

基于深度特征重构增强的光学和SAR图像鲁棒匹配

杨超, 刘畅, 唐腾峰, 叶沅鑫

西南交通大学 地球科学与工程学院, 成都 611756

摘要: 光学和合成孔径雷达SAR (Synthetic Aperture Radar) 图像由于成像原理的根本差异, 影像间呈现出显著的辐射和几何差异, 其自动精确匹配一直是当前国际学术研究的热点。目前, 大多数基于深度学习的匹配方法聚焦于图像深度特征提取, 但这些模型通常忽视了图像多尺度特征融合和共性特征表达, 导致鲁棒性不足, 难以应对复杂多变的地物场景。鉴于此, 本研究提出了一种基于深度特征重构增强的光学和SAR图像鲁棒匹配方法。该方法构建了一个融合多尺度深度特征和图像特征重构的伪孪生共有特征提取网络。首先, 通过多尺度特征提取架构, 网络能够在像素级高效提取光学和SAR图像对的多尺度深度特征。其次, 为光学图像设计了一个伪SAR图像翻译分支结构, 利用深度特征重构图像, 以增强网络学习更鲁棒的共性特征表达能力。最后, 构造了基于多层特征匹配相似度和重构影像平均误差的联合损失函数, 实现了光学和SAR图像的鲁棒匹配。在两种具有不同分辨率且包含多种地物场景(城市、郊区、荒漠、山地、水体)的遥感影像数据集上进行对比实验, 结果表明, 与当前几种最先进的匹配方法相比, 所提出的方法均表现出最高的正确匹配率。

关键词: 遥感, 光学图像, 合成孔径雷达图像, 影像匹配, 深度学习

中图分类号: P237/P2

引用格式: 杨超, 刘畅, 唐腾峰, 叶沅鑫. 2025. 基于深度特征重构增强的光学和SAR图像鲁棒匹配. 遥感学报, 29(8): 2616-2626

Yang C, Liu C, Tang T F and Ye Y X. 2025. Robust matching of optical and SAR images based on deep feature reconstruction enhancement. National Remote Sensing Bulletin, 29(8): 2616-2626 [DOI: 10.11834/jrs.20254295]

1 引言

多模态遥感图像配准任务旨在将来自不同传感器、不同视角或不同拍摄时间的图像进行对齐, 精确匹配图像是实现这一目标的关键, 也是后续联合应用的基础(李星华等, 2023)。光学图像和合成孔径雷达SAR (Synthetic Aperture Radar) 图像作为最常用的两类遥感图像, 广泛应用于图像融合(Li等, 2023; 成飞飞等, 2022)、变化检测(叶沅鑫等, 2023)、目标识别(Wang等, 2021)、飞行器视觉导航(尹奎英等, 2021)等领域。光学图像因丰富的纹理和颜色信息易于理解, 但其成像质量易受到云层遮挡和光照条件等环境因素影响。相比之下, SAR则能有效弥补光学成像系统的不足, 具备全天候观测能力, 但同时SAR影像由于自身成像特性不可避免地会受到斑点噪声

干扰。由于光学和SAR图像的成像特点和图像信息具有互补性, 它们的联合应用可以为成像区域提供有价值的互补信息。

现代遥感系统利用精密的轨道传感器技术, 能够获取具有高地理定位精度的遥感图像。通过预处理, 可以消除如旋转和比例变化等大型全局几何变换, 使得图像间的差异缩减至几十个像素内的平移变换(Merkle等, 2017)。因此, 光学与SAR图像匹配的研究重点在于解决有限的平移估计问题。然而, 由于图像获取特性的差异(图1), 同一场景下光学图像和SAR图像之间存在显著的非线性辐射畸变, 这极大地增加了匹配难度。为了从高度异构的图像对中建立可靠的特征描述或有效的相似性度量, 研究人员主要采用了两类方法: 一是基于手工特征的匹配方法, 二是利用深度学习的匹配方法。

收稿日期: 2024-07-17; 预印本: 2025-04-15

基金项目: 国家重点研发计划(编号: 2024YFC3015402); 国家自然科学基金(编号: 42271446); 四川省中央引导地方科技发展专项(编号: 2024ZYD0084)

第一作者简介: 杨超, 研究方向为遥感图像处理与分析。E-mail: yc18483685462@163.com

通信作者简介: 叶沅鑫, 研究方向为多源遥感图像智能解译。E-mail: yeyuanxin@home.swjtu.edu.cn



图1 同一区域的光学与SAR图像

Fig. 1 Optical and SAR images of the same area

基于手工特征的多模态图像匹配方法主要分为两类：基于特征的方法和基于区域的方法。基于特征的匹配方法通常包括关键点检测、特征描述和特征匹配3个步骤。在光学图像与SAR图像的异源图像匹配中，由于存在显著的非线性辐射差异，研究人员提出了多种解决方案。例如，何梦梦等（2018）针对不同的遥感图像来源使用不同的特征提取方法（光学图像使用Canny边缘提取算子，SAR图像使用Ratio Of Averages算子），并将边缘线特征转化为点特征，提高了配准精度。Li等（2020）针对非线性辐射差异提出了辐射不变特征变换RIFT（Radiation-Variation Insensitive Feature Transform）算法，该算法利用相位一致性进行特征检测，在多种类型图像上均表现良好。随后，众多研究（Gao等，2022；Li等，2022；Zhu等，2023）通过构建多尺度特征检测或优化特征描述图，进一步提升了异源图像特征匹配的性能。基于区域的方法通过模板匹配来测量图像间的相似性。其中，互信息MI（Mutual Information）（Suri和Reinartz，2010）方法在过去几十年中得到了广泛应用。近年来，Ye和Shen（2016）率先指出，基于结构特征的模板匹配方法能有效处理多模态图像间的非线性辐射差异，并进一步（Ye等，2019）提出了名为方向梯度通道特征CFOGs（Channel Features of Orientated Gradients）的快速稳健匹配方法，该方法通过逐像素构造梯度方向描述子来增强对图像结构特征的描述，并利用快速傅立叶变换FFT（Fast Fourier transform）技术来定义基于频域特征表示的相似性度量，从而提高匹配效率。在CFOGs的基础上，Fan等（2021）和Xiang等（2023）通过改进描述子构造，进一步提升了匹配精度。尽管如此，鉴于地物场景的复杂性，手工设计的特征在进一步提升匹配性能

方面的潜力受限。

随着深度学习技术的快速发展，基于学习的匹配方法展现出卓越的性能，逐渐成为图像匹配领域的研究热点。针对图像配对的需求，孪生网络（Siamese Network）和三元组网络已成为卷积神经网络CNN（Convolutional Neural Network）描述符学习领域的主要架构。这些网络通常由两部分组成，即特征提取和相似性度量，通过学习图像的高级语义特征进行图像匹配。Merkle等（2017）首次尝试应用孪生网络学习光学和SAR图像之间的共有特征。在这之后，学者们采用伪孪生CNN架构，专注于像素级深度特征学习，通过计算输出特征的点积来评估输入图像对的相似性（Zhang等，2022a；Zhou等，2022）。Zhang等（2022b）进一步探索了高分辨率光学和SAR匹配的通用网络框架，通过优化特征提取、通道注意模块和自适应加权损失3个方面来提升匹配性能。Xiang等（2022）提出了一种基于特征解耦网络的方法，结合伪孪生全卷积网络和残差去噪网络，有效降低了SAR图像的非线性辐射失真和噪声干扰。近年来，不少学者通过改进网络结构和损失函数，进一步提升了光学和SAR图像的匹配精度（Gazzea等，2023；Ye等，2024）。尽管如此，大多数基于深度学习的匹配方法仅依赖于特征间的距离度量来引导网络学习共性特征，却忽视了多尺度特征融合和特征翻译重构的重要性，这限制了模型学习高质量、广泛适用的共性特征描述的能力。因此，如何从光学和SAR影像中有效地提取出通用且鲁棒的特征描述，已成为亟待解决的科研问题。

综上所述，基于深度学习的光学和SAR图像匹配技术面临若干关键挑战。光学和SAR影像之间存在显著的几何畸变和非线性辐射差异，这对网络学习鲁棒的共性特征构成了重大挑战。此外，不同地物类型的遥感影像特征差异显著，这对算法模型的普适性和准确性提出了更高的要求。因此，开发能够在多种遥感图像地物场景下提取通用且鲁棒的共性特征描述的网络模型，对于提升光学与SAR图像匹配算法的准确性和鲁棒性至关重要。

针对以上问题，本文提出了基于深度特征重构增强的光学和SAR图像鲁棒匹配方法（命名为OS_DFRENet）。本文的主要贡献如下：（1）基于U-Net主干网络，构建了一个伪孪生密集特征提取

网络框架,用于分别获取多尺度深度特征和特征翻译图像。(2)通过网络输出多尺度特征图和伪SAR重构影像,定义了联合多层特征匹配相似度和影像重构损失的新型损失函数,以增强网络学习更通用、更鲁棒的共性特征描述能力。(3)本文在含有多种不同地物类型的光学和SAR影像数据集上进行了广泛测试。测试结果表明,所提出的方法在匹配精度和鲁棒性上优于其他具有代表性的匹配方法。

2 本文方法

2.1 基于深度特征重构增强的匹配框架

在视觉表征学习领域,神经网络架构的创新起着至关重要的作用。卷积神经网络通过引入一种通用的特征学习范式,适用于多种视觉识别任务,取代了以往依赖手动特征工程的方法。CNN已在计算机视觉、生物学、医学成像以及遥感等多个学科领域内展示了其巨大的应用潜力。

U-Net是一种广受认可的语义分割模型,采用对称的编码器—解码器架构设计(Ronneberger等,2015)。编码器通过下采样将输入图像转换为一组特征图,而解码器则通过上采样将这些特征图恢复到输入图像的原始尺寸。通过跳跃连接传递的细节信息有助于克服编码器在输出时面临的分辨

率降低问题,从而有效保留了图像的定位信息。此外,U-Net不受输入图像尺寸的限制,突破了传统卷积神经网络在处理不同尺寸图像数据集时的局限。

针对光学和SAR图像匹配任务,传统算法常常依赖于提取基于结构相似性的逐像素特征描述符。鉴于此,本研究旨在构建一种基于U-Net主干网络的伪孪生密集特征提取匹配框架(图2)。该匹配框架主要由两大部分构成:特征提取部分和特征度量学习部分。详细步骤如下:(1)网络分别将光学图像作为参考图、SAR图像作为模板图,图像均以灰度图形式输入。(2)主干网络包含两个不共享权重的伪孪生U-Net网络分支,这有助于将光学和SAR图像这两种高度异构的数据转换到统一的高维空间,同时避免了影像流之间的相互干扰。(3)在获取光学和SAR图像的多尺度逐像素深度密集特征后,本文为正负样本自适应分配权重,并采用快速傅里叶变换加速的差平方和FFT-SSD(the sum of squared differences accelerated by fast Fourier transform)来衡量特征对之间的相似性。(4)本文将光学图像最后的特征层接入重构网络分支,并构造了伪SAR图的平均误差损失。网络在特征重构损失约束下学习多尺度特征的相似性度量,增强了提取鲁棒特征描述的能力。接下来将详细介绍实施的具体细节。

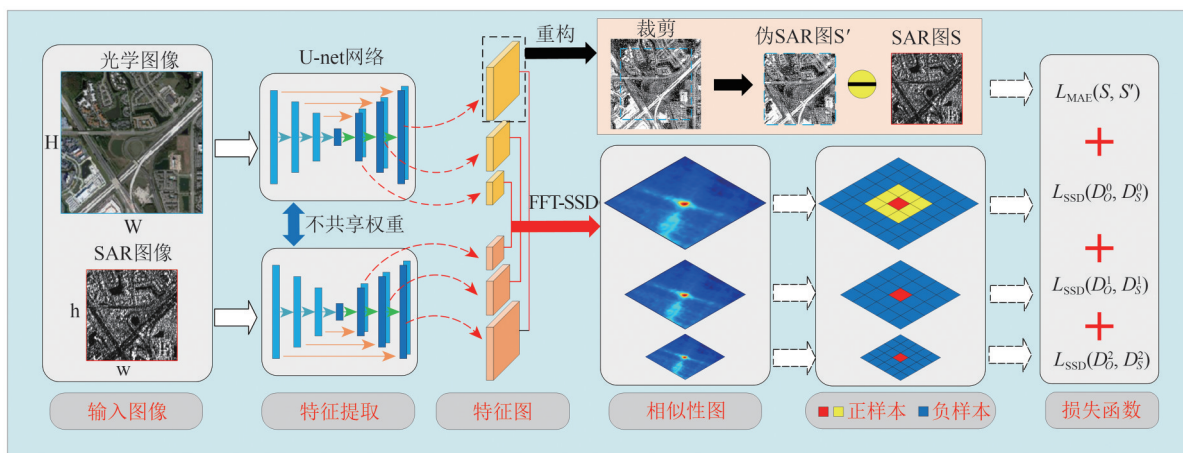


图2 基于深度特征重构增强的匹配网络框架

Fig. 2 The matching network framework based on deep feature reconstruction enhancement

2.2 多尺度特征提取与特征重构

本文所提出的特征提取网络如图3所示,选择U-Net作为特征提取的主干网络,并针对光学和

SAR图像匹配任务对网络结构进行了适当的改进。首先,由于光学与SAR图像间存在巨大的非线性辐射畸变,因此采用伪孪生结构有助于从高度异

构的图像中提取共有特征。伪孪生架构包含两个相同的网络分支，但具有不同的网络参数，这有利于将来自高度异构的光学图像和SAR图像的特征嵌入到同一个高维空间，也有助于防止图像特征之间的相互干扰。伪孪生U-Net提取特征的计算公式可以表达为

$$\begin{aligned} o &= U(O; \theta) \\ s &= U(S; \theta') \end{aligned} \quad (1)$$

式中， $U(O; \theta)$ 和 $U(S; \theta')$ 分别表示光学图像 O 和SAR图像 S 输入到具有不同权重参数 θ 和 θ' 的U-Net网络中。 o 和 s 分别表示对应的特征图。

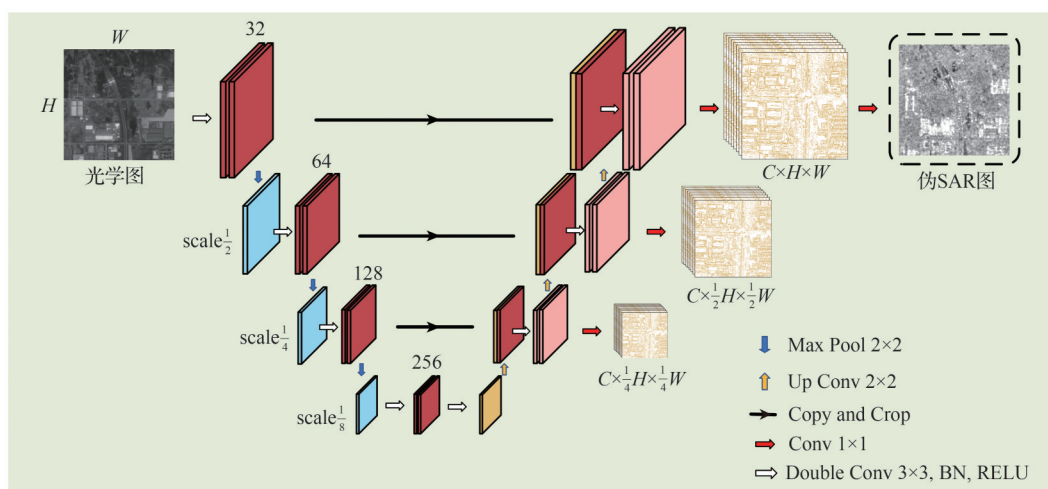


图3 网络结构图

Fig. 3 The network structure diagram

此外，常用的遥感影像尺寸为 256×256 。为了适应图像多尺度特征提取，编码和解码阶段的最后一个卷积模块都被移除，即网络共包括三次下采样和三次上采样过程。这种改进有助于更好地平衡深层语义信息和浅层细节信息。为了融合多层次尺度特征，在解码器部分，与常见的单一特征层输出不同，本文的网络设计允许同时输出多个尺度的图像特征图，从而实现了由粗到精的特征匹配策略。具体来说，小尺寸的图像特征反映更抽象、更本质的图像共有特征，而大尺寸的图像特征则反映更精细的纹理结构特征。通过分别对不同尺度的图像特征进行相似性度量，匹配预测位置由粗到精。这样的多尺度特征融合策略使网络能够学习到更鲁棒的图像多级特征。这种改进不仅增强了模型对不同尺度特征的捕捉能力，而且提高了匹配算法在复杂遥感影像数据上的准确性和鲁棒性。

本文受生成对抗网络 (Isola 等, 2017) 的启发，在训练过程中引入了特征重构分支，以引导网络学习光学和SAR图像间更稳健的共有特征。如图3所示，该分支与特征提取网络共享权重参

数，并通过卷积层与光学图像网络解码器的最后一层相连，以生成尺寸与输入图像相同的重构图像。该分支实现了从光学图像到SAR图像的翻译架构，促使网络所提取的特征能够重构原始的SAR图像。这种特征重构方法不仅增强了模型对不同类型遥感图像特征的学习能力，而且通过实际重构SAR图像，为模型提供了一种有效的监督信号，有助于网络更好地捕捉和表达光学与SAR图像间的共有特征。通过这种方式，网络能够在有限的训练样本下学习到更通用的特征描述，从而提高匹配算法在多场景遥感数据中的性能。

在训练过程中，通常将光学图像作为参考图，SAR图像作为模板图。由于模板图像在参考图像上的位置已知，可以裁剪伪SAR图像以匹配模板尺寸，从而计算伪SAR图像与模板SAR图像之间的平均误差。利用此误差作为损失函数，网络学习到的深度特征表达得以优化，从而增强了模型的鲁棒性和泛化能力。

2.3 匹配度量学习

在提取了SAR与光学图像的共有特征之后，紧接着的一个关键步骤是构造特征间的相似性度

量。由网络结构可知,网络共获取了3对不同尺寸的光学和SAR特征图,其维度分别为 $W \times H \times C$ 和 $w \times h \times C$ 、 $1/2W \times 1/2H \times C$ 和 $1/2w \times 1/2h \times C$ 、 $1/4W \times 1/4H \times C$ 和 $1/4w \times 1/4h \times C$ 。这些特征图分别代表了不同尺度的图像信息,其中, WH 和 wh 分别代表光学图像 O 和SAR图像 S 的宽和高, C 表示通道维度。在计算三维特征图之间的相似性时,采用传统的逐像素滑动策略来计算差平方和SSD (the Sum of Squared Differences) 十分耗时。因此,本文引入了快速傅里叶变换来加速传统的差平方和运算。通过将图像三维特征图转换到频率域,可以将空间域上的模板滑动操作转换为频率域上的点乘,从而实现快速计算。SSD与FFT-SSD的计算公式如下所示:

$$S(\nu) = \sum_i (D_s(i) - D_o(i - \nu)T(\nu))^2 \quad (2)$$

$$S'(\nu) = F^{-1}(F^*(D_o) \cdot F(D_o T))(\nu) - 2 \times F^{-1}(F^*(D_o) \cdot F(D_s T))(\nu) \quad (3)$$

式中, $S(\nu)$ 和 $S'(\nu)$ 分别表示SSD和FFT-SSD所计算的特征相似度热力图, S 中的最小值所在位置表示模板图 T 在对应参考图上的位置。 D_o 和 D_s 分别表示由网络提取的光学和SAR图像的特征图。 F 和 F^{-1} 分别表示傅里叶正变换和逆变换, F^* 表示 F 的复共轭。

在利用FFT-SSD计算得到参考图和模板图特征图之间的相似性热力图后,训练样本挖掘策略对网络性能具有重要影响。简单地考虑所有正负训练样本并非区分特征的最佳选择。三元组损失采用硬负样本挖掘方法,提高了对所学特征的辨别能力。但是,它仅依赖于小部分数据,可能不足以全面描述嵌入空间的全局几何特征,且容易出现过拟合现象。因此,本文根据Zhang等(2022b)的工作引入了自适应样本权重损失。该方法为每个训练样本分配一个动态权重,使得每个样本对相似性得分都有一个自适应的惩罚强度。对于特征距离远离最佳值的样本对,应使用较大的加权系数,因此这些训练样本就会受到较大梯度值的惩罚。反之,对于已经接近最优值的样本,权重系数则较小。

考虑到像素级匹配精度和硬负样本挖掘策略,对于FFT-SSD得分图 $S'(\nu)$ (图2),本文将标签中的像素及其相邻像素设置为正样本,而其余像素设置为负样本。而对于尺度更小的得分图,则仅

将标签位置设置为正样本。值得注意的是,在计算FFT-SSD得分之前,需要学习到的深度密集特征 D 进行L2归一化处理。本文计算特征相似性的损失函数公式如下:

$$L_{SSD} = \log \left(1 + \sum_{i=1}^{K_n} \sum_{j=1}^{K_p} \exp \left(\gamma \left(\omega_n^i (y_n^i - \Delta_n) - \omega_p^j (y_p^j - \Delta_p) \right) \right) \right) \quad (4)$$

式中, y_p 和 y_n 代表正负样本特征对间的相似度, Δ_p 和 Δ_n 分别代表正负样本的边际值, ω 和 γ 分别代表自适应权重因子和尺度因子。

此外,本文还计算了由光学图像特征重构所生成的伪SAR图和原始SAR影像之间的平均绝对误差MAE (Mean Absolute Error),其损失函数公式如下:

$$L_{MAE} = \frac{1}{w \times h} \sum_{i=1}^{w,h} |S_{ij} - S'_{ij}| \quad (5)$$

式中, w 和 h 代表SAR图像的宽和高, S 和 S' 分别代表原始SAR图像和伪SAR图像。

综上所述,网络共得到了3个尺度的相似性损失和一个平均误差损失。因此,本文所构造的联合损失函数如下:

$$L = \lambda_0 L_{SSD}(D_o^0, D_s^0) + \lambda_1 L_{SSD}(D_o^1, D_s^1) + \lambda_2 L_{SSD}(D_o^2, D_s^2) + \lambda_3 L_{MAE}(S, S') \quad (6)$$

式中, D_o 和 D_s 代表光学和SAR图像的特征图,其上标0、1、2分别表示原始尺寸、宽和高缩小1倍和缩小2倍的特征图。 λ_n 则表示不同的超参数。

3 实验设计与结果分析

为了验证本文方法的有效性,本节将所提出的匹配算法与其他具有代表性的光学和SAR图像匹配算法进行对比分析。这些方法包括手工的MI算法(Suri和Reinartz, 2010)和CFOG算法(Ye等, 2019),以及基于学习的DDFN(Zhang等, 2022a)、OSMNet(Zhang等, 2022b)、FFT U-Net(Fang等, 2022)和MoPSI(Liu等, 2023)。本文通过正确匹配率CMR (Correct Matching Rate)进行精度评定,正确匹配率是指与标签真值的L2距离小于特定阈值 T 的匹配图像对所占的百分比。实验数据平台处理器为NVIDIA GeForce RTX 3070 GPU,训练批次大小设置为8,采用AdamW梯度下降优化器,初始学习率设置为0.001,权重衰减设置为 $5E-2$ 。

3.1 实验数据

本研究共采用了两个数据集，分别是公开的SEN1-2数据集和自制的谷歌地球光学图像与高分三号SAR图像数据集（GE-GF3）。以下是对这两个数据集的具体介绍。

SEN1-2数据集最初由Schmitt等（2018）公开，是由Sentinel-1和Sentinel-2卫星获取的大型SAR和光学数据集，分辨率为10 m。该数据集的位置分布遍及全球，森林、沙漠和湖泊等地物覆盖类型占比较大。此外，SEN1-2包含春、夏、秋、冬不同季节的数据，且相邻的两幅图像（256×256像素）的重叠率为50%。由于配准精度低，Xu等（2023）进一步精确配准了该数据集，并公开了一个尺寸为512×512的高精度数据集。本文在此基础上裁剪为256×256大小的数据集，从每张512×512像素的图像中裁剪出5张图像（4张无重叠图像和1张中心区域图像），共得到了20430对图像。

本文自制的GE-GF3数据集是由谷歌地球光学影像和高分三号（GF-3）生成的SAR影像制作而成的，空间分辨率为5 m，拍摄时间为2021年—2022年。光学影像来源于多时相多卫星数据，相邻区域成像色差大，这将进一步增加匹配难度。为了评估匹配算法的鲁棒性和泛化能力，本文共收集了10个不同区域的影像，每幅影像的长宽约1—2万像素。研究区域经纬度分布广泛，土地覆盖类型多样，包含城市、郊区、荒漠、山地、水体等多种地物类型。由于所采集的原始数据地理信息未完全对齐，影像间配准误差在几个像素到几十个像素之间。因此，对于每一对影像，本文通过人工刺点的方式在整幅大图上选取了大量均匀分布的控制点和检查点，然后以SAR影像为参考，拟合投影变换模型对光学影像进行了精确配准，使整幅大图的平均配准精度控制在约1.5个像素以内。最后，裁剪全部影像对的共同区域，制作为256×256大小的非重叠影像块，并从中选取了共30000对影像。

两个数据集均按照70%训练集、20%测试集和10%验证集的比例进行了划分。在训练过程中，将SAR影像设定为模板图像，尺寸设置为128×128像素；而光学影像作为参考图，尺寸为192×192像

素。模板图像确保完全包含在参考图像的范围内，搜索区域被设定为 $[-32, 32]$ 像素。由于图像块的位置是随机选取的，网络在每一轮训练中都会接收到不同的地面实况图输入。这种策略不仅扩充了训练数据集，还有助于提高网络的泛化能力，避免过拟合现象。

3.2 对比实验

3.2.1 定量分析

在本文中，为了全面评估所提出匹配算法的性能，选取了5种具有代表性的光学和SAR图像匹配方法进行比较。其中，基于学习的方法均是针对光学和SAR匹配设计，在本文实验中无需进行结构调整。为了确保公平性，所有模型均采用相同的训练策略和最优的参数设置，并在验证集中选取各自精度最高的轮次，统一比较在测试集中的表现。本节对各方法在两种数据集的测试集上进行性能评估，统计了0—5个像素阈值内的正确匹配率。

图4（a）和图4（b）分别展示了7种方法在SEN1-2和GE-GF3数据集上的实验结果。从图4中可见，基于学习的方法匹配精度显著优于传统手工方法，这表明网络学习的特征在表达图像共有特征方面远超手工设计特征。MI直接统计图像像素值间的互信息来构造相似性度量，在灰度差异显著的光学和SAR图像上表现较差。CFOG通过方向梯度提取图像结构特征，依赖丰富的影像结构信息，但提取的共有特征不够稳健，难以适应复杂地物类型的影像数据。

需特别说明的是，对比两组数据集上匹配方法的表现差异，可见匹配算法在SEN1-2数据集上的5像素内正确匹配率明显优于GE-GF3数据集。分析两组数据集的差异，可以得到以下结论：首先，SEN1-2数据为10 m低分辨率图像，官方数据表明其成像具有较高的地理定位精度。Xu等（2023）所公开的精配准数据经过筛选，剔除了弱纹理、弱特征等匹配难度大的区域，因此在SEN1-2数据集上，匹配算法的5像素内正确率可达90%以上。而GE-GF3数据集选用的是谷歌地球光学影像，该影像由多时相多卫星数据拼接而成，相邻区域成像色差大，影像间相对差异大。

为了更全面评估算法性能,本文选取了多种地物场景,特别是包含大量匹配难度极大的荒漠、山地区域。此外,由于图像分辨率为更高的5 m,且图像模板设置为128×128的小尺寸,图像信息量有限,特征提取难度大。综合以上因素,匹配算法

在GE-GF3数据集上的正确匹配率相较于前者下降了20%以上。这说明训练数据的成像质量、分辨率、场景类别等因素,都会极大地影响算法的匹配性能。

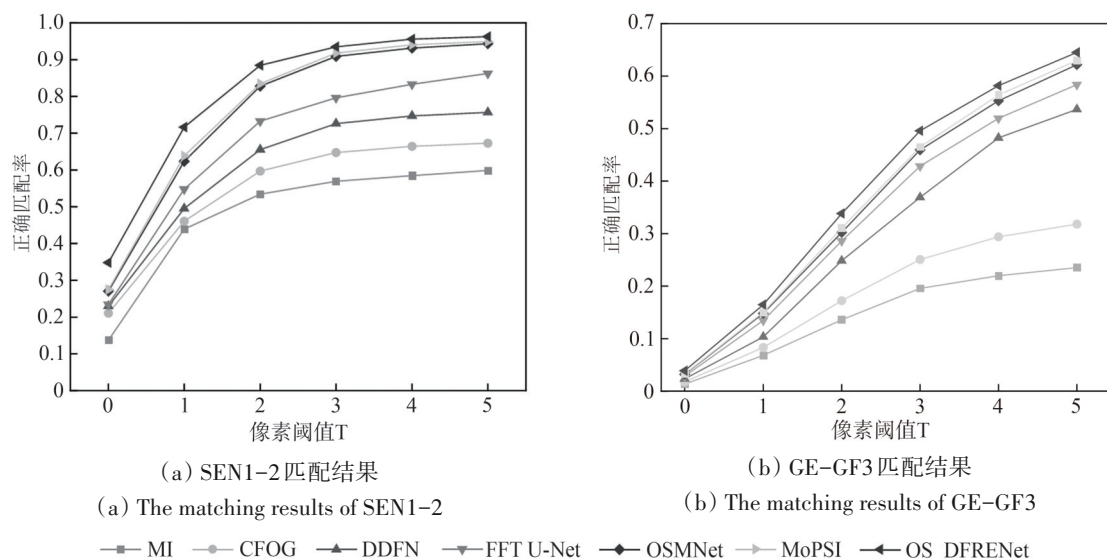


图4 不同匹配方法的对比实验

Fig. 4 Comparative experiment of different matching algorithms

尽管近年来基于学习的匹配方法已取得不错的匹配性能,但本文提出的OS_DFRENNet方法在0—5像素阈值上均实现了最高的正确匹配率。相较于CFOG等经典的传统方法,本文方法在正确匹配率上平均高出约30%。与其他深度学习方法对比,FFT U-Net通过孪生U-Net网络提取异源图像共有特征,但特征流间共享参数不利于共性特征表达。DDFN仅构造连续的卷积块提取特征,忽视了多尺度深度特征。OSMNet在DDFN基础上对特征提取、注意力增强和损失函数3个方面进行了优化,取得了更好的匹配性能。MoPSI使用LOG-Gabor滤波提取图像多方向相位结构信息,构造多尺度融合的U-Net7网络提取多尺度卷积特征,实现了次优的正确匹配率。而本文提出的OS_DFRENNet通过伪孪生的多尺度深度特征提取网络,从不同尺度下提取共有特征,并构建了SAR特征重构分支以增强特征表达,能够提取异源图像间更鲁棒的共有特征。图4显示,本文方法在两组数据集上均获得了最高的正确匹配率,这表明了其在像素级匹配精度上的显著优势。此外,由于训

练和测试数据涵盖了多种地物类型,在有限的训练数据条件下,本文模型相较于其他学习方法展现出更强的鲁棒性,这证明了其在不同遥感图像数据上具备泛化能力。

3.2.2 定性分析

为了定性评价不同方法所学习特征的鲁棒性,本节展示了7种方法所生成的特征相似性热力图。如图5所示,从左到右,本文从GE-GF3数据中选取了包含城市、郊区、荒漠、山地、水体等5种不同地物场景的影像,其中红色虚线框代表参考图的图像范围,黄色虚线框代表模板图的图像范围。从上到下,分别展示了7种方法计算得到的相似性热力图。由于采用SSD策略的预测位置为得分图的最小值,为便于直观对比观察,展示时进行了取反操作。参考图和模板图的尺寸分别为192和128,因此热力图尺寸为搜索域范围65×65,标签真值理论上为图像的中心点,即图5中红色点所示。热力图上颜色越亮的区域表示特征相似性得分越高,反之颜色越深表示特征差异越大,匹配算法的预测值标记为图5中黑色点。

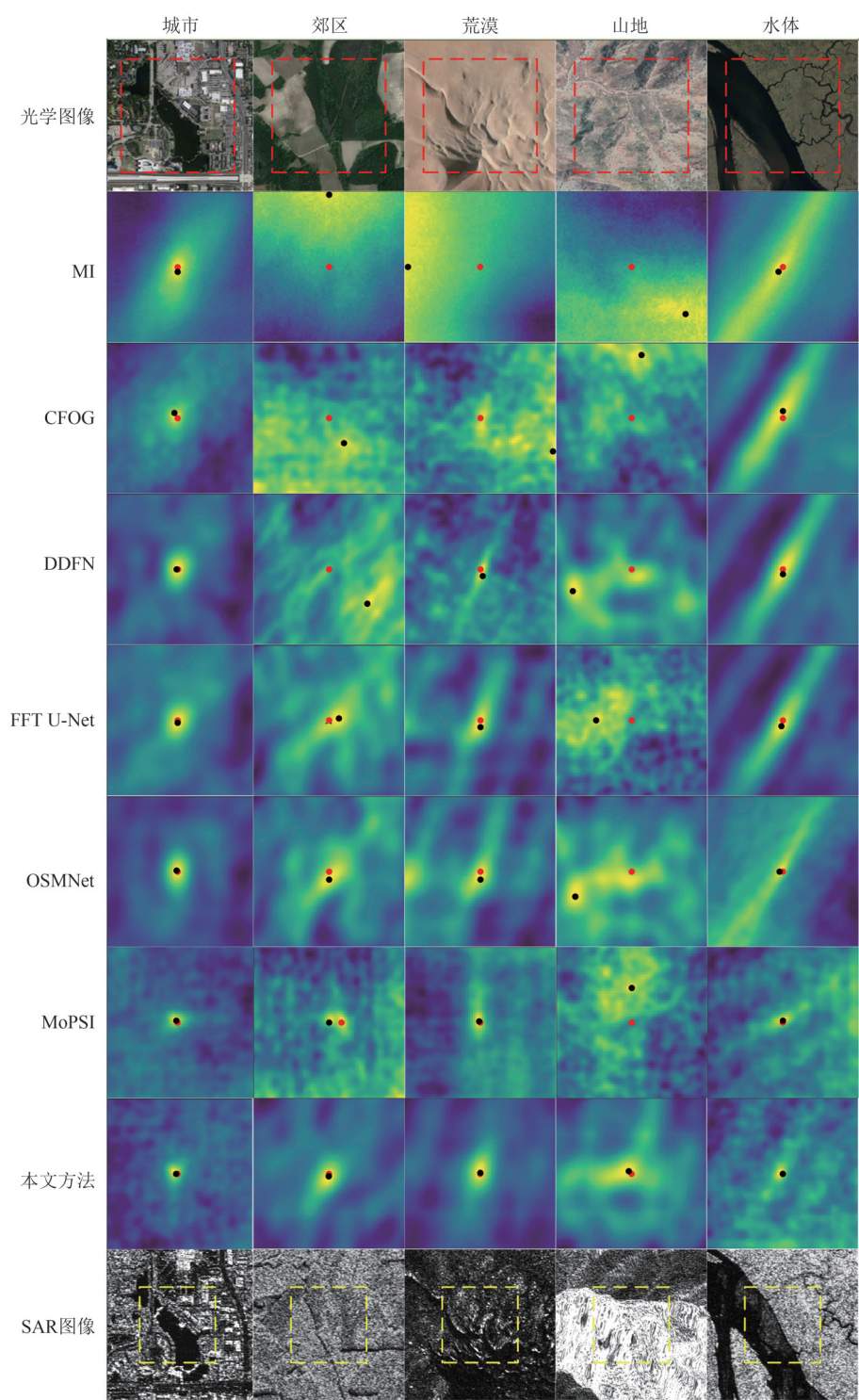


图5 对比不同匹配方法在多种地物场景上的热力图(红色点为标签位置,黑色点为预测位置)
Fig. 5 Comparing different matching methods on heatmaps in various land feature scenes (Red dots represent label positions, black dots represent predicted positions)

从图5中可以看出,在特征丰富的城区和地形平坦的水域,匹配算法均能获取正确的匹配结果。然而,针对SAR影像在山地等大地形起伏区域存在的局部几何畸变问题,传统的手工方法如CFOG,在基于方向梯度的特征表达面对影像叠影

导致地物结构方向不一致的情况下几乎失效。而现有的深度学习方法如DDFN则忽视了多尺度深度特征,而小尺度下局部特征畸变带来的影响会显著降低。本文方法不仅考虑了多尺度深度特征,还基于深度特征重建增强了特征表达,对图像间

非线性辐射差异和局部畸变具有更强的抵抗性。本文方法在5组图像上均得到了正确的匹配结果。与其他方法相比,所提方法生成的热力图峰值范围更小、更尖锐,这表明本文算法能准确描述图像特征,清晰区分正负样本,在复杂场景的影像数据集上仍能保持鲁棒的匹配性能。

3.3 消融实验

在本节中,验证了所提出方法中模块的有效

性。如图6所示,本文在两组数据上进行了4组实验,旨在验证本文引入的多尺度特征提取和特征重构模块对正确匹配率的影响。实验设置包括:Baseline,即采用伪孪生U-Net网络;方案1,在Baseline基础上增加多尺度特征提取模块;方案2,在Baseline基础上增加特征重构模块;本文方法,即在Baseline基础上增加了所提出的两个模块。

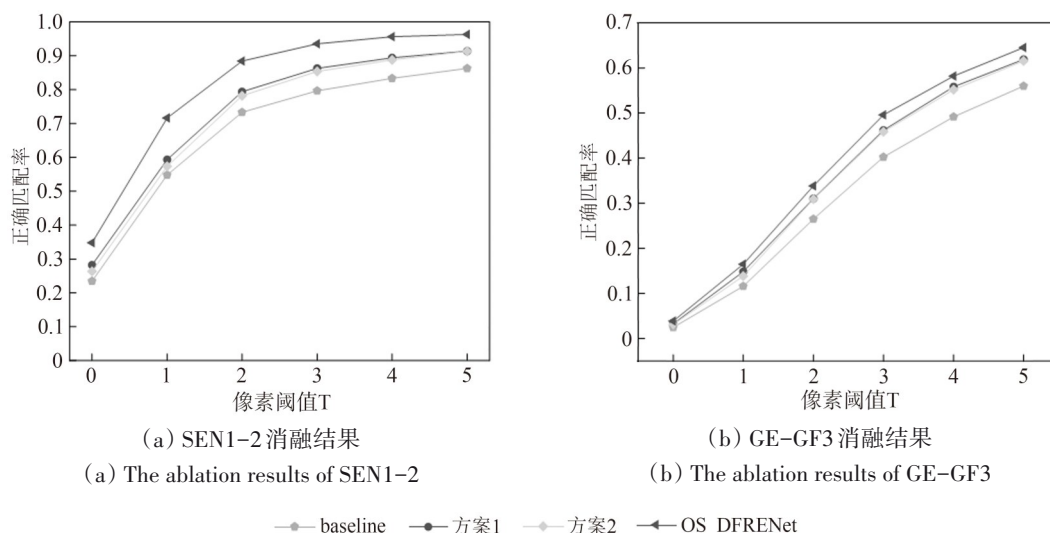


图6 所提出方法框架的消融实验

Fig. 6 The ablation experiment

从图6中对比结果可以看出,消融实验验证了本文所引入模块的有效性。多尺度特征提取模块通过网络输出图像的多尺度特征,然后同时对不同尺度的结果进行监督学习。与单一尺度的基准框架相比,该模块可有效提升网络的特征提取能力。特征重构分支模块则是在特征提取的结果上通过对图像的逆表达生成翻译图像,从而引导网络学习异源图像的共性特征。实验结果也验证了该分支结构可有效增强特征表达的稳健性,显著提高模型的正确匹配率。本文方法整合了以上两个模块,在相同实验条件下获得了最高的正确匹配率。

此外,本文还给出了输出特征维度参数 C 的消融实验,如图7所示。从结果中可以看出,所构建网络输出的特征维度的最优值为8。维度设置较小时,图像特征信息被压缩,导致图像共有特征描述不充分,正确匹配率显著下降;维度较大时则存在较多的冗余信息,一定程度上影响了特征的显著性。

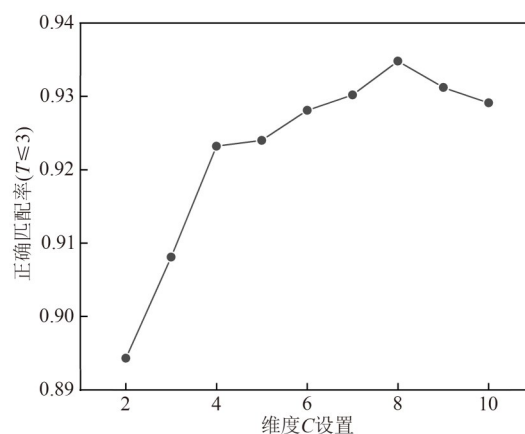


图7 特征维度参数的消融实验

Fig. 7 The ablation experiment of feature dimension parameter

而不同损失函数之间的超参数设置则是为了控制网络不同分支结构的权重,使模型在多尺度特征表达和特征重构之间达到平衡,以实现更好的训练效果。通过实验经验和参数调整, λ_{0-3} 最优的参数设置分别为0.6、0.3、0.1和0.5。

4 结 论

本文针对光学与SAR图像的自动精确匹配问题，基于伪孪生特征提取网络框架和U-Net骨干模块，提出了一种融合多尺度特征提取与特征重构增强的新匹配方法。主要结论如下：所提出的匹配方法能有效抵抗光学和SAR影像间显著的非线性辐射差异，适用于包含复杂场景的影像数据。实验结果表明，所提出的优化策略有效地提升了模型性能，且本文方法的匹配精度均优于其他几种具有代表性的匹配方法。

然而，本文提出的匹配算法存在一定的局限性，无法有效处理具有显著几何畸变（如旋转和尺度变化）的异源图像。在匹配之前，通常需要利用传感器姿态参数进行初始校正，以消除图像间的旋转和尺度差异。因此，未来研究将进一步探索具有几何不变性的光学和SAR图像智能匹配方法和技术。

参考文献(References)

- Cheng F F, Fu Z T, Huang L, Niu B S, Chen P D, Wang L G and Ji X R. 2022. Review of deep learning in optical and SAR image fusion. *National Remote Sensing Bulletin*, 26(9): 1744-1756 (成飞, 付志涛, 黄亮, 牛宝胜, 陈朋弟, 王雷光, 季欣然. 2022. 深度学习在光学和SAR影像融合研究进展. *遥感学报*, 26(9): 1744-1756) [DOI: 10.11834/jrs.20211136]
- Fan Z L, Zhang L, Liu Y X, Wang Q D and Zlatanova S. 2021. Exploiting high geopositioning accuracy of SAR data to obtain accurate geometric orientation of optical satellite images. *Remote Sensing*, 13(17): 3535 [DOI: 10.3390/rs13173535]
- Fang Y Y, Hu J, Du C, Liu Z B and Zhang L. 2022. SAR-optical image matching by integrating siamese U-Net with FFT correlation. *IEEE Geoscience and Remote Sensing Letters*, 19: 4016505 [DOI: 10.1109/LGRS.2021.3100531]
- Gao C Z, Li W, Tao R and Du Q. 2022. MS-HLMO: multiscale histogram of local main orientation for remote sensing image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5626714 [DOI: 10.1109/TGRS.2022.3193109]
- Gazzea M, Sommervold O and Arghandeh R. 2023. MARU-Net: multi-scale attention gated residual U-Net with contrastive loss for SAR-optical image matching. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 4891-4899 [DOI: 10.1109/JSTARS.2023.3277550]
- He M M, Guo Q, Li A, Chen J, Chen B and Feng X X. 2018. Automatic fast feature-level image registration for high-resolution remote sensing images. *Journal of Remote Sensing (in Chinese)*, 22(2): 277-292 (何梦梦, 郭擎, 李安, 陈俊, 陈勃, 冯旭祥. 2018. 特征级高分辨率遥感图像快速自动配准. *遥感学报*, 22(2): 277-292) [DOI: 10.11834/jrs.20186420]
- Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE: 5967-5976 [DOI: 10.1109/CVPR.2017.632]
- Li J J, Zhang J C, Yang C, Liu H Y, Zhao Y G and Ye Y X. 2023. Comparative analysis of pixel-level fusion algorithms and a new high-resolution dataset for SAR and optical image fusion. *Remote Sensing*, 15(23): 5514 [DOI: 10.3390/rs15235514]
- Li J Y, Hu Q W and Ai M Y. 2020. 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, Xu W Y, Shi P C, Zhang Y J and Hu Q W. 2022. LNIFT: locally normalized image for rotation invariant multimodal feature matching. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5621314 [DOI: 10.1109/TGRS.2022.3165940]
- 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 (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) [DOI: 10.11834/jrs.20235012]
- Merkle N, Luo W J, Auer S, Müller R and Urtasun R. 2017. Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images. *Remote Sensing*, 9(6): 586 [DOI: 10.3390/rs9060586]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- 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]
- 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]
- Wang N, Li B, Wei X X, Wang Y H and Yan H Q. 2021. Ship detection in spaceborne infrared image based on lightweight CNN and multisource feature cascade decision. *IEEE Transactions on Geoscience and Remote Sensing*, 59(5): 4324-4339 [DOI: 10.1109/TGRS.2020.3008993]
- Xiang D L, Xie Y Z, Cheng J D, Xu Y H, Zhang H and Zheng Y P. 2022. Optical and SAR image registration based on feature decoupling network. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5235913 [DOI: 10.1109/TGRS.2022.3211858]
- Xiang Y M, Wang X Q, Wang F, You H J, Qiu X L and Fu K. 2023. A global-to-local algorithm for high-resolution optical and SAR image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5215320 [DOI: 10.1109/TGRS.2023.3309855]
- Xu W Y, Yuan X H, Hu Q W and Li J Y. 2023. SAR-optical feature matching: a large-scale patch dataset and a deep local descriptor. *International Journal of Applied Earth Observation and Geoinformation*, 122: 103433 [DOI: 10.1016/j.jag.2023.103433]
- Ye Y X, Bruzzone L, Shan J, Bovolo F and Zhu Q. 2019. Fast and robust matching for multimodal remote sensing image registration.

- IEEE Transactions on Geoscience and Remote Sensing, 57(11): 9059-9070 [DOI: 10.1109/TGRS.2019.2924684]
- Ye Y X and Shen L. 2016. HOPC: a novel similarity metric based on geometric structural properties for multi-modal remote sensing image matching. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, III-1: 9-16 [DOI: 10.5194/isprs-annals-III-1-9-2016]
- Ye Y X, Sun M M, Zhou L, Yang C, Liu T Y and Hao S Y. 2023. Main body, edge decomposition and reorganization network for building change detection. Acta Geodaetica et Cartographica Sinica, 52(1): 71-81 (叶沅鑫, 孙苗苗, 周亮, 杨超, 刘天逸, 郝思媛). 2023. 面向建筑物变化检测的主体边缘分解与重组神经网络. 测绘学报, 52(1): 71-81 [DOI: 10.11947/j.AGCS.2023.20210350]
- Ye Y X, Yang C, Gong G Q, Yang P Z, Quan D and Li J Y. 2024. Robust optical and SAR image matching using attention-enhanced structural features. IEEE Transactions on Geoscience and Remote Sensing, 62: 5610212 [DOI: 10.1109/TGRS.2024.3366247]
- Yin K Y, Huang G and Huang Z Y. 2021. SAR platform's scene matching positioning method based on generalized pseudorange. Radar Science and Technology, 19(5): 589-597 (尹奎英, 黄冠, 黄照悠). 2021. 基于广义伪距定位的SAR景象匹配导航方法. 雷达科学与技术, 19(5): 589-597 [DOI: 10.3969/j.issn.1672-2337.2021.05.016]
- Zhang H, Lei L, Ni W P, Tang T, Wu J Z, Xiang D L and Kuang G Y. 2022a. Optical and SAR image matching using pixelwise deep dense features. IEEE Geoscience and Remote Sensing Letters, 19: 6000705 [DOI: 10.1109/LGRS.2020.3039473]
- Zhang H, Lei L, Ni W P, Tang T, Wu J Z, Xiang D L and Kuang G Y. 2022b. Explore better network framework for high-resolution optical and SAR image matching. IEEE Transactions on Geoscience and Remote Sensing, 60: 4704418 [DOI: 10.1109/TGRS.2021.3126939]
- Zhou L, Ye Y X, Tang T F, Nan K and Qin Y. 2022. Robust matching for SAR and optical images using multiscale convolutional gradient features. IEEE Geoscience and Remote Sensing Letters, 19: 4017605 [DOI: 10.1109/LGRS.2021.3105567]
- Zhu B, Yang C, Dai J K, Fan J W, Qin Y and Ye Y X. 2023. R₂FD₂: fast and robust matching of multimodal remote sensing images via repeatable feature detector and rotation-invariant feature descriptor. IEEE Transactions on Geoscience and Remote Sensing, 61: 5606115 [DOI: 10.1109/TGRS.2023.3264610]

Robust matching of optical and SAR images based on deep feature reconstruction enhancement

YANG Chao, LIU Chang, TANG Tengfeng, YE Yuanxin

Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China

Abstract: The automatic precise registration between optical and Synthetic Aperture Radar (SAR) imagery remains a major challenge in remote sensing due to fundamental differences in their imaging mechanisms. The inherent modality gap manifests as substantial radiometric discrepancies (speckle noise vs. photometric consistency) and geometric distortions (side-looking geometry vs. nadir projection), which introduce critical obstacles for conventional feature matching approaches. While existing deep learning-based methods have progressed in extracting deep features, most architectures inadequately address two crucial aspects: multiscale feature fusion across different imaging characteristics and cross-modal invariant feature representation. This insufficiency leads to compromised robustness in complex geographical scenarios. To address this concern, we propose a robust matching method based on deep feature reconstruction for optical and SAR images. Our method features a pseudo-Siamese network that integrates multiscale deep features and image reconstruction. First, a multiscale feature extraction architecture efficiently obtains multiscale deep features at the pixel level. This architecture allows the network to capture detailed information from various scales, which is crucial for understanding the complex patterns present in remote sensing images. Second, a pseudo-SAR translation branch for optical images is designed to reconstruct images from deep features, which enhances the ability of the network to learn robust common features. This branch mimics the characteristics of SAR images. This feature enables the network to find shared features between the two image types more effectively. Through this translation process, the network learns to focus on the essential elements that are common to optical and SAR images, which improves the matching accuracy. Finally, a joint loss function based on multilayer feature matching similarity and reconstructed image average error is constructed for robust matching. This loss function ensures that the network not only matches features accurately across different layers but also maintains a high degree of fidelity in reconstructing the original images. By combining the two aspects, the network can achieve a balance between feature similarity and image reconstruction quality, which leads to more reliable matching results. Experiments on two remote sensing image datasets with different resolutions and diverse terrain scenes (urban, suburban, desert, mountain, and water) show that our method outperforms several state-of-the-art matching methods in correct matching rate. The proposed method demonstrates superior performance in various environments, which indicates its versatility and effectiveness in real-world applications. The ability to handle different resolutions and terrain types is particularly important for practical remote sensing tasks, where conditions can vary widely.

This research advances cross-modal image analysis in two aspects: (1) a new paradigm combining feature learning with cross-domain translation is provided, and (2) practical solutions for SAR-optical registration in challenging environments are proposed.

Key words: remote sensing, optical image, SAR image, image matching, deep learning

Supported by National Key Research and Development Program of China (No.2024YFC3015402); National Natural Science Foundation of China (No. 42271446); Central Guidance for Local Science and Technology Development Fund Projects of Sichuan Province (No. 2024ZYD0084)