

强化位置感知的光学与SAR图像一体化配准方法

杨玉婷¹ 赵凌君^{*1} 赵路路² 张 晗^{1,3} 熊博莅¹ 计科峰¹

(1. 国防科技大学电子科学学院CEMEE国家重点实验室, 湖南长沙 410073;
2. 中国人民解放军第95369部队, 广东佛山 528000; 3. 西北核技术研究院, 陕西西安 710024)

摘要: 图像配准是光学与SAR图像信息融合的基础。现有典型的配准方法大多依赖于特征点检测与匹配来实现, 对不同场景区域的适用性较差, 容易出现误匹配点多或有效同名点不足以致配准失效的情况。针对该问题, 本文提出了一种强化位置感知的光学与SAR图像一体化配准方法, 利用深度网络直接回归图像间的几何变换关系, 在不依赖特征点检测与匹配的情况下, 实现端到端的高精度配准。具体地, 首先在骨干网络中利用融合坐标注意力的特征提取模块, 捕获输入图像对中具有位置敏感性的细粒度特征; 其次, 融合骨干网络输出的多尺度特征, 兼顾浅层特征的定位信息与高层特征的语义信息; 最后提出联合位置偏差与图像相似性的损失函数优化配准结果。基于高分辨率光学与SAR图像配准公开数据集OS-Dataset的实验结果表明, 与现有典型的OS-SIFT、RIFT2、DHN及DLKFM四种算法相比, 所提方法对于城市、农田、河流、重复纹理及弱纹理等不同场景区域均具有良好的稳健性, 在配准的目视效果以及定量的精度指标上均优于现有算法。其中平均角点误差小于3个像素的百分比与四种算法中精度最高的DLKFM相比提高了25%以上; 配准速度与四种算法中最快的DHN基本相当, 可实现高精度、高效率的光学与SAR图像配准。

关键词: 图像配准; 位置感知; 多尺度特征; 特征融合; 图像相似性

中图分类号: TP751.1 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2024.03.014

引用格式: 杨玉婷, 赵凌君, 赵路路, 等. 强化位置感知的光学与SAR图像一体化配准方法[J]. 信号处理, 2024, 40(3): 557-568. DOI: 10.16798/j.issn.1003-0530.2024.03.014.

Reference format: YANG Yuting, ZHAO Lingjun, ZHAO Lulu, et al. Integrated registration method with enhanced position awareness for optical and SAR images[J]. Journal of Signal Processing, 2024, 40(3): 557-568. DOI: 10.16798/j.issn.1003-0530.2024.03.014.

Integrated Registration Method with Enhanced Position Awareness for Optical and SAR Images

YANG Yuting¹ ZHAO Lingjun^{*1} ZHAO Lulu² ZHANG Han^{1,3} XIONG Boli¹ JI Kefeng¹

(1. State Key Laboratory of CEMEE, College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China;
2. Army 95369 of PLA, Foshan, Guangdong 528000, China;
3. Northwest Institute of Nuclear Technology, Xi'an, Shaanxi 710024, China)

Abstract: Image registration is the basis for optical and SAR image information fusion. Most of the existing typical registration methods rely on feature-point detection and matching. However, because of their poor applicability to different

收稿日期: 2023-06-27; 修回日期: 2023-08-22

*通信作者: 赵凌君 nudzlj@163.com *Corresponding Author: ZHAO Lingjun, nudzlj@163.com

基金项目: 湖南省自然科学基金项目(2021JJ40684); 卫星信息智能处理与应用技术重点实验室自主研究基金项目(2022-ZZKY-JJ-10-02)
Foundation Items: Hunan Provincial Natural Science Foundation of China (2021JJ40684); and Independent Research Fund of Key Laboratory of Satellite Information Intelligent Processing and Application Technology (2022-ZZKY-JJ-10-02)

scene regions, these methods are prone to problems such as excessive mismatched points and insufficiently effective matched points, resulting in invalid registration. Therefore, this study investigated an integrated registration method with enhanced position awareness for optical and SAR images. This method utilizes a deep neural network to directly regress the geometric transformation relationship between images. The proposed method achieves end-to-end high-precision registration without relying on feature-point detection. First, a feature-extraction module that integrates coordinate attention is used in the backbone network to extract position-sensitive fine-grained features from the input image pairs. Second, the multiscale features of the backbone network output are fused, taking into account the positional information of low-level features and semantic information of high-level features. Finally, a loss function that combines the position deviation and image similarity is used to optimize the registration results. Experimental results based on a publicly available high-resolution optical and SAR dataset (OS-Dataset) demonstrated that compared with four existing typical algorithms (OS-SIFT, RIFT2, DHN, and DLKFM), the proposed method had good robustness for different scene areas such as urban, farmland, river, repetitive texture, and weak texture scenes, and outperformed the existing algorithms in terms of visual effects and a quantitative precision metric. The percentage of average corner errors of fewer than 3 pixels was more than 25% better than that of DLKFM, which had the highest precision among the four algorithms. The registration speed was comparable to that of DHN, which was the fastest of the four algorithms. The proposed method could achieve high-precision and high-efficiency optical and SAR image registration.

Key words: image registration; position awareness; multi-scale feature; feature fusion; image similarity

1 引言

光学图像与合成孔径雷达(Synthetic Aperture Radar, SAR)图像是对地观测的两种主要数据源,其反映的地物特征信息存在明显的互补性^[1]。通过结合光学与SAR图像的特性进行融合分析,在变化检测、地物分类、目标识别等方面具有重要的应用前景^[2],而图像配准是融合分析实现的重要技术前提。图像配准技术广泛应用于弹载景象匹配制导、飞行器定位导航及地物变化检测等任务,如何实现精度高、实时性强的图像配准是该领域当前的研究重点之一。由于成像机理迥异,除了同源图像之间存在的几何结构、旋转、分辨率等差异外,光学图像与SAR图像之间还存在显著的非线性辐射差异(Nonlinear Radiation Distortions, NRD)^[3],适用于同源图像配准的方法用于异源配准时,性能往往大幅下降。近年来,针对异源遥感图像配准问题,诸多行之有效的方法被相继提出,主要分为传统方法和基于深度学习的方法两大类。

传统的光学与SAR图像配准方法根据配准所利用的信息不同,可分为基于特征的方法和基于区域的方法两大类。在基于特征的方法中,目前大多基于SIFT(Scale Invariant Feature Transformation)算法^[4]进行适应性改进,如OS-SIFT^[5]、PCG-SIFT^[6]等。该类算法主要通过图像的局部邻域梯度信息检测特征点并构建描述符,利用描述符之间的相似性度量得到初始匹配,再通过迭代估计的方法进行误匹配剔除。由于光学与SAR图像辐射差异较大,

根据局部邻域梯度信息所构建的特征并不稳健,因此配准效果较差。基于区域的方法重点在于相似性度量准则的选择^[7],对于存在显著非线性灰度差异的光学与SAR图像,互信息(Mutual Information, MI)^[8]相较于平方差之和^[9]、归一化互相关^[10]等相似性度量更具有稳健性,因此,诸多基于区域的配准研究围绕MI进行改进^[11-12]。此外,基于频域的相位一致性特征^[13]可有效克服异源图像间的NRD。文献[2]提出利用相位一致性特征图代替原图进行特征点检测及描述,取得了较好的配准效果。HOPC(Histogram of Orientated Phase Congruency)^[14]利用相位一致性特征值和方向构建区域特征描述符,并以模板匹配的方式完成精确配准,但只适用于已消除明显几何差异的图像。总的来说,上述算法皆依赖于人工设计特征,对于不同图像场景泛化性不够强,且实时性较差。

近年来,基于深度学习的光学与SAR图像配准方法得到了广泛的研究,相关方法主要分为两类:一类是深度学习与传统配准相结合的方法,另一类是端到端的配准方法。深度学习与传统配准相结合通常是将神经网络结合到传统的图像配准流程中。该类方法又可细分为两类:一是基于风格迁移的配准方法^[15-16],即利用深度网络(如生成对抗网络(Generative Adversarial Network, GAN))统一待配准图像对的模态,缓解光学与SAR图像间的NRD影响,再通过同源配准方法进行配准,该类方法在风格迁移的过程中很难模拟真实的遥感图像特征,且过于依赖训练数据,泛化能力较弱,应用范围受限;二

是基于深度特征描述符的方法,主要思想是利用深度特征代替经典的手工特征,构建深度特征描述符或特征图,通过欧氏距离或相关性热力图等进行精确匹配。文献[17]将裁剪出的异源图像块对输入权值共享的孪生网络得到深度特征图,通过模板匹配实现同名点的定位及误匹配点对剔除;Hughes提出了一种端到端的光学与SAR图像匹配框架^[18],分别训练了粗匹配、精匹配、误匹配点对剔除等三个卷积神经网络;文献[19]利用卷积神经网络同时完成了特征点检测与描述两个任务;OSM-Net(Optical and SAR image Matching Network)^[20]通过伪孪生网络,同时采用多层特征融合、多频通道激励及自适应加权损失等策略对输入图像对进行像素级密集特征提取,实现高分辨率光学与SAR图像的高精度匹配。该类方法大多聚焦于图像块的匹配,对于配准任务来说存在以下问题:部分方法只适用于预测图像间的平移量,对于显著的仿射变换鲁棒性较差;部分方法利用多阶段网络获得匹配结果后仍需要进行迭代参数估计,配准过程繁琐复杂,实时性受到影响。

为了提高配准的实时性,学者们开始研究端到端的配准方法,该类方法旨在通过深度网络预测参考图像与待配准图像之间的几何变换关系。根据网络的结构设计不同,可细分为两类。一类是多阶段网络的流程化方法,即利用深度网络模拟传统配准过程中的特征检测与匹配、参数估计,多阶段网络按流程串联,实现端到端配准。文献[21]利用孪生卷积网络提取图像对的特征图,并通过内积得到特征图的相关度信息,最后利用回归网络得出变换参数。但由于特征图分辨率较低,根据特征图间的匹配关系进行配准的精度不高;文献[22]利用皮尔逊相关性对匹配关系进行改进,提升了光学遥感图像的配准精度。另一类是直接回归参数的一体化方法,即利用神经网络对变换参数直接回归,且其中并不划分如特征检测、特征匹配等任务模块。DHN(Deep Homography Network)^[23]是该类方法的一项开创性工作,通过人工生成仿真数据获取监督信息以实现变换参数的直接回归,从而完成自然图像的配准;文献[24]通过多尺度级联的方式对三个DHN进行有效堆叠以提升参数估计精度;Voxel-Morph^[25]利用输入图像对之间的局部相关性及约束正则化项构造损失函数,以无监督学习的方式进行优化,完成医学图像的配准。由于图像特性不一致,且应用场景不同(如这类方法在用于医学配准时通常要求图像对预先仿射对齐),端到端的配准

方法在光学与SAR图像配准领域中尚未广泛应用。

上述研究工作大多需要通过特征点检测、特征提取、特征匹配及参数迭代估计来完成整个配准流程,过程繁琐的同时,依赖于良好的特征点检测结果,难以解决不同场景中光学与SAR图像存在显著几何变换的精确配准问题(如同名点不足的弱纹理区域)。基于DHN的配准算法利用深度卷积网络直接估计图像间的几何变换,实现了端到端的配准,但该类方法进行参数回归时只关注了深层特征的语义信息,忽略了浅层特征的定位信息,难以实现图像位置的精确对齐。针对以上问题,本文在DHN算法端到端估计几何变换参数的基础思想之上,提出一种强化位置感知的光学与SAR图像一体化配准方法。利用卷积神经网络提取位置敏感的语义信息,自动学习待配准图像对之间的空间变换关系,无须进行特征点检测及迭代后处理,适用于各种不同的图像场景区域。

2 本文方法

为实现异源图像端到端的高精度配准,提升特征的定位能力尤为重要。不同于特征检测-特征匹配-参数估计的多阶段处理方法,本文提出了一种强化位置感知的一体化配准模型,利用融合坐标注意力的特征提取模块强化位置信息,多尺度特征融合模块聚合全局和局部信息,提高模型的位置敏感性,直接回归图像间的变换关系以获取精确的配准结果。本节将详细介绍一体化配准框架、融合坐标注意力的特征提取模块、多尺度特征融合模块以及联合位置偏差与图像相似性的损失函数四个部分。

2.1 一体化配准框架

所提一体化配准方法包括特征提取、特征融合、参数回归及图像重采样四个部分,总体框架如图1所示。首先,输入图像对在通道方向叠加,通过融合坐标注意力(Coordinate Attention, CA)^[26]的特征提取模块得到位置信息敏感的深度特征表达。为进一步提高特征的定位能力,通过多尺度特征融合模块(Multi-Scale Feature Fusion, MSFF)对不同层级的深度特征图进行有效聚合,获得同时具有深层语义表征能力以及精确定位信息的显著特征。最后利用全连接层进行几何变换参数的映射,并通过空间变换模块(Spatial Transformer, ST)^[27]对待配准图像进行仿射变换,得到配准后图像。在训练阶段,网络的输入图像对由预先对齐的配准数据集进行随机仿射变换所得,网络优化依靠随机仿射变

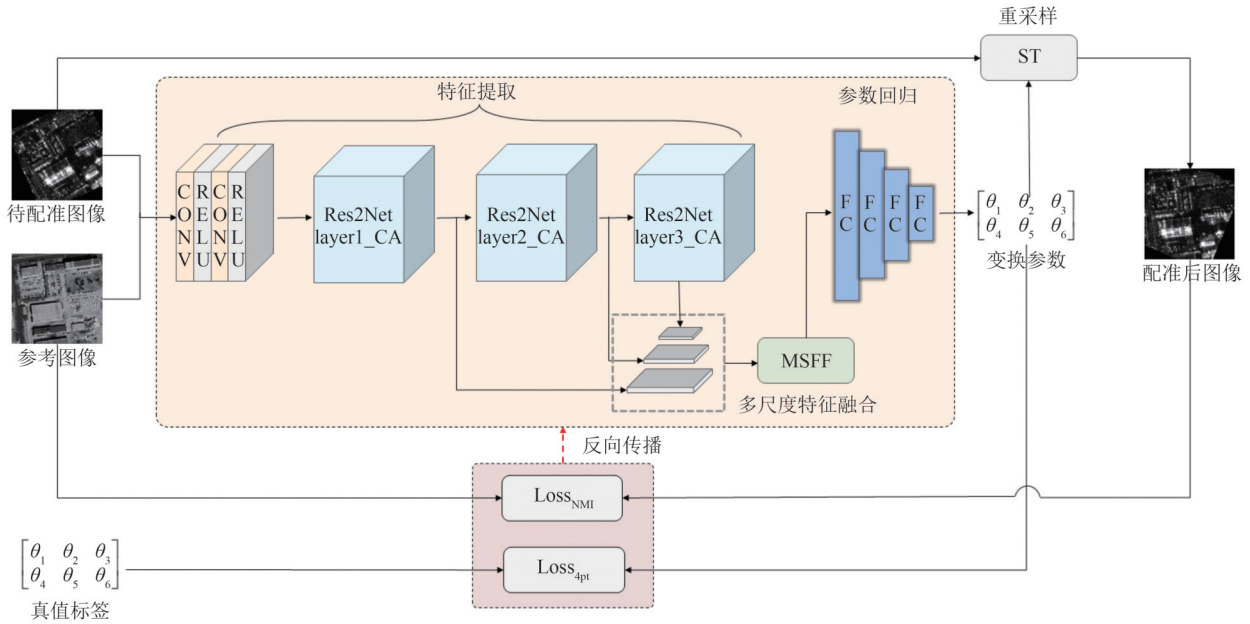


图1 强化位置感知的光学与SAR图像一体化配准框架

Fig. 1 Integrated registration framework with enhanced position awareness for optical and SAR images

换的参数标签以及参考图像和配准后图像的相似性度量组合实现。

2.2 融合坐标注意力的特征提取模块

卷积网络所提取的深层语义信息一定程度上能够适应光学和SAR图像的辐射差异,且特征感受野的适当扩充,更有利于异源图像深层特征语义的表征。然而,随着卷积网络层数的加深,特征定位依赖的位置信息有所丢失。为提取具有较强泛化能力以及位置信息的语义特征,采用具备多尺度特征表示能力的Res2Net网络结构^[28]来提取输入图像对的深层语义信息,并通过融合坐标注意力机制^[26],显式地对特征位置信息进行精确建模,提升特征定位能力。

Res2Net是ResNet的变体网络,以多个Res2Net module级联的方式实现深度特征提取。Res2Net module是对普通ResNet架构使用的bottleneck block的重建,通过在单个残差块内构建分层的特征组并进行类残差连接,以取代通用的 3×3 卷积核组,扩大网络层的感受野。如图2左侧所示,在输入特征图经过 1×1 卷积后,将其按通道数均匀分割成 n 个特征子集,即每个子集空间大小相同,通道数为 $1/n$ 。除 x_i 之外,每个 x_i 都有对应的 3×3 卷积组 H_i ,将经过 H_i 的输出特征记为 y_i ,则块内的类残差连接可表述为:特征子集 x_i 与前一组的输出 y_{i-1} 相加,经过卷积组 H_i 后得到该组输出 y_i ,相应的数学表

达为:

$$y_t = \begin{cases} x_t & t=1; \\ H_t(x_t) & t=2; \\ H_t(x_t + y_{t-1}) & 2 < t \leq n \end{cases} \quad (1)$$

上述过程重复多次,最后将来自所有卷积组的输出特征图 $\{y_1, y_2, \dots, y_n\}$ 进行拼接,再次经过 1×1 卷积,以实现特征信息的充分融合。由于组合效应, y_t 具有比 x_t 更大的感受野,等效感受野增加,重新拼接后产生细粒度的多尺度特征。

经过上述的Res2Net module进行特征提取,可获得较为抽象的高级语义特征。但由于采用了卷积、池化等操作,通常来说,这种语义特征所包含的有利于定位的特征信息(例如边缘、纹理等)不够充分,从而对于位置的敏感性不强,特征定位精度较差。然而,与分类和场景识别等任务不同,配准任务对特征的精确位置信息具有较高的需求。因此,为使网络进一步挖掘有利于输入图像对空间关系映射的重要信息,充分利用特征间的空间位置关系,引入坐标注意力机制对Res2Net module进行改进。通过一维全局自适应池化生成水平和垂直方向的方向感知特征图,实现坐标信息的嵌入。方向特征图进一步变换生成的坐标注意力图,引导特征学习聚焦于位置信息,从而强化位置信息在最终回归决策中的作用。具体结构如图2所示。

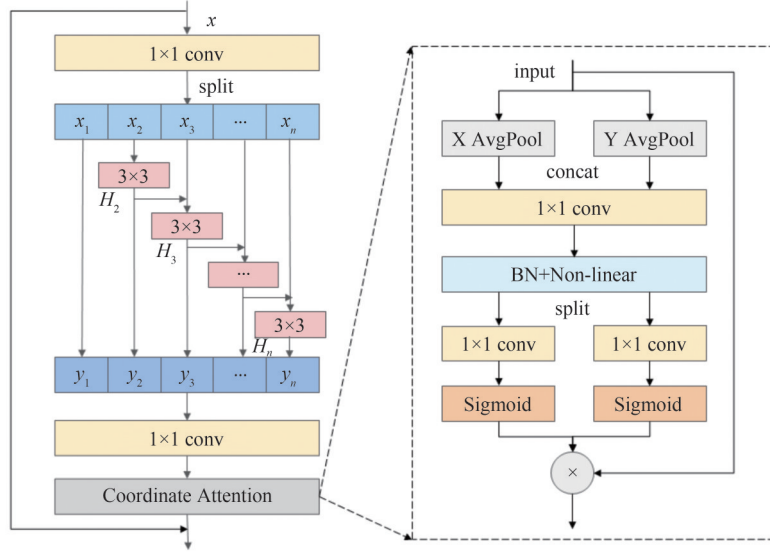


图 2 融合坐标注意力的 Res2Net module

Fig. 2 Res2Net module integrating coordinate attention

在 CA 模块中,通过对输入特征 x 分别进行 X、Y 方向的一维全局自适应池化,将水平和垂直方向的输入特征聚合为两个相互独立的方向感知特征图,以在获取沿某一方向的远程依赖关系的同时,保留沿另一方向的精确位置信息。该过程称之为坐标信息嵌入,可理解为对通道关系和空间远程依赖关系的特征编码,相应数学表达为:

$$\begin{aligned} z_c^h(h) &= \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \\ z_c^w(w) &= \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \end{aligned} \quad (2)$$

式中, $z_c^h(h)$ 、 $z_c^w(w)$ 分别为高度 h 、宽度 w 的第 c 个通道的输出。为了有效利用信息嵌入过程中获取的表征信息,需要对上述的两个方向感知特征图进行变换,生成坐标注意力图。具体地,首先将特征图 z^h 、 z^w 进行级联,并使用 1×1 卷积变换 K_1 得到中间特征图 f ,即:

$$f = \delta(K_1([z^h, z^w])) \quad (3)$$

式中, δ 表示非线性激活单元(ReLU)。按空间维度将其分割为 $f^h \in \mathbb{R}^{(C/r) \times H}$ 及 $f^w \in \mathbb{R}^{(C/r) \times W}$,通过 1×1 卷积调整通道数,使其与输入通道数相同,得到注意力权重如下:

$$\begin{aligned} g^h &= \sigma(K_1^h(f^h)) \\ g^w &= \sigma(K_1^w(f^w)) \end{aligned} \quad (4)$$

式中, σ 为 Sigmoid 激活函数。 g^h 和 g^w 是注意力权

重,将其与输入特征 x 相乘,即可得到最终输出:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (5)$$

通过结合 Res2Net module 的细粒度特征提取能力及坐标注意力的位置感知能力,特征提取模块可以自动地学习图像间特征的空间位置信息,帮助网络更加精确地捕捉图像间存在的几何变换关系。

2.3 多尺度特征融合模块

深度卷积神经网络在架构设计中遵循特征图空间大小逐步减小的方式,来获取抽象的高级语义特征。这些深层特征具有较大的感受野和丰富的语义信息,但分辨率较低,忽略了图像边缘、角点等辅助图像配准的定位信息。若采取浅层特征预测,特征图的分辨率虽然较高,但其中的干扰信息也较多。尤其是对于具有显著非线性灰度差异的光学与 SAR 图像来说,浅层特征中可用于配准的异源图像语义信息十分有限。同时,浅层特征的感受野较小,忽略了全局信息的应用。因此,本文以 Res2Net layer1_CA、Res2Net layer2_CA 及 Res2Net layer3_CA 的输出特征图 F_1, F_2, F_3 构建特征金字塔,输入多尺度特征融合模块 MSFF,以聚合层级的多尺度局部信息和全局信息。在对异源图像深层语义信息表达良好的同时,捕获一定程度的位置信息,提升配准精度,具体结构如图 3 所示。

由前述内容可知,浅层特征图 F_1 分辨率较高,通道数较少,包含有利于定位的特征信息,而特征图 F_3 分辨率较低,通道数较多,包含更多的区分性

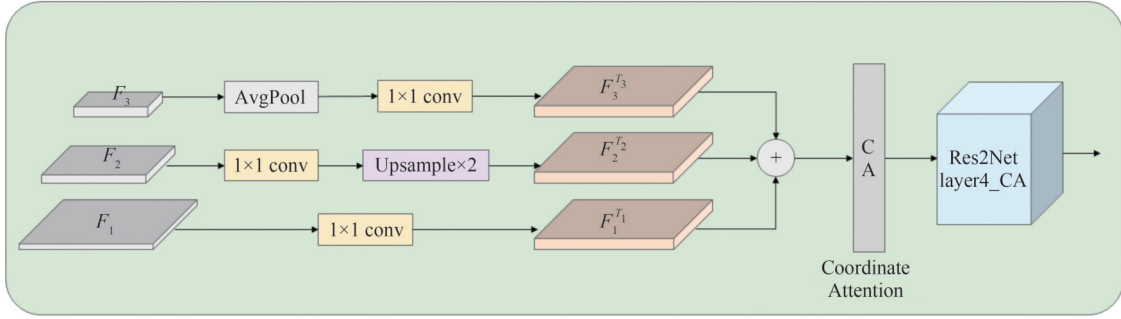


图3 多尺度融合模块(MSFF)

Fig. 3 Multi-scale feature fusion module

语义信息。因此,在融合的同时,为防止特征信息丢失及分辨率降低,分别对特征图 F_1, F_2, F_3 进行不同的变换得到分辨率与 F_1 相同,通道数与 F_3 相同的特征图 $F_1^{T_1}, F_2^{T_2}, F_3^{T_3}$ 。具体地, F_1 特征图利用 1×1 卷积扩展通道数得到特征图 $F_1^{T_1}$, F_2 特征图利用 1×1 卷积扩展通道数后进行两倍上采样得到特征图 $F_2^{T_2}$, F_3 特征图首先利用全局自适应平均池化得到全局特征信息,进行 1×1 卷积变换后扩展得到特征图 $F_3^{T_3}$ 。得到上述特征图后,将 $F_1^{T_1}, F_2^{T_2}, F_3^{T_3}$ 组合相加带来的增益十分有限,甚至更差,这是因为不同尺度层级的特征图之间存在语义鸿沟,其中的语义信息可能会互相干扰。为此,将特征图 $F_1^{T_1}, F_2^{T_2}, F_3^{T_3}$ 相加后的结果特征输入坐标注意力模块,再输入改进的 Res2Net layer4_CA,将前述的层级特征进一步深度融合。

如2.2节所述,用于特征提取的 Res2Net module 具有细粒度的多尺度特征表示能力,融合坐标注意力后嵌入精确的位置信息,而本节的 MSFF 通过融合不同层级的多尺度特征图,聚合局部和全局信息,保证了输出特征的分辨率,缓解了浅层定位信息的丢失。两者组合,提升了特征的识别能力和位置感知能力,为空间位置关系的回归映射提供有效依据。将 MSFF 输出的特征图进行全局自适应平均池化,然后送入参数回归网络,可得到最终的空间变换参数,利用空间变换模块,可获得配准后图像。

2.4 联合位置偏差与图像相似性的损失函数

光学与 SAR 图像间的同名点标注困难,真实的变换参数难以获取,因此对预先配准对齐的光学与 SAR 图像数据集进行随机仿射变换,以获取真值标签。具体地,假设光学参考图像为 I_r , 对应预先配准好的 SAR 图像为 I_s , 经过随机变换 $\Theta_{T \rightarrow S}$ 后的 SAR 图像为 I_s , 此处有 $\Theta_{R \rightarrow S} = \Theta_{T \rightarrow S}$, 则参考图像 I_r 与配准图像 I_s 的空间位置关系为:

$$(u', v', 1)^T = \Theta_{R \rightarrow S}(u, v, 1)^T, \Theta_{R \rightarrow S} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

其中 $(u, v) \in I_r, (u', v') \in I_s$ 为对应图像中像素点的坐标。在对网络参数进行迭代更新时,由于变换参数 Θ 混合了旋转、平移及尺度等分量,且这些分量有不同的量纲,若直接通过仿射参数真值 $\Theta_{S \rightarrow R}^{gt}$ 与网络预测的变换参数 $\Theta_{S \rightarrow R}^{pre}$ 之间的均方误差来进行优化,旋转、尺度与平移分量之间难以平衡,网络较难收敛。因而,通过借鉴 DHN 算法^[23]思想,对仿射参数进行四点参数化,设图像 I_r 的角点为 $C_r = \{(x_i, y_i), i = 1, 2, 3, 4\}$, 经仿射变换 $\Theta_{R \rightarrow S}$ 后得到对应点坐标 $C_s = \{(x'_i, y'_i), i = 1, 2, 3, 4\}$, 可计算出图像变换前后的4个角点位移量 C_{4pt} , 如式(7)所示,其中 $\Delta x_i = x'_i - x_i$ 。根据仿射参数 $\Theta_{S \rightarrow R}^{gt}$ 和 $\Theta_{S \rightarrow R}^{pre}$ 分别计算出 C_{4pt}^{gt} 和 C_{4pt}^{pre} , 并依据均方误差 MSE 来构建四角点偏移损失 $Loss_{4pt}$, 如式(8)所示。

$$C_{4pt} = \begin{bmatrix} \Delta x_1 & \Delta y_1 \\ \Delta x_2 & \Delta y_2 \\ \Delta x_3 & \Delta y_3 \\ \Delta x_4 & \Delta y_4 \end{bmatrix} \quad (7)$$

$$Loss_{4pt} = \frac{1}{8} \left\| C_{4pt}^{gt} - C_{4pt}^{pre} \right\|_F^2 \quad (8)$$

上述四角点偏移损失函数可以使网络收敛到一个较优值,对于自然图像的配准具有较高的精度,但本文配准任务的目标数据是存在显著非线性辐射差异的光学与 SAR 图像。因此,除损失函数 $Loss_{4pt}$ 外,以异源图像间的相似性度量为基础构建损失函数 $Loss_{NMI}$, 辅助配准网络在目标数据上更好地收敛,提升配准精度。在对异源图像计算相似性以完成配准任务时,互信息 MI 可以克服图像间的辐射差异,取得

较好的效果。然而由于MI受两幅图像的重合区域大小影响较大,MI值达到最大时无法保证配准结果正确。为使目标函数能更加准确反映MI值与配准效果之间的关系,采用归一化互信息(NMI)^[12]作为图像间的相似性度量,其相应数学表达为:

$$\text{NMI}(A, B) = \frac{-\sum_a p_A(a) \log_2 p_A(a) - \sum_b p_B(b) \log_2 p_B(b)}{-\sum_{a,b} p_{AB}(a, b) \log_2 p_{AB}(a, b)} \quad (9)$$

式中 $p_A(a)$ 、 $p_B(b)$ 分别为图像A、B的边缘概率分布, $p_{AB}(a, b)$ 为图像A与B的联合概率分布,图像A、B在文中代指的是参考图像与配准后图像。为进一步提高NMI值反映配准效果的稳健性,考虑到图像的配准方向,结合输入图像对 (I_R, I_S) 之间的双向转换,构建损失函数 Loss_{NMI} ,即:

$$\text{Loss}_{\text{NMI}} = e^{-[\text{NMI}(I_R, F_\theta(I_S)) + \text{NMI}(I_S, F_\theta^{-1}(I_R))]/2} \quad (10)$$

其中, F_θ 是待配准图像 I_S 向参考图像 I_R 配准的变换函数, F_θ^{-1} 是逆向变换函数。在获取网络输出的变换参数后,通过可微的空间变换模块 $\text{ST}^{[27]}$ 来对图像进行重采样,以便梯度可以反向传播。

总的损失函数是上述四角点偏移损失 $\text{Loss}_{4\text{pt}}$ 和图像间的相似性损失 Loss_{NMI} 之和,即:

$$\text{Loss}_{\text{total}} = \text{Loss}_{4\text{pt}} + \text{Loss}_{\text{NMI}} \quad (11)$$

3 实验结果与分析

对所提方法进行实验验证。3.1节和3.2节分别介绍了实验环境、实验数据以及评价指标,3.3节是本文方法的配准结果,与现有典型的算法进行了定性和定量的对比分析,验证所提方法优势。此外,3.4节通过对方法各部分设计进行消融实验,进一步验证所提部分的有效性。

3.1 实验数据与环境

实验采用公开的高分辨率光学与SAR图像配准数据集OS-Dataset^[29]进行网络的训练和测试。OS-Dataset数据集包含10692对图像(训练集8044对,验证集952对,测试集1696对),图像大小为 256×256 像素,主要包含城市、河流、农田地表起伏不大的场景。其中SAR图像来源于高分三号,分辨率为1m。光学图像收集于谷歌地球平台,并重新采样到1m分辨率。对原配准图像对中的SAR图像进行随机仿射变换,生成具有仿射变换关系的光学与SAR图像数据及相应的参数真值标签。其中,旋转角度范围为 $-30^\circ \sim 30^\circ$,尺度变换范围为

0.75~1.25倍原图宽高,平移范围为 $-30 \sim 30$ 像素。所提一体化配准方法基于Pytorch框架实现,采用AdamW优化器,权重衰减为 1×10^{-4} ,初始学习率为 2.5×10^{-4} ,批处理大小为12。所有实验均在处理器为i7-12700KF,显卡为GeForce RTX 4090的PC机上完成。

3.2 评价指标

本文从定性和定量两个角度评估算法的配准精度。在定性评估方面,以配准后的图像在参考图像中位置的边界线与真值边界线的重合程度为评价标准。边界线重合程度越高,说明所求得的变换参数越接近真值,图像对应位置越准,配准精度越高。

关于配准精度的量化评价指标,遵循文献[23],采用图像四个角点位移计算平均角点误差(Average Corner Error, ACE),ACE越小说明配准精度越高。由2.4节可知,依据变换参数 $\theta_{S \rightarrow R}^{\text{gt}}$ 和 $\theta_{S \rightarrow R}^{\text{pre}}$ 可计算出四角点偏移量 $C_{4\text{pt}}^{\text{gt}}$ 和 $C_{4\text{pt}}^{\text{pre}}$,则ACE的数学计算公式为:

$$\text{ACE} = \frac{1}{4} \sum_{i=1}^4 \sqrt{(\Delta x_i^{\text{pre}} - \Delta x_i^{\text{gt}})^2 + (\Delta y_i^{\text{pre}} - \Delta y_i^{\text{gt}})^2} \quad (12)$$

此外,为考察配准的实时性,以输入图像对到输出变换参数的运行时间作为量化指标,分析算法的配准效率。

3.3 配准结果与分析

为验证所提方法的配准性能,采用传统配准方法OS-SIFT^[5]与RIFT2(Speeding-up Radiation-variation Insensitive Feature Transform)^[30],以及两种深度配准模型DHN^[23]与DLKFM(Deep Lucas-Kanade Feature Map)^[31]进行对比。其中,OS-SIFT是专门为光学与SAR图像配准而设计的经典方法,该方法基于图像的梯度信息进行特征检测并完成配准;RIFT2基于相位一致性特征进行图像特征匹配,对于多模态遥感图像间存在的NRD具有较强的稳健性,实现了优越的异源配准性能;DHN是首次进行图像间几何参数估计的端到端深度学习方法,对于单模自然图像具有较好的性能;DLKFM是一种实现多模态遥感图像像素对齐的深度算法(该方法高精度的实现需要提供先验估计,本文以DHN的结果作为DLKFM的先验估计)。为保证结果的对比性,使用同一数据集在相同的硬件平台上对上述方法进行训练和测试。

3.3.1 配准结果定性评估

为体现方法的稳健性与泛化性能,从OS-Dataset数据集中挑选出不同场景的光学与SAR图

像进行可视化配准测试,场景类型包括城市区域、河流区域、农田区域、重复纹理区域及弱纹理区域,待配准图像对与相应的配准结果如图4所示。图中方框线表示配准后图像在参考图像中的对应位置,红色线为实验预测值,绿色为真值,两方框线重合程度越高,配准结果越精确。图4第1行是针对城市区域的配准;第2行是针对河流区域的配准;第3行是针对农田区域的配准;第4行是针对重复纹理区域的配准;第5行是针对弱纹理区域的配准。从图中可知,无论是对于因具有大量人造地物而导致非线性灰度差异较大的城市区域,还是纹理重复或几乎无纹理的区域,本文方法的预测方框线与真值线重合度最高,取得了相比于其他方法更好的效果。

从对比实验的结果中,可以看出基于特征点检测及匹配流程的传统方法效果较差。其中,OS-SIFT在所有测试图像中皆未能有效配准,这是因为依赖图像梯度信息的配准方法对于具有显著非线性差异的光学与SAR图像而言,鲁棒性较差。RIFT2基于相位一致性检测特征点并匹配,对于城市区域和普通农田区域可以得到相对较多的正确同名点对,实现配准。但对于重复纹理区域,RIFT2易产生大量误匹配点对;对于弱纹理区域,获取的可靠同名点对数量不足,因此配准性能皆大幅下降。基于深度学习的

配准方法对于不同场景区域有较好的适应性。DHN作为自然图像配准中端到端学习几何参数的经典方法,对于具有辐射差异的光学与SAR图像也有一定的泛化性,对于测试图像可进行粗略对齐,但无法得到精确对齐的结果。考虑应是DHN用于参数回归的深度特征为VGG网络输出的高层特征,丢失了用于精确定位的浅层特征信息。DLKFM是关于多模遥感图像对齐的深度学习算法,通过不同尺度的特征提取器提取特征,并由Lucas-Kanade算法迭代,逐步计算出不同尺度下的几何变换参数,由粗到精地实现图像对齐。如图4所示,DLKFM在DHN提供良好的先验参数时,可以取得较DHN更为精确的配准结果,但也存在不小的偏差。

为进一步验证所提方法的配准性能,对具有旋转、平移、尺度单项变换关系的光学与SAR图像对进行测试。图5给出了不同算法对于单项变换的配准结果。该部分展示的图像皆为含较多匹配同名点的普通场景区域,以有效展示传统基于点特征的配准算法效果,客观分析不同算法的性能。从图中可以看出,RIFT2对于图像的平移、旋转畸变具有较高的配准精度,但对于尺度变化较为敏感,因此可以判断影响RIFT2性能的因素除图像场景区域外,还有图像的尺度畸变。在对比的深度学习方法中,

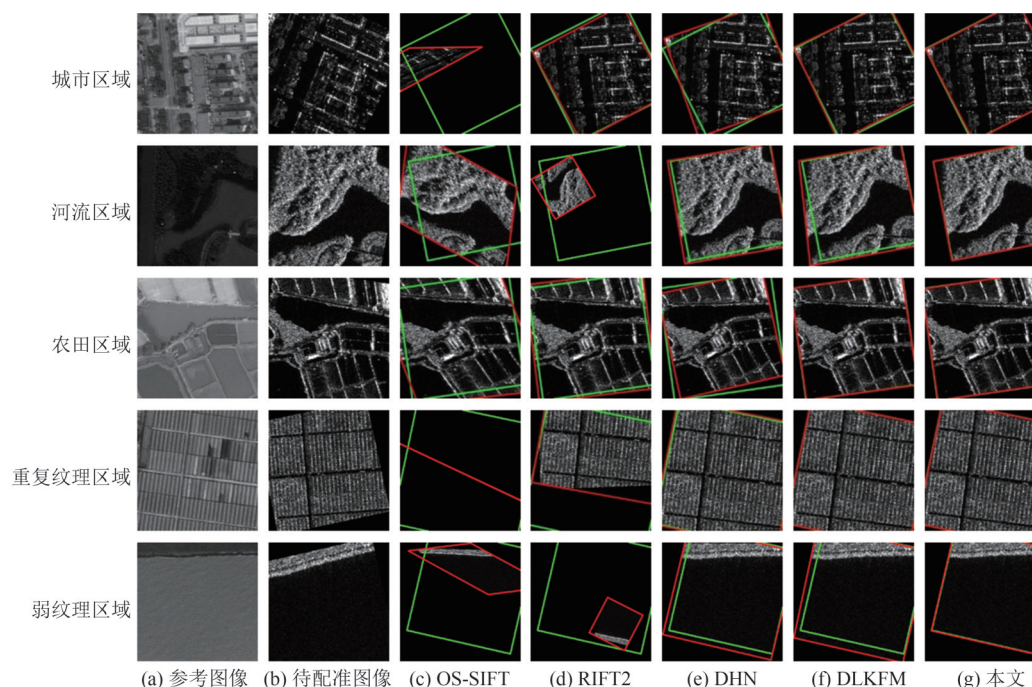


图4 不同算法在不同场景区域的配准结果示例(红框表示配准后图像位置的预测值,绿框表示配准后图像位置的真值)
Fig. 4 Examples of registration results of different algorithms in different scene regions (the red line represents the predicted value of the image position, and the green line represents the ground truth of the image position)

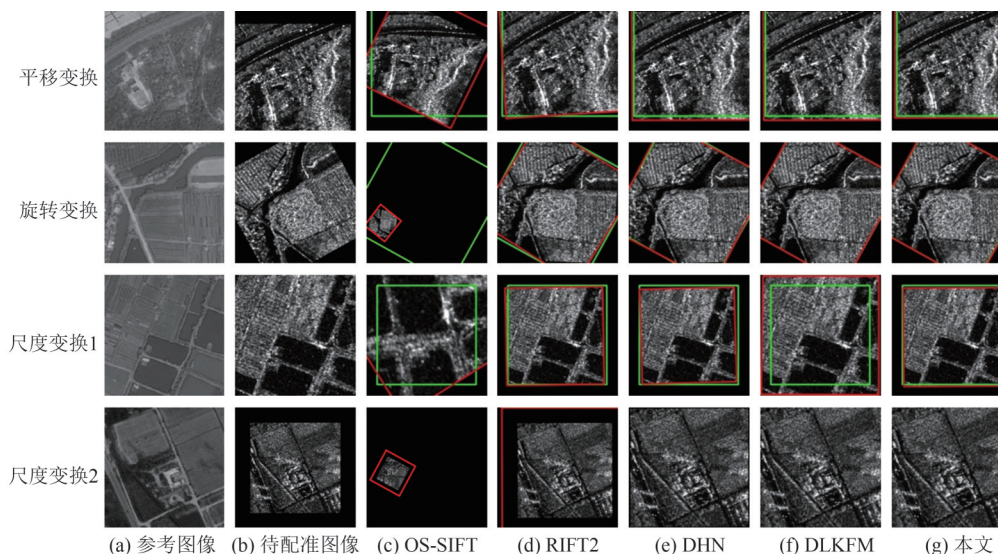


图 5 不同算法对于不同变换的配准结果示例(红框表示配准后图像位置的预测值,绿框表示配准后图像位置的真值)
 Fig. 5 Examples of registration results of different algorithms under different transformations (the red line represents the predicted value of the image position, and the green line represents the ground truth of the image position)

各方法对于旋转变换都具有较高的鲁棒性,但对于平移变换,皆存在少量的偏差,这可能是由卷积网络中的“平移不变性”所带来的影响。但从图 5 中可以看出本文方法的平移偏差相较于 DHN 有所减少,接近于真值。对于尺度变换,DLKFM 出现了明显的失效,说明其配准性能受到尺度畸变的影响, DHN 实现了大致对齐,但本文方法方框线重合度更高,配准结果更为精确。

3.3.2 量化指标对比

采用 3.2 节所述的评价指标 ACE 值对配准算法性能进行定量评估,表 1 和图 6 是不同算法对于 1696 对仿射变换测试图像的配准结果。以平均角点误差像素值范围为限,计算其占总测试集的百分比。从表 1 中可以看出,本文方法的配准精度领先于所有对比算法,平均角点误差在 3 pixel 以内的比例超出 DLKFM 算法 26.8%,即 25% 以上。由此可见,本文

方法具有良好的稳健性和较高的配准精度。

除配准精度外,配准速度也是关注的重点指标,因此对所有对比方法进行算法参数量和运行时间比较,运行时间为 1696 对测试图像配准耗时的平均值,结果如表 2 所示。相比于 RIFT2, OS-SIFT 的计算速度更快,原因是其检测出的有效同名点很少,大大降低了计算的复杂性,但同时也存在配准性能急剧下降的问题。在深度学习方法中, DLKFM 的模型参数量是最少的,但配准时间却最长,这是因为 DLKFM 在利用模型求取特征图后进行协方差计算,仍需要利用 Lucas-Kanade 算法来迭代出变换参数,影响了配准的实时性。本文方法的配准时间虽然将近 DHN 的两倍,但仍然处于一个非常小的量级,而本文方法的平均角点误差小于 3 pixel 的百分比比较 DHN 更高。这说明所提方法可以更好地提取稳健特征以提升配准精度,同时具备

表 1 不同算法的配准精度评估

Tab. 1 Precision evaluations of different registration algorithms

| 定量指标 | OS-SIFT ^[5] | RIFT2 ^[30] | DHN ^[23] | DLKFM ^[31] | 本文方法 |
|----------------|------------------------|-----------------------|---------------------|-----------------------|--------------|
| ACE<20 pixel/% | 2.24 | 34.61 | 87.03 | 87.38 | 99.12 |
| ACE<15 pixel/% | 1.30 | 26.24 | 80.31 | 81.96 | 98.64 |
| ACE<10 pixel/% | 0.41 | 15.86 | 68.51 | 74.00 | 96.70 |
| ACE<5 pixel/% | 0 | 3.48 | 47.52 | 63.97 | 91.86 |
| ACE<3 pixel/% | 0 | 0.94 | 26.65 | 58.61 | 85.44 |

注:加粗字体为最优结果。

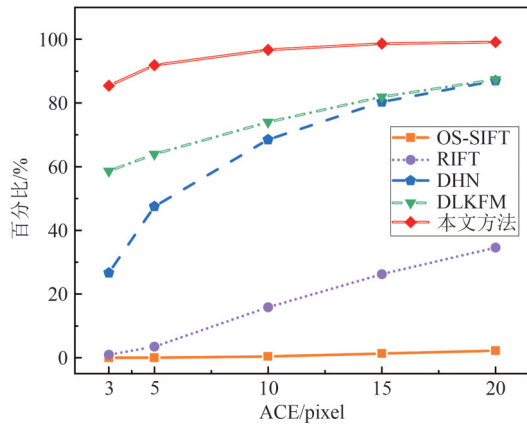


图6 不同算法的配准精度评估

Fig. 6 Precision evaluations of different registration algorithms

表2 不同算法的参数量与运行时间比较

Tab. 2 Running time and parameter comparison of different algorithms

| 定量指标 | OS-SIFT ^[5] | RIFT ^[30] | DHN ^[23] | DLKFM ^[31] | 本文方法 |
|----------------|------------------------|----------------------|---------------------|-----------------------|--------|
| Parameters/M | - | - | 514.44 | 7.39 | 200.41 |
| Mean of Time/s | 0.683 | 3.970 | 0.029 | 0.347 | 0.059 |

了配准任务所需的实时性。

3.4 消融实验

为验证所提方法中各部分的有效性,有针对性地设置相关消融实验。为保证结果的公平性,所有实验均在相同的参数设置及数据集上进行,评估结果如表3所示。表3第1行给出了使用Res2Net进行特征提取并回归的基准框架在测试数据上的平均角点误差;第2行是特征提取模块融合坐标注意力后的网络框架所得的配准效果,可以看出,平均角点误差小于10 pixel、5 pixel及3 pixel的比例相比于第1行均有所提升,说明通过坐标注意力嵌入位

表3 消融实验结果

Tab. 3 Results of the ablation study

| Methods | ACE<10 pixel/% | ACE<5 pixel/% | ACE<3 pixel/% | Mean of ACE/pixel |
|--|-------------------|------------------|------------------|----------------------|
| Res2Net module | 91.57 | 83.02 | 75.12 | 3.04 |
| Res2Net module+CA | 94.75 | 86.03 | 78.18 | 2.50 |
| Res2Net module+CA+ MSFF | 96.88 | 91.86 | 85.14 | 1.93 |
| Res2Net module+CA+ MSFF+Loss _{NMI} | 96.70 | 91.86 | 85.44 | 1.84 |

注:CA为坐标注意力机制,MSFF为多尺度特征融合模块,Loss_{NMI}为图像间的相似性损失。

置信息,可提升特征定位的精度;第3行为加入特征融合模块防止浅层信息丢失,并聚合全局信息和局部信息后的配准精度,ACE小于5 pixel和3 pixel的比例相比于第2行分别提高了5.83%和6.96%;最后一行是进一步引入相似性度量损失后的配准结果,ACE小于3 pixel的比例有所提升,测试数据的ACE均值降低了0.1 pixel左右。

实验结果证明,融合坐标注意力的特征提取模块、多尺度特征融合模块以及基于图像间相似性度量的损失函数等设计在一定程度上都能够提升配准精度,最终的配准精度较基准框架更是提升了9%左右,实现了高分辨率光学与SAR图像的高精度、端到端配准。

4 结论

针对存在显著几何差异和辐射差异的光学与SAR图像配准问题,本文提出了一种强化位置感知的一体化配准方法,直接通过深度网络回归变换参数,无须特征点检测与匹配,所提方法在传统方法容易失效的重复纹理区域与弱纹理区域,具有良好的稳健性。而通过深度网络对光学与SAR图像间的空间关系进行非线性映射时,仅利用高层特征或简单级联的多尺度特征来回归变换参数的方法配准效果不佳。本文提出结合坐标注意力的细粒度的多尺度特征提取模块,嵌入位置信息,并通过高效地融合定位信息丰富的浅层特征与语义信息丰富的深层特征,提高了特征定位精度,显著提升了配准性能。在公开的高分辨率数据集OS-Dataset上的实验表明,本文方法的配准精度明显高于其他四种代表性算法,且保证了较高的配准速度,为光学与SAR图像配准定位任务奠定了良好的基础。后续针对山区等存在局部畸变的地表起伏数据,所提方法将根据图像特点优化设计更为精细、自由度更高的配准模型以拟合数据的复杂变换。

参考文献

- [1] 张良培,何江,杨倩倩,等.数据驱动的多源遥感信息融合研究进展[J].测绘学报,2022,51(7):1317-1337. ZHANG Liangpei, HE Jiang, YANG Qianqian, et al. Data-driven multi-source remote sensing data fusion: Progress and challenges [J]. Acta Geodaetica et Cartographica Sinica, 2022, 51(7):1317-1337. (in Chinese)
- [2] 陈新,彭科举,周东翔,等.一种利用SAR和可见光图像融合检测目标的方法[J].信号处理,2010,26(9):

- 1408-1413.
- CHEN Xin, PENG Keju, ZHOU Dongxiang, et al. A fusion of target detection method from SAR image and optical image [J]. *Signal Processing*, 2010, 26(9): 1408-1413. (in Chinese)
- [3] LI Jiayuan, HU Qingwu, AI Mingyao. RIFT: Multimodal image matching based on radiation-variation insensitive feature transform [J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2019: 3296-3310.
- [4] NG P C, HENIKOFF S. SIFT: Predicting amino acid changes that affect protein function [J]. *Nucleic Acids Research*, 2003, 31(13): 3812-3814.
- [5] XIANG Yuming, WANG Feng, YOU Hongjian. OS-SIFT: a robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(6): 3078-3090.
- [6] JIANG Shuai, JZANG U, WANG Bingnan, et al. Registration of SAR and optical images by weighted sift based on phase congruency [C]//IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia, Spain. IEEE, 2018: 8885-8888.
- [7] 余先川, 吕中华, 胡丹. 遥感图像配准技术综述 [J]. *光学精密工程*, 2013, 21(11): 2960-2972.
- YU Xianchuan, LÜ Zhonghua, HU Dan. Review of remote sensing image registration techniques [J]. *Optics and Precision Engineering*, 2013, 21(11): 2960-2972. (in Chinese)
- [8] COLE-RHODES A A, JOHNSON K L, LEMOIGNE J, et al. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient [J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2003, 12(12): 1495-1511.
- [9] ZITOVÁ B, FLUSSER J. Image registration methods: A survey [J]. *Image and Vision Computing*, 2003, 21(11): 977-1000.
- [10] MA Jianglin, CHAN J C W, CANTERS F. Fully automatic subpixel image registration of multiangle CHRIS/proba data [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 48(7): 2829-2839.
- [11] SURI S, REINARTZ P. Mutual-information-based registration of TerraSAR-X and ikonos imagery in urban areas [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 48(2): 939-949.
- [12] KNOPS Z F, MAINTZ J B A, VIERGEVER M A, et al. Normalized mutual information based registration using k-means clustering and shading correction [J]. *Medical Image Analysis*. 2006, 10(3): 432-439.
- [13] KOVESI P. Phase congruency detects corners and edges [C]//The Australian Pattern Recognition Society Conference: DICTA. 2003.
- [14] YE Yuanxin, SHAN Jie, BRUZZONE L, et al. Robust registration of multimodal remote sensing images based on structural similarity [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(5): 2941-2958.
- [15] MERKLE N, AUER S, MÜLLER R, et al. Exploring the potential of conditional adversarial networks for optical and SAR image matching [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(6): 1811-1820.
- [16] DU Wenliang, ZHOU Yong, ZHAO Jiaqi, et al. K-means clustering guided generative adversarial networks for SAR-optical image matching [J]. *IEEE Access*, 2020, 8: 217554-217572.
- [17] ZHANG Han, NI Weiping, YAN Weidong, et al. Registration of multimodal remote sensing image based on deep fully convolutional neural network [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(8): 3028-3042.
- [18] HUGHES L H, MARCOS D, LOBRY S, et al. A deep learning framework for matching of SAR and optical imagery [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 169: 166-179.
- [19] 蓝朝桢, 卢万杰, 于君明, 等. 异源遥感影像特征匹配的深度学习算法 [J]. *测绘学报*, 2021, 50(2): 189-202.
- LAN Chaozhen, LU Wanjie, YU Junming, et al. Deep learning algorithm for feature matching of cross modality remote sensing images [J]. *Acta Geodaetica et Cartographica Sinica*, 2021, 50(2): 189-202. (in Chinese)
- [20] ZHANG Han, LEI Lin, NI Weiping, et al. Explore better network framework for high-resolution optical and SAR image matching [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-18.
- [21] ROCCO I, ARANDJELOVIC R, SIVIC J. Convolutional neural network architecture for geometric matching [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017: 6148-6157.
- [22] KIM D G, NAM W J, LEE S W. A robust matching network for gradually estimating geometric transformation on remote sensing imagery [C]//2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). IEEE, 2019: 3889-3894.
- [23] DETONE D, MALISIEWICZ T, RABINOVICH A. Deep image homography estimation [EB/OL]. 2016: arXiv: 1606.03798. <https://arxiv.org/abs/1606.03798>.
- [24] LE H, LIU Feng, ZHANG Shu, et al. Deep homography estimation for dynamic scenes [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 7649-7658.
- [25] BALAKRISHNAN G, ZHAO A, SABUNCU M R, et

- al. VoxelMorph: A learning framework for deformable medical image registration [J]. IEEE Transactions on Medical Imaging, 2019: 1788-1800.
- [26] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 13708-13717.
- [27] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks [EB/OL]. 2015: arXiv: 1506.02025. <https://arxiv.org/abs/1506.02025>.
- [28] GAO Shanghua, CHENG Mingming, ZHAO Kai, et al. Res2Net: A new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(2): 652-662.
- [29] XIANG Yuming, TAO Rongshu, WANG Feng, et al. Automatic registration of optical and SAR images via improved phase congruency model [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 5847-5861.
- [30] LI Jiayuan, SHI Pengcheng, HU Qingwu, et al. RIFT2: speeding-up RIFT with a new rotation-invariance technique [EB/OL]. 2023: arXiv: 2303.00319. <https://arxiv.org/abs/2303.00319>.
- [31] ZHAO Yiming, HUANG Xinming, ZHANG Ziming. Deep lucas-kanade homography for multimodal image alignment [C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021: 15945-15954.

作者简介



杨玉婷 女, 1998年生, 江西南昌人。国防科技大学电子科学学院 CEMEE 国家重点实验室硕士研究生, 主要研究方向为遥感图像配准。
E-mail: 2395482066@qq.com



赵凌君 女, 1981年生, 安徽安庆人。国防科技大学电子科学学院 CEMEE 国家重点实验室副教授、硕士生导师。主要研究方向为遥感信息处理、合成孔径雷达目标自动识别等。
E-mail: nudtzlj@163.com



赵路路 男, 1984年生, 山东东营人。中国人民解放军第95369部队工程师, 主要研究方向为遥感图像处理、信号分析等。
E-mail: 270498854@qq.com



张晗 女, 1989年生, 河南商丘人。西北核技术研究院助理研究员, 国防科技大学电子科学学院 CEMEE 国家重点实验室在读博士研究生, 主要研究方向为多源遥感影像配准技术。
E-mail: zhang.han@aliyun.com



熊博莅 男, 1981年生, 湖南益阳人。国防科技大学电子科学学院副教授, 电子信息系统复杂电磁环境效应 (CEMEE) 国家重点实验室副主任, 主要从事遥感图像解译、SAR 图像目标识别、图像配准及变化检测等领域研究工作。
E-mail: bolixiong@qq.com



计科峰 男, 1974年生, 陕西长武人。博士, 国防科技大学电子科学学院 CEMEE 国家重点实验室教授, 博士生导师, 主要研究方向为 SAR 图像解译、目标检测与识别、特征提取、SAR 和 AIS 匹配。
E-mail: jikefeng@nudt.edu.cn

(责任编辑: 刘建新)