提升特征点定位精度的基础上, 有效减小了计算量。

值得注意的是, 无论是基于 Score Map 的特征点提取策略还是基于区域匹配的特征点提取策略, 均存在定位精度不高的问题。Lee 等 (2021) 提供了一种基于区域的配准方法, 通过精校正待配准影像块的方式提升匹配特征点定位精度。

表 4 基于深度学习的特征提取

Table 4 Feature detection based on deep learning

参考文献	网络	训练方式	训练数据	测试数据	主要贡献 / 结论
Verdie 等, 2015	Regressor	监督	—	Oxford EF	对于季候、光照条件等具有强鲁棒性
DeTone 等, 2018	FCN	监督	MS-COCO 2014	HPatches	提出了一种自监督方法; 对光照及成像角度不敏感
Barroso-Laguna 等, 2019	CNN	监督	—	HPatches	提出了一种融合影像浅层与深层特征的提取网络; 比 SuperPoint 参数少且对成像角度变化的鲁棒性更好
Rocco 等, 2017	Siamese CNN	监督	—	Proposal Flow	基于模糊匹配的思路, 利用 correlation-layer 匹配影像特征点; 利用 CNN 预测变换模型参数
Kim 等, 2019	Siamese CNN	监督	—	ISPRS Google Earth	利用 Siamese Networks 与 correlation layer 提取、匹配特征; 两阶段估计转换模型参数
Zeng 等, 2020	Siamese CNN	监督	—	—	将配准与拼接视为一个整体任务; 相较于基于 SIFT 的配准方法, 在配准精度与效率上均有显著提升
Ma 等, 2019b	CNN	监督	—	—	提出了一种“由粗到精”的特征点定位方法

特征描述。特征描述的目的是对影像特征进行区分,使得同名特征在特征空间中的位置尽可能靠近,非同名特征则尽可能的分散。所以,如何构建稳健、准确且具有代表性的特征描述符是此类方法的研究重点。VGGNet (Visual Geometry Group NetWork) (Simonyan 和 Zisserman, 2015) 是特征描述中使用较多的一类 CNN。众多研究表明 (Yang等, 2018; Ye等, 2018; Dong等, 2019), VGGNet具有较深的网络结构,可以准确提取深层特征,且具有很强的泛化性与适用性。DescNet (Dong等, 2019) 是最为简单且直观的特征描述网

络。DescNet对输入影像块进行多层次卷积池化操作,得到影像块深层特征,最后输出128维向量作为特征描述符。Ye等 (2018)、Blendowski 和 Heinrich (2019) 与 Fan等 (2021) 在考虑了影像深层特征的基础上,联合SIFT等浅层特征,共同作为影像块特征描述符。实验证明,相较于单一的基于深度学习的或手工设计的相似性测度,联合特征具有更好鲁棒性,能够有效提高特征匹配数量与准确性。与上述方法类似,Yang等 (2018) 通过联合不同卷积层的特征描述符,提升了特征点定位精度以及特征描述符的鲁棒性。

表5 基于深度学习的特征描述

Table 5 Feature description based on deep learning

参考文献	网络	训练方式	主要贡献/结论
Yang等,2018	VGGNet	监督	使用预训练的VGGNet生成多尺度特征描述符;提出了一种点集配准策略
Ye等,2018	VGGNet	监督	提出了一种联合影像深层特征与SIFT特征的特征描述方法,并将其嵌入到PSO-SIFT方法
Dong等,2019	VGGNet	监督	减少了训练时间,提升了算法鲁棒性;相比于MatchNet与传统SIFT算法,DescNet能够正确匹配到更多特征点
Fan等,2021	Siamese	监督	深层残差网络可以提取深层特征;深层-浅层联合特征显著提升了多模态遥感影像配准精度

人工设计的特征提取算子对特征有一个固定定义,通常只能检测某一类的特征,当影像存在复杂几何、辐射畸变时,其特征定位、描述效果并不理想。基于深度学习的特征提取方法无需定义“标准特征”,通过对不同训练样本的学习,同一网络模型可以适用于不同类型特征提取问题。且神经网络模型具有大量的参数,能够描述传统方法难以考虑或顾及的某些隐性影响因素,所以基于深度学习的特征提取方法往往具有更好的鲁棒性。但目前来看,基于深度学习的特征提取仍然存在特征点定位精度不足、难以适用于大尺寸影像等问题。所以,针对遥感影像特征提取与描述的研究仍然是一项十分有价值且具有挑战性的工作。

4.2 基于深度学习的特征匹配

特征匹配是指按照一定的搜索策略匹配影像特征,使得影像同名特征点一一对应的过程。匹配特征点的数量及准确性决定了转换模型的精度。传统匹配方法一般是基于两阶段匹配策略:首先利用特征描述符计算特征点相似度,并根据某个阈值筛选出预匹配点对;再基于某种几何约束条件对预匹配点对进行精筛选。一方面,受制于特

征描述符与相似性测度的不稳定性,预匹配阶段难以保证匹配特征点对的单射关系,常常会出现误匹配问题;另一方面,当误匹配点对占据比例过大时,传统稳健估计方法难以保证最优解。

特征匹配质量很大程度上依赖于特征描述符与相似性测度的质量,所以基于深度学习的特征匹配往往将特征描述与匹配作为一个整体任务。根据匹配对象与匹配阶段的不同,基于深度学习的特征匹配可以划分为:单模态遥感影像特征匹配、多模态遥感影像特征匹配与误匹配点剔除。

单模态遥感影像特征匹配。目前,基于深度学习的单模态遥感影像特征匹配已形成了两套较为成熟的匹配框架。(1) 基于区域的特征匹配。该方法将特征提取、描述与匹配统一成同一个任务,利用双分支孪生网络分别提取影像深层特征,再引入相关层,逐区域计算两景影像的相关性,最后输出匹配结果。本文在特征提取一节中,已对此类方法进行介绍。(2) 基于相似性度量的特征匹配。此类方法的核心思想是将特征匹配视为二元分类问题,即利用权值共享的孪生网络构建输入影像的特征描述符,并将其输入全连接层中,

最后由全连接层判断影像是否匹配。Wang等(2018)与Fan等(2021)提出的匹配算法是此类方法的典型算法。Wang等(2018)针对遥感影像训练样本不足的问题,提出了一种自监督的训练策略,通过DBN判断两景影像块是否匹配。相较于DBN,CNN能够考虑到影像空间结构、光谱特征且具有更好的深层特征提取能力(Lee和Oh,2021)。(Zhu等,2019a;He等,2019;Lee和Oh,2021;Lee等,2021)将DBN替换为CNN,使用深层特征描述符代替影像灰度值作为全连接层的输入,提升了配准精度。

目前来看,相较于传统匹配方法,深度学习类方法主要通过优化特征描述符与相似性测度的方式提升匹配精度。具体来说,深度学习类方法

不再显式定义相似性测度,而是基于端到端训练的方式,使得数据、特征描述符与相似性测度具有更好的适配性。但是,目前绝大多数深度学习类匹配方法仍然是基于暴力匹配的搜索策略。一方面,暴力匹配大大提升了计算量与内存开支,显著降低了匹配效率,另一方面,暴力匹配无法规避一对多的误匹配问题。针对这一问题,Zhu等(2019a)提出了基于特征点强度与空间结构特征的搜索策略,有效提升了配准效率;He等(2018)则提出基于“由粗到精”的匹配策略,利用耦合二叉树模型提升匹配效率。此外,Lu等(2021)提出了一种基于实例分割的高分辨率遥感影像特征匹配策略,适用于特定场景下的尺寸、高空间分辨率遥感影像配准。

表6 基于深度学习的单模态遥感影像特征匹配

Table 6 Feature matching of monomodal remote sensing images based on deep learning

参考文献	网络	训练方式	训练数据	测试数据	主要贡献/结论
Wang等,2018	DBN	监督	—	—	提出了一种自学习的训练策略;影像浅层特征丰富但可分性差,深层特征包含更多语义特征,具有较强的可分性
He等,2018	Siamese CNN	监督	—	—	提出利用亚像素Harris算法精确定位匹配特征点;提出基于高斯金字塔耦合二叉树的匹配搜索策略
He等,2019	Siamese CNN	监督	—	—	引入“多支持块卷积层”、采用模板匹配代替模糊匹配策略,正确匹配点对数量约为SIFT算法的10倍
Zhu等,2019a	Siamese CNN	监督	Canada	Canada Yellow River MS_Canada	针对孪生网络特征匹配效率低的问题,提出了基于特征点强度(Sp-SGS)与空间结构特征(Sp-OSM)的搜索策略
Lu等,2021	CNN	监督	Vaihingen	WHU Building Dataset	提出了一种基于实例分割的自动配准框架,以解决由于空间定位不准确导致的误匹配问题
Fan等,2021	Siamese DRE	监督	—	DIR-Lab	提出了一种联合深层与浅层特征的多分辨率遥感图像匹配框架;相较于浅层局部特征,可以更充分地利用特征点邻近区域信息

多模态遥感影像特征匹配。多模态影像配准一直是遥感影像配准领域的研究难点与热点。相较于单模态影像配准,由于成像波段、成像机理不同,不同模态影像在辐射、几何上存在明显差异。就光学—SAR影像配准而言,SAR影像存在透视收缩、阴影、斑点噪声等现象,光学影像受成像条件影响严重,影像可能存在云、地物阴影等噪声。一方面,多模态影像存在严重的非线性灰度映射问题,传统基于区域的配准方法难以实现高精度配准;另一方面,不同模态影像特征显著但相互独立。如何提取不同模态影像特征并将其统一到同一度量标准下,以实现提高特征匹

精度、数量的目的,是此类方法的重要研究方向。

目前,针对异源遥感影像特征差异大的问题,多模态遥感影像特征匹配算法也大致可以划分为两个类别:(1)沿用单模态影像特征匹配思路。Hughes等(2018)与Shabanov等(2020)利用全连接层替代传统方法的L2范数计算影像相似性测度;CMM-Net(蓝朝桢等,2021)采用VGGNet提取异源影像特征,将异源影像特征描述符进行L2范式归一化,使得异源影像特征描述符尽可能地统一,最后实现异源影像特征匹配;Zhang等(2019a)与Zhu等(2019b)基于区域匹配的方式实现光学—SAR影像特征匹配。虽然此类方法在光

学-SAR影像配准问题上取得了一些进展,但是异源影像特征差异大的问题仍然十分显著,当影像存在较大尺度、角度差异时,此类方法仍难以保证其鲁棒性。(2)通过引入GAN等网络模型,将不同模态影像转换为同一模态影像,以抑制由于成像机

理等原因引起的特征差异。Zhang等(2019b)与Zeng等(2021)使用伪孪生网络将SAR、光学影像迁移到同一种风格,以削弱光学与SAR影像之间的辐射、几何差异,再充分利用现有同源遥感影像配准方法,有效降低了多模态影像配准难度。

表7 基于深度学习的多模态遥感影像特征匹配

Table 7 Feature matching of multi-modal remote sensing images based on deep learning

参考文献	网络	训练方式	主要贡献/结论
Zhang等,2019a	Siamese FCN	监督	设计了一种基于FCN的孪生神经网络,采用最大正负样本特征距离作为损失函数,提升了匹配效果
Zhang等,2019b	CNN	监督	使用风格迁移的方法对影像进行预处理,以抑制影像由于成像机制等原因造成的几何、灰度上的差异
Zhu等,2019b	CNN	监督	双通道CNN具有良好的泛化能力;密集连接卷积层将影像匹配性能提高了10%以上,证明了CNN浅层特征的有效性
Shabanov等,2020	Siamese CNN	监督	在Siamese Networks特征匹配的基础上,将相关系数作为全连接层的输入,提升了网络性能;无法配准有明显旋转错位的影像
Zeng等,2021	Siamese CNN	监督	利用CNN对影像进行语义分割并提取影像区域特征,对多源遥感影像配准效果较好
蓝朝桢等,2021	VGGNet	监督	CMM-Net具有较强的稳健性与适用性,在匹配点数量、分布与效率等方面具有优势

误匹配点剔除。迭代加权最小二乘算法IRSL (iteratively reweighted least-squares) (Holland和Welsch, 1977)与随机抽样一致算法RANSAC (Random Sample Consensus) (Fischler和Bolles, 1981)是传统方法中常用的稳健估计方法。当误匹配点对比例过高时或影像存在严重非刚性变化时,IRSL与RANSAC难以实现最优模型估计(Li等, 2020b; Jiang等, 2021b)。随着深度学习的发展,有学者将深度学习应用于误匹配点剔除。

LFGC (Yi等, 2018)根据已知的匹配点集与相机内方位元素,通过弱监督的方式训练网络模型。依据转换模型构建损失函数,最后输出匹配点集中的内点、外点以及转换模型参数。但是LFGC方法需要相机内方位元素作为输入,且过于依赖假定的转换模型,难以处理复杂几何形变的情况。针对上述问题,LMR (Ma等, 2019c)将误匹配点剔除视为一个二元分类问题,利用训练好的通用分类器判断匹配的正确性。实验证明,从定性与定量的角度分析,LMR方法在效率与准确率上都优于LFGC方法。最近, Jiang等(2021b)基于空间一致性假设检测误匹配点,并通过多次迭代的策略逐步剔除错点,在效率上可以达到时间线性复杂度。此外,该方法为稀疏点集处理问题提供了新的研究思路。由于匹配点数据是典型的非欧几里得数据结构,具有无序性与分散性,

所以如何利用深度学习提取匹配点的空间结构信息仍然是一个研究难点(Jiang等, 2021b)。

特征匹配是基于特征的配准方法的一个重要环节,特征匹配的准确性与匹配效率很大程度上决定了配准算法的精度与适用性。一方面,基于深度学习的特征匹配方法将特征描述与匹配作为一个整体任务,通过端对端的训练策略,在一定程度上解决了传统匹配方法匹配准确率低、匹配特征数量不足的问题;但另一方面,目前此类方法仍然存在匹配效率低、标准数据集不足等问题。

5 配准数据集

网络模型的训练需要大量样本,虽然目前有大量遥感影像数据可供使用,但是人工标注数据匮乏的问题仍然十分突出。针对上述问题,目前常见有3种解决策略:(1)基于传统方法,通过自动或半自动的方式构建数据集。如利用Harris角点提取算子(Harris和Stephens, 1988)、SIFT算法或人工标记方法,构建影像特征数据集;或者基于传统方法匹配影像特征,再通过人工审查筛选样本,以构建匹配样本数据集。此类方法易于理解与实现,且容易获取少量标注样本,但难以构建大样本数据集。一方面,无论是人工标注还是后续审查,都需要耗费大量成本;另一方面,传统方法难以准确获取真值,其准确性与精度难以

保证。(2) 通过人工合成的方式构建数据集。利用随机生成的转换模型或通过添加噪声的方式, 对已有影像数据进行变换。此类方法易于构建大样本数据集, 且数据集内任意影像间的空间对应关系明确, 但是人工合成数据集难以准确模拟真实影像间复杂的空间、灰度映射关系, 很大程度上制约了网络的适用性与鲁棒性。(3) 通过非监督的方式训练网络。此类方法以参考影像与待配准影像之间的相似性测度或者某种几何约束条件引导网络优化。一方面, 目前遥感影像的获取渠道广泛, 易于构建大样本且数据类型丰富的无标签数据集, 但另一方面, 此类方法也存在对网络初始参数敏感、收敛速度慢与易于陷入局部最优解等问题。

深度学习方法能否取得好的效果, 在很大程度上取决于是否有好的训练数据集。一般认为, 评价数据集质量好坏取决于3点:(1) 样本数量是否合理;(2) 样本种类是否丰富且均衡;(3) 标签信息是否准确。本文总结了目前几个常用且开源的数据集, 主要为遥感影像数据集, 另辅以少量自然影像数据集作为补充, 并对数据集的特点进行分析。

光学影像数据集。University-1652 (Zheng等, 2020) 包含1652个建筑物近120000景影像, 包括通过Google Map等平台获取的地面摄影影像数据、卫星影像数据以及通过Google Earth模拟生成的无人机视角影像数据。University-1652数据源、成像视角丰富、具有明确的标注信息, 能够为多视角、异源影像匹配研究提供丰富的数据资料。但是, 对于单模态遥感影像配准研究, University-1652存在卫星影像数据少、覆盖地物类型单一等问题。

HPatches (Balntas等, 2017) 利用三维重建方法获得特征点位置和匹配关系, 是一个标准数据集。HPatches通过对匹配影像对添加不同的噪声, 如几何变换、光照变换等, 构建影像数据集, 根据视角和光照变化情况, 将影像分为3个子数据集: EASY、HARD和TOUGH, 每个子数据集包括 2×10^5 个正样本与 1×10^6 个负样本。常用于自然影像特征提取、特征描述与特征匹配等领域。

Google Earth Dataset (Park等, 2020) 是一个人工合成数据集。Park等通过对多时相航空影像

添加随机仿射变换模型生成标注样本, 并构建数据集。该数据集包含9000个多时相影像对以及其对应的转换模型。

SAR-光学影像数据集。SEN1-2 (Schmitt等, 2018) 包含4组数据, 分别对应四季, 共282384组光学-SAR遥感影像对, 样本均为Sentinel系列卫星于2017年拍摄影像, 在空间上覆盖全球5大洲不同地表区域。其中光学影像由Sentinel-2卫星获取, 包括可见光三波段, 空间分辨率为10m; SAR影像由Sentinel-1卫星获取, 极化方式为垂直极化, 空间分辨率为30m。SEN1-2使用Google Earth Engine提取影像并完成影像预处理, 再通过两轮人工审查, 筛除存在大面积云雾覆盖等问题的影像。在保证大量数据的前提下, 显著提升了样本质量。SEN1-2具有样本数量大、样本种类丰富且时空分布均衡、样本质量高等优势, 在异源遥感影像配准及相关领域研究中发挥了重要作用。但是, 由于SEN1-2光学影像并没有做大气校正且仅仅包含RGB三波段, 所以对于一些对地表反射率反演、多光谱影像配准等研究来说, 仍存在一定局限性。

SARptical数据集 (Wang和Zhu, 2018) 是一个免费公开的光学-SAR影像数据集。样本为UltraCAM与TerraSAR-X传感器于2009年—2013年获取的德国柏林影像, 共10000余组影像对。与SEN1-2相比, SARptical样本具有更高的空间分辨率, 其中光学影像空间分辨率为0.2 m, SAR影像空间分辨率为1 m, 可以为基于深度学习的异源遥感影像配准研究提供数据支持。此外, 其光学影像除可见光三波段外, 还包括全色与近红外波段, 在一定程度上弥补了SEN1-2数据集仅包含RGB波段的问题。但是SARptical存在覆盖区域小, 地物类型单一, 影像数据量较少等问题。

OS-dataset (Xiang等, 2020) 包含10692组光学-SAR影像对, 影像尺寸为 256×256 。其中SAR影像为GF-3获取, 空间分辨率为1 m; 光学影像则是从Google Earth平台获取, 空间分辨率采样到1 m。相较于SEN1-2、SARptical, OS-dataset具有空间分辨率高、空间分布广、地物类型丰富等优点。此外, 通过数据增强的方式, OS-dataset数据样本可以进一步得到提升。

表8 配准数据集
Table 8 Registration datasets

数据集	影像	特点	获取渠道
University-1652 (Zheng等,2020)	光学遥感 自然	数据量大、成像视角丰富、具有明确标注信息;适用于多视角或多模态影像配准	https://github.com/layumi/University1652-Baseline
Hpatches (Balntas等,2017)	自然	基准数据集;适用于自然影像配准、遥感影像配准网络预训练等	https://github.com/hpatches
Google Earth Dataset (Park等,2020)	光学遥感	人工合成数据集,适用于遥感影像配准	https://github.com/jaehyunnn/DeepAerialMatching
SEN1-2 (Schmitt等,2018)	光学-SAR	样本数量大、质量高、种类丰富且时空分布均匀;光学影像未做大气校正且仅包含可见光波段,难以训练网络挖掘光谱信息;适用于多模态影像配准	https://mediatum.ub.tum.de/1436631
SARptical (Wang和Zhu,2018)	光学-SAR	空间分辨率高、光学影像包括全色、可见光及近红外波段;覆盖区域小,地物类型单一;数据量较少;适用于多模态影像配准	http://www.sipeo.bgu.tum.de/downloads/SARptical_data.zip
OS-dataset (Xiang等,2020)	光学-SAR	数据量丰富,空间覆盖广,地物信息丰富;用于多模态影像配准	https://pan.baidu.com/s/14bqaJhMSZEy7EXcXVAc77w with the code "vriw."

6 遥感影像配准技术展望

得益于对地观测技术的巨大进步,越来越多的遥感影像被投入应用。影像配准是遥感影像预处理的关键步骤,也是定量遥感分析的基础研究内容。虽然目前已有大量针对遥感影像配准的研究成果,但针对高效率、高精度、高鲁棒性算法的研究仍然是遥感影像配准领域的主流研究方向。近些年来,基于深度学习的遥感影像配准算法的研究呈现出逐年增长的趋势,但仍处于初期,研究思路与方法框架尚不成熟。概括来说,基于深度学习的遥感影像配准包括但不限于以下不足:(1)缺少大样本且开源的标准数据集;(2)难以应用于大范围遥感影像;(3)对于遥感影像的地理空间信息、光谱数据信息等利用不充分;(4)网络训练时间长,计算开销大。从数据、方法的角度来看,深度学习在遥感影像配准领域存在巨大研究价值。

6.1 从数据的角度

神经网络对数据具有高度依赖性,训练数据的质与量直接影响网络性能。与自然影像相比,遥感影像覆盖地物信息更加复杂,当成像条件、传感器类型、影像分辨率发生变化时,同一地物在影像上往往呈现出不同的光谱或几何特征。所以构建配准数据集需要更加精细、复杂的标签信息对遥感影像进行描述、分类与组织。

遥感影像配准数据集。数据集生产是深度学习算法研究的基础性工作,其重要性不言而喻。就遥感影像配准任务而言,构建配准数据集存在几个明显问题:(1)遥感影像种类多、尺度跨度大、几何与辐射畸变复杂,所以需要海量数据构建数据集,以应对不同配准任务;(2)目前已有方法局限性大,缺少适用于大多数配准任务的标准算法,无法实现全自动构建数据集;(3)难以获得真值。在一些配准任务中,如基于深度学习的转换模型估计等,即使利用人工标注的方法,也难以获取真值数据。概括来说,同源遥感影像易于获取,相关配准方法也较为成熟,易于构建小样本数据集,但构建大样本同源遥感影像数据集成本大。异源遥感影像获取、筛选难度大,虽然目前已有如SEN1-2、SARptical等高质量、开源的异源遥感影像数据集,但现有数据集也存在覆盖研究范围较窄等问题。如何将已有配准算法与人工结合,兼顾精度与生产效率,使“大量、无序”的遥感影像数据归顺为“有序”的数据集,是未来遥感影像配准领域的重要研究内容与目标。

6.2 从方法的角度

基于模型驱动的传统配准方法与基于数据驱动的深度学习方法各具优势。传统配准方法需要依赖专家对问题的先验理解,以构建具有物理或数学意义的模型,但是无法顾及所有影响影像几何形变的因素。在实际应用中,往往需要通过人

工干预的方式对模型流程或参数进行调整,以减小隐性因素的影响。神经网络基于“学习—反馈—调整”策略,能够从大量相同模式的训练数据中抽象出数据的共性特征,且由于网络具有足够多的参数,通过多次迭代训练,一些难以被人工提取、描述的隐性影响因素也能够被纳入模型,在一定程度上解决了专家依赖的问题。从另一个角度看,传统配准方法经历了数十年的发展,已经形成了成熟的影像配准流程。相对于深度学习方法,传统方法过程透明,稳定性高,易于调整、优化,且对硬件要求低,在处理简单配准任务时,能够达到更高的配准精度与效率。

目前,深度学习已广泛应用于遥感影像配准领域。(1)从形式上来看,可以将其划分为“模块化网络”与端对端网络。“模块化网络”是指利用深度学习解决配准流程中某一阶段性任务,如基于深度学习的特征提取等。此类方法易于训练,灵活性强,但难以避免多阶段任务中易出现的误差累计问题,且可能陷入局部最优;端对端网络则利用一个网络直接实现影像配准,如基于深度学习的位移场模型估计。相较于“模块化”网络,端对端网络的训练目标一致,能够得到全局最优解,但也存在训练难度大,解释性差等问题;(2)从应用上来看,可以将其划分为基于区域与基于特征的配准方法。一方面,许多研究表明,将深度学习应用于传统配准框架内,可以有效提升算法精度及鲁棒性。但另一方面,传统配准框架的局限性仍然存在。如基于深度学习特征的配准方法仍无法避免误差累计的问题,基于深度学习区域的配准方法在处理几何差异较大的影像时,难以保证鲁棒性等。目前来看,两类配准方法的优缺点明显,且适用范围、对象也存在差异。如何充分发挥深度学习的完全数据驱动优势,以提升两类方法的配准精度与效率,是今后配准研究的重要内容。(3)从训练方式的角度看,可以将已有算法划分为监督学习与无监督学习两个类别。一方面,在影像配准领域,无监督学习的研究基础较弱,损失函数的构建难度大,且存在训练时间长,易于陷入局部最优等问题;另一方面,相较于无监督学习与强化学习,基于监督学习的深度学习往往能取得更好的精度,人工合成数据集的方式在很大程度上缓解了标注样本不足的问题,所以监督学习是目前且在较长一段时间内都

将是影像配准领域的主流训练策略。

基于混合模型的遥感影像配准方法。将传统配准方法与深度学习相互结合,发挥不同方法的优势,是遥感影像配准的重要发展方向之一。比如,使用深度学习代替人工设计特征提取算子。Wang等(2018)将深度学习嵌入到传统基于特征的配准框架中,利用全连接网络实现特征匹配,通过提高特征匹配精度的方式提升最终配准效果。此外,也可以参考Feng等(2019)提出的特征灰度联合配准方法的思路,利用传统基于特征方法粗配准全局影像,再利用深度学习对影像进行局部精校正。一方面,深度学习仅用来精校正局部小尺寸影像块,可以有效降低网络训练难度;另一方面,传统配准框架中融入了物理模型,保证了算法的鲁棒性与可解释性。

基于不同网络模型的遥感影像配准方法。在遥感影像配准领域中,CNN是目前使用最为广泛的一类模型(Kuppala等,2020)。针对其他模型的研究仍相对较少,未来可以尝试将其他网络模型应用于遥感影像配准领域,比如:在异源遥感影像配准中使用GAN,以削弱异源影像之间的差异;使用Transformer(Vaswani等,2017)代替CNN以实现特征匹配等。此外,基于端到端网络的遥感影像配准方法也是遥感影像配准的重要研究方向。相对于多阶段模型,端到端网络具有很多显著的优势,如:不存在误差累计问题,不易陷入局部最优解,网络训练更简单等。在自然影像配准领域,目前已有许多针对端到端网络的研究,并取得了十分优秀的成果(Yi等,2016;Liu等,2019a,2019b)。

基于小样本的神经网络训练策略。在数据集匮乏的情况下,可以通过迁移学习、半监督学习等弱监督方法(Zhou,2018)对网络进行训练,也可以通过数据增强(马崇霖等,2021)的方式,对现有小样本数据集进行扩充。基于上述方法,可以充分利用已有资源,降低生产数据集所需要的时间、人力成本。

7 结 语

随着传感器技术的不断发展,遥感影像的观测尺度更加精细,覆盖地物信息也更加复杂。由于地形起伏等原因引起的局部畸变问题也愈加受到重视(冯蕊涛等,2021),很多在大尺度遥感影

像配准上取得良好效果的传统配准方法,并不适用于小尺度遥感影像的高精度配准。

近些年来,深度学习为遥感影像配准提供了新的研究方向。相对于模型驱动的传统配准方法,基于数据驱动的深度学习方法具有许多显著优势。首先,深度学习不需要人工预先定义影像特征,通过对大量训练数据的学习,可以提取到影像的深层语义特征。在特征提取、匹配等应用中,深度学习方法取得了许多优秀成果(Zhang和Rusinkiewicz, 2018; Wang等, 2018; Hughes等, 2018; Zhang等, 2019b)。其次,深度学习具有更好的适用性。神经网络由众多小神经元组成,通过训练对神经元交互过程进行调整,同一网络可以处理不同的任务(Yang等, 2018; 蓝朝桢等, 2021)。此外,通过优化硬件的方式,深度学习方法可以实现更高效的计算。

本文在简要介绍传统遥感影像配准方法的优缺点与遥感影像配准中几种常见网络的基础上,重点介绍了深度学习方法在基于特征的方法、基于区域的方法两方面的重要进展,并介绍了几种公开的遥感影像配准数据集。最后对遥感影像配准技术进行了展望。目前,基于深度学习的遥感影像配准方法研究思路与配准框架尚未成熟。但已有研究表明,基于深度学习的遥感影像配准方法在配准精度、配准效率上都可以达到甚至超越当前传统配准方法的最优水平。可以预见,基于深度学习的遥感影像配准方法会成为实时、高精度遥感影像配准问题的重要解决方法。

参考文献(References)

- Balntas V, Lenc K, Vedaldi A and Mikolajczyk K. 2017. HPatches: a benchmark and evaluation of handcrafted and learned local descriptors//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii: IEEE: 3852-3861 [DOI: 10.1109/CVPR.2017.410]
- Barroso-Laguna A, Riba E, Ponsa D and Mikolajczyk K. 2019. Key-Net: keypoint detection by handcrafted and learned CNN filters. arXiv: 1904.00889v3
- Blendowski M and Heinrich M P. 2019. Combining MRF-based deformable registration and deep binary 3D-CNN descriptors for large lung motion estimation in COPD patients. International Journal of Computer Assisted Radiology and Surgery, 14(1): 43-52 [DOI: 10.1007/s11548-018-1888-2]
- Chen H M, Varshney P K and Arora M K. 2003. Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 41(11): 2445-2454 [DOI: 10.1109/TGRS.2003.817664]
- Cheng X, Zhang L and Zheng Y F. 2018. Deep similarity learning for multimodal medical images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 6(3): 248-252 [DOI: 10.1080/21681163.2015.1135299]
- Chopra S, Hadsell R and LeCun Y. 2005. Learning a similarity metric discriminatively, with application to face verification//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE: 539-546 [DOI: 10.1109/CVPR.2005.202]
- de Vos B D, Berendsen F F, Viergever M A, Sokooti H, Staring M and Išgum I. 2019. A deep learning framework for unsupervised affine and deformable image registration. Medical Image Analysis, 52: 128-143 [DOI: 10.1016/j.media.2018.11.010]
- DeTone D, Malisiewicz T and Rabinovich A. 2018. SuperPoint: self-supervised interest point detection and description. arXiv: 1712.07629v4
- Dong Y Y, Jiao W L, Long T F, Liu L F, He G J, Gong C J and Guo Y T. 2019. Local deep descriptor for remote sensing image feature matching. Remote Sensing, 11(4): 430 [DOI: 10.3390/rs11040430]
- Fan R B, Hou B C, Liu J B, Yang J H and Hong Z L. 2021. Registration of multiresolution remote sensing images based on L2-Siamese model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 237-248 [DOI: 10.1109/JSTARS.2020.3038922]
- Feng R T, Du Q Y, Li X H and Shen H F. 2019. Robust registration for remote sensing images by combining and localizing feature- and area-based methods. ISPRS Journal of Photogrammetry and Remote Sensing, 151: 15-26 [DOI: 10.1016/j.isprsjprs.2019.03.002]
- Feng R T. 2020. The Registration Algorithm for Remote Sensing Images Covering the Complex Terrain by Multi-Model Union. Wuhan: Wuhan University: 1-122 (冯蕊涛. 2020. 复杂地形条件下的遥感影像多模型联合配准方法. 武汉: 武汉大学: 1-122)
- Feng R T, Shen H F, Bai J J and Li X H. 2021. Advances and opportunities in remote sensing image geometric registration: a systematic review of state-of-the-art approaches and future research directions. IEEE Geoscience and Remote Sensing Magazine, 9(4): 120-142 [DOI: 10.1109/MGRS.2021.3081763]
- Feng R T, Du Q Y, Luo H, Shen H F, Li X H and Liu B. 2021. A registration algorithm based on optical flow modification for multi-temporal remote sensing images covering the complex-terrain region. National Remote Sensing Bulletin, 25(2): 630-640 (冯蕊涛, 杜清运, 罗恒, 沈焕锋, 李星华, 刘波. 2021. 基于光流校正的复杂地形区多时相遥感影像配准. 遥感学报, 25(2): 630-640) [DOI: 10.11834/jrs.20209280]
- Fischler M A and Bolles R C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6): 381-

- 395 [DOI: 10.1145/358669.358692]
- Girard N, Charpiat G and Tarabalka Y. 2019. Aligning and updating cadaster maps with aerial images by multi-task, multi-resolution deep learning//14th Asian Conference on Computer Vision. Perth: Springer: 675-690 [DOI: 10.1007/978-3-030-20873-8_43]
- Gong M G, Zhao S M, Jiao L C, Tian D Y and Wang S. 2014. A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. *IEEE Transactions on Geoscience and Remote Sensing*, 52(7): 4328-4338 [DOI: 10.1109/TGRS.2013.2281391]
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial networks. arXiv: 1406.2661v1
- Harris C and Stephens M. 1988. A combined corner and edge detector// *Proceedings of the Alvey Vision Conference*. Manchester: AVC: 147-151 [DOI: 10.5244/C.2.23]
- Haskins G, Kruecker J, Kruger U, Xu S, Pinto P A, Wood B J and Yan P K. 2019. Learning deep similarity metric for 3D MR - TRUS image registration. *International Journal of Computer Assisted Radiology and Surgery*, 14(3): 417-425 [DOI: 10.1007/s11548-018-1875-7]
- He H Q, Chen M, Chen T and Li D J. 2018. Matching of remote sensing images with complex background variations via Siamese convolutional neural network. *Remote Sensing*, 10(2): 355 [DOI: 10.3390/rs10020355]
- He H Q, Chen M, Chen T, Li D J and Cheng P G. 2019. Learning to match multitemporal optical satellite images using multi-support patches Siamese networks. *Remote Sensing Letters*, 10(6): 516-525 [DOI: 10.1080/2150704x.2019.1577572]
- Hinton G E and Salakhutdinov R R. 2006. Reducing the dimensionality of data with neural networks. *Science*, 313(5786): 504-507 [DOI: 10.1126/science.1127647]
- Holland P W and Welsch R E. 1977. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-Theory and Methods*, 6(9): 813-827 [DOI: 10.1080/03610927708827533]
- Hughes L H, Schmitt M, Mou L C, Wang Y Y and Zhu X X. 2018. Identifying corresponding patches in SAR and optical images with a Pseudo-Siamese CNN. *IEEE Geoscience and Remote Sensing Letters*, 15(5): 784-788 [DOI: 10.1109/LGRS.2018.2799232]
- Hui T W, Tang X O and Loy C C. 2018. LiteFlowNet: a lightweight convolutional neural network for optical flow estimation//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 8981-8989 [DOI: 10.1109/CVPR.2018.00936]
- Hui T W and Loy C C. 2020. LiteFlowNet3: resolving correspondence ambiguity for more accurate optical flow estimation//16th European Conference on Computer Vision. Glasgow: Springer: 169-184 [DOI: 10.1007/978-3-030-58565-5_11]
- Hui T W, Tang X O and Loy C C. 2021. A lightweight optical flow CNN—revisiting data fidelity and regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8): 2555-2569 [DOI: 10.1109/TPAMI.2020.2976928]
- Jiang X Y, Ma J Y, Xiao G B, Shao Z F and Guo X J. 2021a. A review of multimodal image matching: methods and applications. *Information Fusion*, 73: 22-71 [DOI: 10.1016/j.inffus.2021.02.012]
- Jiang X Y, Ma J Y, Fan A X, Xu H P, Lin G, Lu T and Tian X. 2021b. Robust feature matching for remote sensing image registration via linear adaptive filtering. *IEEE Transactions on Geoscience and Remote Sensing*, 59(2): 1577-1591 [DOI: 10.1109/TGRS.2020.3001089]
- Johnson K, Cole-Rhodes A, Zavorin I and Le Moigne J. 2001. Mutual information as a similarity measure for remote sensing image registration//*Proceedings Volume 4383, Geo-Spatial Image and Data Exploitation II*. Orlando: SPIE: 51-61 [DOI: 10.1117/12.428251]
- Kim D G, Nam W J and Lee S W. 2019. A robust matching network for gradually estimating geometric transformation on remote sensing imagery//2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). Bari: IEEE: 3889-3894 [DOI: 10.1109/SMC.2019.8913881]
- Kuppala K, Banda S and Barige T R. 2020. An overview of deep learning methods for image registration with focus on feature-based approaches. *International Journal of Image and Data Fusion*, 11(2): 113-135 [DOI: 10.1080/19479832.2019.1707720]
- Lan C Z, Lu W J, Yu J M and Xu Q. 2021. Deep learning algorithm for feature matching of cross modality remote sensing images. *Acta Geodaetica et Cartographica Sinica*, 50(2): 189-202 (蓝朝桢, 卢万杰, 于君明, 徐青. 2021. 异源遥感影像特征匹配的深度学习算法. *测绘学报*, 50(2): 189-202) [DOI: 10.11947/j.AGCS.2021.20200048]
- LeCun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278-2324 [DOI: 10.1109/5.726791]
- LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436-444 [DOI: 10.1038/nature14539]
- Lee W, Sim D and Oh S J. 2021. A CNN-based high-accuracy registration for remote sensing images. *Remote Sensing*, 13(8): 1482 [DOI: 10.3390/rs13081482]
- Lee W J and Oh S J. 2021. Remote sensing image registration using equivariance features//2021 International Conference on Information Networking (ICOIN). Bangkok: IEEE: 776-781 [DOI: 10.1109/ICOIN50884.2021.9333861]
- Li J Y, Hu Q W and Ai M Y. 2020a. RIFT: multi-modal image matching based on radiation-variation insensitive feature transform. *IEEE Transactions on Image Processing*, 29: 3296-3310 [DOI: 10.1109/TIP.2019.2959244]
- Li J Y, Zhao P C, Hu Q W and Ai M Y. 2020b. Robust point cloud registration based on topological graph and Cauchy weighted l_q -norm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160: 244-259 [DOI: 10.1016/j.isprsjprs.2019.12.008]
- Li L, Ji S, Yu Y and Zhang Y S. 2020. A multi-feature-based registration method adapted to multi-source remote sensing images. *Journal of Geomatics Science and Technology*, 37(1): 74-78 (李力, 纪松, 于英, 张永生. 2020. 一种基于组合特征的异源遥感影像配

- 准方法. 测绘科学技术学报, 37(1): 74-78 [DOI: 10.3969/j.issn.1673-6338.2020.01.014]
- Li X H, Feng R T, Guan X B, Shen H F and Zhang L P. 2019. Remote sensing image mosaicking: achievements and challenges. *IEEE Geoscience and Remote Sensing Magazine*, 7(4): 8-22 [DOI: 10.1109/MGRS.2019.2921780]
- Li X H, Du Z S, Huang Y Y and Tan Z Y. 2021. A deep translation (GAN) based change detection network for optical and SAR remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 179: 14-34 [DOI: 10.1016/j.isprsjprs.2021.07.007]
- Liang J Y, Liu X P, Huang K N, Li X, Wang D G and Wang X W. 2014. Automatic registration of multisensor images using an integrated spatial and mutual information (SMI) metric. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1): 603-615 [DOI: 10.1109/TGRS.2013.2242895]
- Liang Y, Sheng Y H, Zhang K and Yang L. 2014. Linear feature matching method based on local affine invariant and Epipolar constraint for close-range images. *Geomatics and Information Science of Wuhan University*, 39(2): 229-233 (梁艳, 盛业华, 张卡, 杨林. 2014. 利用局部仿射不变及核线约束的近景影像直线特征匹配. 武汉大学学报(信息科学版), 39(2): 229-233) [DOI: 10.13203/j.whugis.20120611]
- Liao R, Miao S, de Tournemire P, Grbic S, Kamen A, Mansi T and Comaniciu D. 2016. An artificial agent for robust image registration. arXiv: 1611.10336v1
- Liu P P, Irwin K, Lyu M R and Xu J. 2019a. DDFlow: learning optical flow with unlabeled data distillation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1): 8770-8777 [DOI: 10.1609/aaai.v33i01.33018770]
- Liu P P, Lyu M, King I and Xu J. 2019b. SelfFlow: self-supervised learning of optical flow//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE: 4571-4580 [DOI: 10.1109/CVPR.2019.00470]
- Lowe D G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91-110 [DOI: 10.1023/b:visi.0000029664.99615.94]
- Lu J Y, Jia H G, Li T, Li Z Q, Ma J Y and Zhu R F. 2021. An instance segmentation based framework for large-sized high-resolution remote sensing images registration. *Remote Sensing*, 13(9): 1657 [DOI: 10.3390/rs13091657]
- Luo Y, Cao W M, He Z Q, Zou W L and He Z H. 2021. Deformable adversarial registration network with multiple loss constraints. *Computerized Medical Imaging and Graphics*, 91: 101931 [DOI: 10.1016/j.compmedimag.2021.101931]
- Ma D A, Tang P, Zhao L J and Zhang Z. 2021. Review of data augmentation for image in deep learning. *Journal of Image and Graphics*, 26(3): 487-502 (马崇昇, 唐婷, 赵理君, 张正. 2021. 深度学习图像数据增广方法研究综述. 中国图象图形学报, 26(3): 487-502) [DOI: 10.11834/jig.200089]
- Ma F L. 2018. Research on Color-To-Gray Conversion based on Normalized Cross Correlation. Lanzhou: Lanzhou University: 1-61 (马方龙. 2018. 基于归一化积相关匹配的彩色图像灰度化研究. 兰州: 兰州大学: 1-61)
- Ma J Y, Jiang X Y, Jiang J J, Zhao J and Guo X J. 2019c. LMR: learning a two-class classifier for mismatch removal. *IEEE Transactions on Image Processing*, 28(8): 4045-4059 [DOI: 10.1109/TIP.2019.2906490]
- Ma J Y, Jiang X Y, Fan A X, Jiang J J and Yan J C. 2021. Image matching from handcrafted to deep features: a survey. *International Journal of Computer Vision*, 129(1): 23-79 [DOI: 10.1007/s11263-020-01359-2]
- Ma L, Liu Y, Zhang X L, Ye Y X, Yin G F and Johnson B A. 2019a. 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]
- Ma W P, Zhang J, Wu Y, Jiao L C, Zhu H and Zhao W. 2019b. A novel two-step registration method for remote sensing images based on deep and local features. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7): 4834-4843 [DOI: 10.1109/TGRS.2019.2893310]
- Ni H, Feng Z, Guan Y, Jia X Y, Chen W, Jiang T, Zhong Q Y, Yuan J, Ren M, Li X N, Gong H, Luo Q M and Li A N. 2021. DeepMapi: a fully automatic registration method for mesoscopic optical brain images using convolutional neural networks. *Neuroinformatics*, 19(2): 267-284 [DOI: 10.1007/s12021-020-09483-7]
- Niu R G, Sun X, Tian Y, Diao W H, Chen K Q and Fu K. 2020. Hybrid multiple attention network for semantic segmentation in aerial images. arXiv: 2001.02870v3
- Park J H, Nam W J and Lee S W. 2020. A two-stream symmetric network with bidirectional ensemble for aerial image matching. *Remote Sensing*, 12(3): 465 [DOI: 10.3390/rs12030465]
- Rahaghi A I, Lemmin U, Sage D and Barry D A. 2019. Achieving high-resolution thermal imagery in low-contrast lake surface waters by aerial remote sensing and image registration. *Remote Sensing of Environment*, 221: 773-783 [DOI: 10.1016/j.rse.2018.12.018]
- Ranjan A and Black M J. 2017. Optical flow estimation using a spatial pyramid network//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii: IEEE: 2720-2729 [DOI: 10.1109/CVPR.2017.291]
- Rocco I, Arandjelovic R and Sivic J. 2017. Convolutional neural network architecture for geometric matching//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Hawaii: IEEE: 39-48 [DOI: 10.1109/CVPR.2017.12]
- Savinov N, Seki A, Ladicky L, Sattler T and Pollefeys M. 2017. Quad-networks: unsupervised learning to rank for interest point detection. arXiv: 1611.07571v2
- Schmidhuber J. 2015. Deep learning in neural networks: an overview. *Neural Networks*, 61: 85-117 [DOI: 10.1016/j.neunet.2014.09.003]
- Schmitt M, Hughes L H and Zhu X X. 2018. The SEN1-2 dataset for deep learning in SAR-optical data fusion. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1: 141-146 [DOI: 10.5194/isprs-annals-IV-1-141-2018]

- Seo S, Choi J S, Lee J, Kim H H, Seo D, Jeong J and Kim M. 2020. UPSNet: unsupervised pan-sharpening network with registration learning between panchromatic and multi-spectral images. *IEEE Access*, 8: 201199-201217 [DOI: 10.1109/ACCESS.2020.3035802]
- Shabanov A, Gladilin S and Shvets E. 2020. Optical-to-SAR image registration using a combination of CNN descriptors and cross-correlation coefficient//Twelfth International Conference on Machine Vision (ICMV). Amsterdam: SPIE: 440-449 [DOI: 10.1117/12.2558414]
- Shelhamer E, Long J and Darrell T. 2017. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4): 640-651 [DOI: 10.1109/TPAMI.2016.2572683]
- Simonovsky M, Gutiérrez-Becker B, Mateus D, Navab N and Komodakis N. 2016. A deep metric for multimodal registration//19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens: Springer: 10-18 [DOI: 10.1007/978-3-319-46726-9_2]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556v6
- Smith S M and Brady J M. 1997. SUSAN-a new approach to low level image processing. *International Journal of Computer Vision*, 23(1): 45-78 [DOI: 10.1023/A:1007963824710]
- Suri S and Reinartz P. 2010. Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2): 939-949 [DOI: 10.1109/TGRS.2009.2034842]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2017. Attention is all you need. arXiv: 1706.03762v5
- Verdie Y, Yi K M, Fua P and Lepetit V. 2015. TILDE: a temporally invariant learned detector//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE: 5279-5288 [DOI: 10.1109/CVPR.2015.7299165]
- Wang S, Quan D, Liang X F, Ning M D, Guo Y H and Jiao L C. 2018. A deep learning framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145: 148-164 [DOI: 10.1016/j.isprsjsprs.2017.12.012]
- Wang Y Y and Zhu X X. 2018. The SARptical dataset for joint analysis of SAR and optical image in dense urban area. arXiv: 1801.07532
- Wang Z H and Wu F C. 2009. Mean-standard deviation descriptor and line matching. *Pattern Recognition and Artificial Intelligence*, 22(1): 32-39 (王志衡, 吴福朝. 2009. 均值-标准差描述子与直线匹配. 模式识别与人工智能, 22(1): 32-39) [DOI: 10.3969/j.issn.1003-6059.2009.01.006]
- Xiang Y M, Tao R S, Wang F, You H J and Han B. 2020. Automatic registration of optical and SAR images via improved phase congruency model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 5847-5861 [DOI: 10.1109/JSTARS.2020.3026162]
- Yang J, Yang J H, Zhao F and Zhang W J. 2021. An unsupervised multi-scale framework with attention-based network (MANet) for lung 4D-CT registration. *Physics in Medicine and Biology*, 66(13): 135008 [DOI: 10.1088/1361-6560/ac0afc]
- Yang Z Q, Dan T T and Yang Y. 2018. Multi-temporal remote sensing image registration using deep convolutional features. *IEEE Access*, 6: 38544-38555 [DOI: 10.1109/ACCESS.2018.2853100]
- Yao M Q and Hu J. 2020. Robust multimodal medical image registration using deep recurrent reinforcement learning. *Journal of Computer-Aided Design and Computer Graphics*, 32(8): 1236-1247 (姚明青, 胡靖. 2020. 基于深度强化学习的多模态医学图像配准. 计算机辅助设计与图形学学报, 32(8): 1236-1247) [DOI: 10.3724/SP.J.1089.2020.17847]
- Ye F M, Su Y F, Xiao H, Zhao X Q and Min W D. 2018. Remote sensing image registration using convolutional neural network features. *IEEE Geoscience and Remote Sensing Letters*, 15(2): 232-236 [DOI: 10.1109/LGRS.2017.2781741]
- Ye Y X and Shan J. 2014. A local descriptor based registration method for multispectral remote sensing images with non-linear intensity differences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 90: 83-95 [DOI: 10.1016/j.isprsjsprs.2014.01.009]
- Yi K M, Trulls E, Lepetit V and Fua P. 2016. LIFT: learned invariant feature transform//14th European Conference on Computer Vision (ECCV). Amsterdam: Springer: 467-483 [DOI: 10.1007/978-3-319-46466-4_28]
- Yi K M, Trulls E, Ono Y, Lepetit V, Salzmann M and Fua P. 2018. Learning to find good correspondences//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE: 2666-2674 [DOI: 10.1109/CVPR.2018.00282]
- Yu X C, Lü Z H and Hu D. 2013. Review of remote sensing image registration techniques. *Optics and Precision Engineering*, 21(11): 2960-2972 (余先川, 吕中华, 胡丹. 2013. 遥感图像配准技术综述. 光学精密工程, 21(11): 2960-2972) [DOI: 10.3788/OPE.20132111.2960]
- Yuan Z, Guo H T, Lu J, Lu W and Lin Y Z. 2021. High-resolution remote sensing image change detection technology based on UNet++ and attention mechanism. *Journal of Geomatics Science and Technology*, 38(2): 155-159 (袁洲, 郭海涛, 卢俊, 路威, 林雨准. 2021. 融合UNet++网络和注意力机制的高分辨率遥感影像变化检测算法. 测绘科学技术学报, 38(2): 155-159) [DOI: 10.3969/j.issn.1673-6338.2021.02.008]
- Zampieri A, Charpiat G, Girard N and Tarabalka Y. 2018. Multimodal image alignment through a multiscale chain of neural networks with application to remote sensing//15th European Conference on Computer Vision (ECCV). Munich: Springer: 679-696 [DOI: 10.1007/978-3-030-01270-0_40]
- Zeng L, Du Y L, Lin H P, Wang J, Yin J J and Yang J. 2021. A novel region-based image registration method for multisource remote sensing images via CNN. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 1821-1831 [DOI: 10.1109/JSTARS.2020.3047656]
- Zeng Y, Ning Z H, Liu P, Luo P L, Zhang Y and He G J. 2020. A mosa-

- ic method for multi-temporal data registration by using convolutional neural networks for forestry remote sensing applications. *Computing*, 102(3): 795-811 [DOI: 10.1007/s00607-019-00716-5]
- Zhang H, Ni W P, Yan W D, Xiang D L, Wu J Z, Yang X L and Bian H. 2019a. Registration of multimodal remote sensing image based on deep fully convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8): 3028-3042 [DOI: 10.1109/JSTARS.2019.2916560]
- Zhang J, Ma W P, Wu Y and Jiao L C. 2019b. Multimodal remote sensing image registration based on image transfer and local features. *IEEE Geoscience and Remote Sensing Letters*, 16(8): 1210-1214 [DOI: 10.1109/LGRS.2019.2896341]
- Zhang L G and Rusinkiewicz S. 2018. Learning to detect features in texture images//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE: 6325-6333 [DOI: 10.1109/CVPR.2018.00662]
- Zhao L Y, Lü B Y, Li X R and Chen S H. 2015. Multi-source remote sensing image registration based on scale-invariant feature transform and optimization of regional mutual information. *Acta Physica Sinica*, 64(12): 124204 (赵辽英, 吕步云, 厉小润, 陈淑涵. 2015. 基于尺度不变特征变换和区域互信息优化的多源遥感图像配准. *物理学报*, 64(12): 124204) [DOI: 10.7498/aps.64.124204]
- Zheng Z D, Wei Y C and Yang Y. 2020. University-1652: a multi-view multi-source benchmark for drone-based geo-localization. arXiv: 2002.12186v2
- Zhou Z H. 2018. A brief introduction to weakly supervised learning. *National Science Review*, 5(1): 44-53 [DOI: 10.1093/nsr/nwx106]
- Zhu H, Jiao L C, Ma W P, Liu F and Zhao W. 2019a. A novel neural network for remote sensing image matching. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9): 2853-2865 [DOI: 10.1109/TNNLS.2018.2888757]
- Zhu R J, Yu D W, Ji S P and Lu M. 2019b. Matching RGB and infrared remote sensing images with densely-connected convolutional neural networks. *Remote Sensing*, 11(23): 2836 [DOI: 10.3390/rs11232836]

Survey of remote sensing image registration based on deep learning

LI Xinghua¹, AI Wenhao¹, FENG Ruitao^{1,2}, LUO Shaojie³

1. School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China;

2. School of Geography and Tourism, Shaanxi Normal University, Xi'an 710119, China;

3. State Grid Hangzhou Power Supply Company, Hangzhou 310020, China

Abstract: Remote sensing image registration is the process of spatial alignment of two or more images through geometric transformation. It is an important preprocessing operation for image fusion, change detection, agricultural monitoring and other remote sensing applications. Considering that remote sensing images have the characteristics of large-scale changes, complex ground covers and imaging modalities, although a large number of registration methods have been developed, there is still a lack of methods that can be widely used in different scenarios. Therefore, research on registration algorithms with high efficiency, high robustness, high precision and wide applicability is of great significance. In recent years, deep learning, which has achieved great success in the field of natural image and medical image registration, has provided a new method for remote sensing image registration. First, we introduced two kinds of traditional registration methods and analyzed the advantages and disadvantages of area-based and feature-based registration methods in detail from the aspects of registration accuracy, efficiency and algorithm robustness. Generally, there are two main problems in traditional methods: poor applicability and insufficient utilization of the deep semantic information of the image. Second, we focused on the important progress of deep learning in area-based registration methods and feature-based registration methods. According to the specific application purpose of deep learning, we made a more detailed division of the above two methods and summarized the advantages and disadvantages of the existing research. In addition, considering the importance of datasets for deep learning, we sorted and shared some public datasets for remote sensing image registration. Due to the great progress of earth observation technology, an increasing number of remote sensing images are being applied. Image registration is the key step of remote sensing image preprocessing and the basic research content of quantitative remote sensing analysis. In recent years, research on remote sensing image registration algorithms based on deep learning has shown an increasing trend, but it is still in the early stage, and the framework is not mature. It mainly includes but is not limited to the following shortcomings: (1) lack of open source standard datasets; (2) difficult to apply to large-scale remote sensing images; (3) insufficient utilization of geospatial information and spectral information of remote sensing images; and (4) long training time and the large computing overhead. From the perspective of data and methods, we looked forward to the application of deep learning in the field of remote sensing image registration and put forward four main research directions: (1) remote sensing image registration datasets; (2) registration methods based on hybrid models; (3) registration methods based on different neural networks; and (4) training strategies based on small samples.

Key words: deep learning, image registration, area-based, feature-based, registration datasets

Supported by State Grid Science and Technology Project (No.5700-202019162A-0-0-00)