第 42 卷 第 24 期/2022 年 12 月/光学学报 研究论文

光学学报

基于卷积与图神经网络的合成孔径雷达与可见光图像配准

刘磊，李元祥*，倪润生，张宇轩，王艺霖，左宗成

上海交通大学航空航天学院，上海 200240

摘要 由于卫星遥感合成孔径雷达(SAR)与可见光图像之间存在显著的非线性辐射差异，故现有的SAR与可见光图像配准算法存在实时性差、旋转与尺度不变性弱等问题。针对目前算法只关注图像局部特征的外观信息，而忽略几何结构信息的问题，提出了一种结合卷积与图神经网络(GNN)的SAR与可见光图像匹配方法。该方法采用卷积神经网络进行特征检测与描述的同时，引入了GNN进行特征匹配。与最近邻匹配算法仅利用局部描述符信息相比，GNN先将特征点的位置坐标嵌入到描述符中，使得描述符具有几何位置信息，再利用注意力机制进一步聚合特征描述符的几何上下文信息，最后利用可微分的最优传输算法直接输出特征点的匹配结果，保证了网络可进行端到端的训练。实验结果表明：所提方法在大范围旋转与尺度变化的配准任务上，获得了最先进的性能；与目前主流配准算法辐射不变特征变换相比，所提方法在提升匹配精度的同时，计算速度也提高了50倍以上。

关键词 图像处理；异源图像匹配；合成孔径雷达与可见光图像配准；卷积神经网络；图神经网络；最优传输

中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/AOS202242.2410002

Synthetic Aperture Radar and Optical Images Registration Based on Convolutional and Graph Neural Networks

Liu Lei, Li Yuanxiang*, Ni Runsheng, Zhang Yuxuan, Wang Yilin, Zuo Zongcheng

School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240, *China*

Abstract Due to the significant nonlinear radiometric differences between synthetic aperture radar (SAR) and optical images obtained by satellite remote sensing, the current SAR and optical images registration algorithms are weakened by their poor real-time performance and weak rotation and scale invariance. To address the problem that the current algorithms only focus on the appearance information on the local features of images and ignore the geometric structure information, a SAR and optical image matching method combining the convolutional and graph neural network (GNN) is proposed. The method uses the convolutional neural network for feature detection and description, and adopts the GNN for feature matching. In contrast to the nearest neighbor matching algorithm that merely uses local descriptor information, the GNN embeds the location coordinates of feature points into the descriptors, thereby providing the descriptors with geometric location information. Then, the geometric context information of the feature descriptors is further aggregated with the attention mechanism. Finally, the matching results of the feature points are directly output by the differentiable optimal transport algorithm to ensure that the network can be trained in an end-to-end manner. The experimental results show that the proposed method achieves state-of-the-art performance on the registration task featuring rotation and scale variation in a large range. In addition, compared with the current mainstream registration algorithm radiation-invariant feature transform, the proposed method not only improves matching accuracy, but also increases the computational speed by more than 50 times.

Key words image processing; multimodal image matching; synthetic aperture radar and optical images registration; convolutional neural network; graph neural network; optimal transport

收稿日期：2022-04-18；**修回日期**：2022-06-17；**录用日期**：2022-07-11

通信作者：*yuanxli@sjtu.edu.cn

1 引 言

合成孔径雷达（SAR）图像和可见光图像是卫星遥感领域中常见的两类异源图像，两者融合将提供更丰富的信息[1]。不同卫星平台获取的SAR图像和可见光图像间通常存在视角差异，故进行图像融合之前需进行图像配准。SAR图像因存在大量斑点噪声，与可见光图像之间存在显著的非线性辐射畸变（NRD）[2]，使得基于同源图像的经典配准算法（如SIFT[3]、ORB[4]等）失效或性能大幅度下降。近年来，遥感异源图像配准问题得到了较为广泛的关注[5]，根据是否采用深度学习方法，可分为基于人工设计的配准方法和基于深度学习的配准方法。

在基于人工设计的配准方法中，较早的遥感SAR与可见光图像配准方法包括两类：基于区域的互信息模板匹配方法[5]和基于特征的改进SIFT方法[6]，但这两类方法配准效果较差且计算代价高昂。近期研究发现，基于频域的相位一致性（PC）特征[7]对NRD具有更好的鲁棒性，在遥感领域得到了较为广泛的应用：HOPC方法[8]通过拓展PC特征构建区域描述符，采用模板匹配方法进行配准，但只对异源图像间的小范围平移问题有效；辐射不变特征变换（RIFT）[2]直接利用PC特征图进行特征点检测与描述，实现了配准的旋转不变性；后续研究者进一步提高了PC特征的旋转与尺度不变性[9-11]，但这类配准算法普遍较慢、不具有实时性。

为了提高配准精度与实时性，近年来出现了一些基于深度学习的SAR与可见光图像配准研究工作，大致可以分为三类：1）基于风格迁移的配准方法[12-13]，即采用生成对抗网络（GAN）等模型，将SAR与可见光图像迁移到同一种模态下，降低异源图像之间的差异，再利用同源配准方法进行匹配，这类网络目前泛化能力不强，且容易丢失细节信息；2）基于图像区域的配准方法，Hughes等[14]和Zhang等[15]将SAR与可见光图像块输入到孪生卷积神经网络中，利用特征图的相关性预测图像块之间的平移量，但这类算法在SAR与可见光影像之间存在显著的旋转与尺度差异，难以保证配准的鲁棒性；3）基于深度学习特征的配准方法，这类方法主要利用卷积神经网络同时进行特征点的检测与描述，典型的网络包括SuperPoint[16]与D2Net[17]，目前主要用于自然图像配准工作中。其中，D2Net方法实现了昼夜图像的配准，CMM-Net[18]是D2Net的改进方法，用于SAR与可见光图像的配准任务。

在上述研究工作中，通常只考虑图像区域或局部特征的外观相似性，忽略了特征点在空间位置上的关系，难以解决SAR与可见光图像在显著旋转与尺度变化下的配准问题。本文受基于上下文描述符的ContextDesc算法[19]与SuperGlue算法[20]启发，提出了一种基于卷积与图神经网络（CGNet）的配准方法。CGNet先利用CNN进行局部特征点的检测与描述，再结合特征点的位置信息，通过图神经网络（GNN）聚合描述符的几何上下文信息，最后利用图匹配方法完成特征点匹配，构建了端到端的SAR与可见光配准网络。

2 SAR与可见光图像配准网络架构

本文的SAR与可见光图像配准网络架构如图1所示，其中CNN用于特征点检测与描述任务，GNN用于特征点匹配任务。

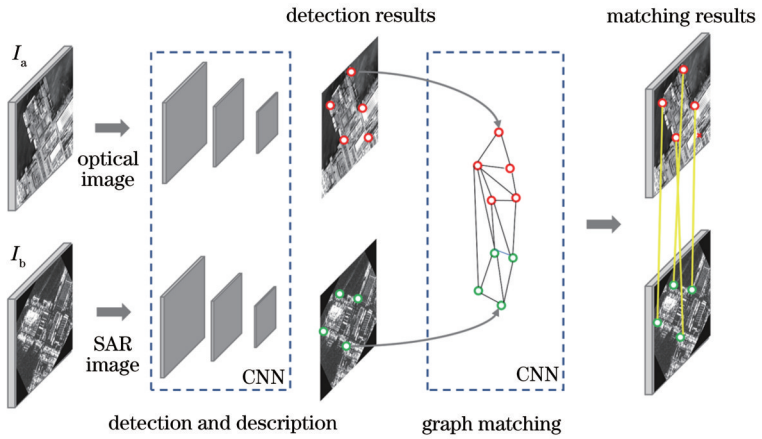

图1 SAR与可见光图像配准网络架构

Fig. 1 Architecture of SAR and optical images registration network

2.1 特征点检测与描述网络

如图2所示，图像特征点的检测与描述直接采用SuperPoint[16]网络结构，该网络由一个用于特征提取的编码器和两个分别用于检测与描述的解码器组成，两个解码器通过共享编码器减少了模型整体的参数量。

1）特征编码器

该编码器利用类似VGG[21]的结构进行特征提取与降维，由卷积层、池化空间降采样和非线性激活函数三部分构成。编码器共使用了三个最大池化层，即在

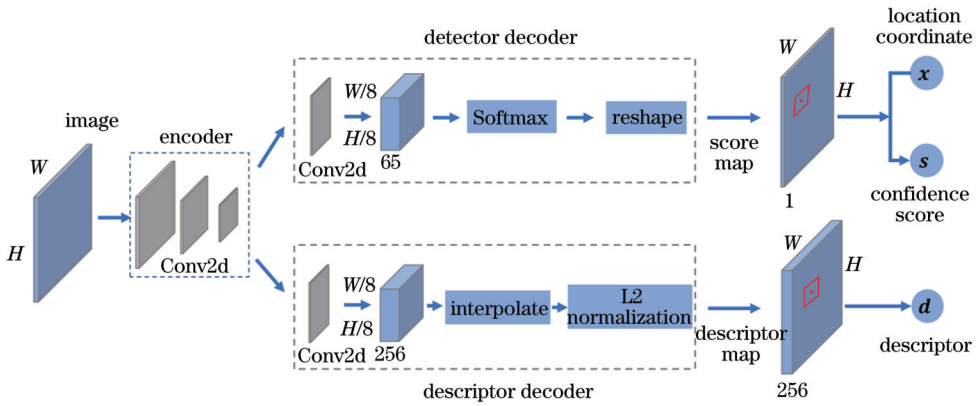

图 2　特征检测与描述网络

Fig. 2　Feature detection and description network

空间维度上经过了三次两倍下采样，假设输入图像大小为 $H\times W$，则经过编码器后的张量图维度为 $H_c\times W_c\times F$，其中 $H_c=H/8$，$W_c=W/8$，F 为通道数，这说明张量图上的每一个像素点都代表原图上大小为 8×8 的局部图像块。

2）特征检测解码器

编码器输出的张量图通过该解码器的两层卷积模块后，中间张量的维度大小变为 $H_c\times W_c\times 65$，其中 65 个通道中的前 64 个用于预测原图上 8×8 的局部区域中每个像素作为特征点的概率，最后一个通道代表这个区域内没有特征点的概率。对中间张量进行 Softmax 操作，去掉最后一个通道，利用 reshape 操作将张量的维度大小从 $H_c\times W_c\times 64$ 变成 $H\times W$，获得一张原图大小的特征点置信度得分图。根据置信度得分 $\boldsymbol{s}$，设定得分大于阈值 s_{th} 的像素点为特征点，并将对应的像素位置设定为特征点的位置坐标 $\boldsymbol{x}$。

3）特征描述解码器

该解码器包含两个基本卷积层，输出的张量维度为 $H_c\times W_c\times 256$，直接利用双线性插值操作，将维度变为 $H\times W\times 256$，最后对每个像素点的通道进行 L2 归一化，获得单位化的描述符，根据特征点的位置进行采样，可获得特征点描述符 $\boldsymbol{d}$。

假设输入图像分别为 I_a 和 I_b，经过检测与描述网络后，输出特征点集合 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，则每个特征点都包含位置坐标、置信度、描述符三个属性，即 $(\boldsymbol{x}_i,\boldsymbol{s}_i,\boldsymbol{d}_i)\in\boldsymbol{A}$，$(\boldsymbol{x}_j,\boldsymbol{s}_j,\boldsymbol{d}_j)\in\boldsymbol{B}$。

2.2　特征点匹配网络

特征点匹配网络由位置信息编码器、GNN、最优传输三部分组成，结构如图 3 所示。

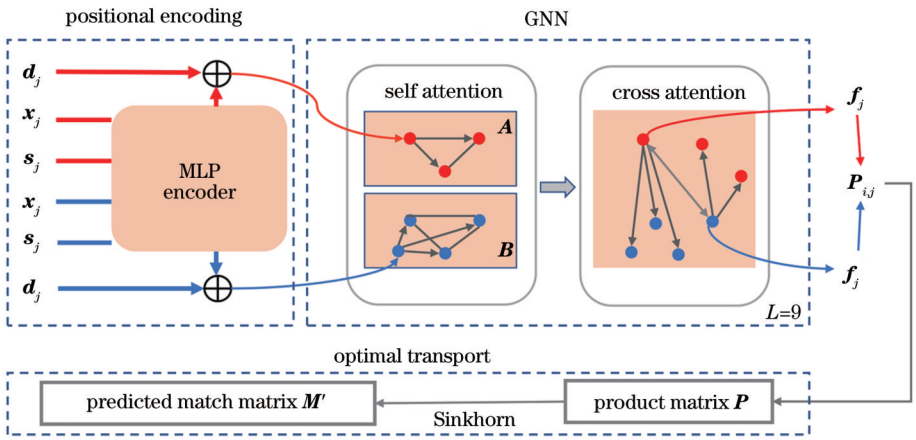

图 3　特征点匹配网络

Fig. 3　Feature point matching network

1）位置信息编码器

为使得局部描述符具有几何上下文信息，本文将特征点 i 的位置 $\boldsymbol{x}_i$ 与置信度 $\boldsymbol{s}_i$ 输入到多层感知机（MLP）网络，目的是将信息嵌入到高维的描述符向量 $\boldsymbol{d}_i$ 中，得到新的特征描述符 $\boldsymbol{f}_i$，相应表达式为

$$\boldsymbol{f}_i^{(0)}=\boldsymbol{d}_i+\mathrm{MLP}(\boldsymbol{x}_i,\boldsymbol{s}_i),\tag{1}$$

式中：MLP(·)表示利用多层感知机进行信息编码。

2）GNN

为了使得特征描述符能进一步聚合几何上下文信息，需要采用自注意力层和交叉注意力层进行信息传

递，相应表达式为

$$\boldsymbol{f}_i^{(l+1)}=\boldsymbol{f}_i^{(l)}+\mathrm{MLP}\left(\boldsymbol{f}_i^{(l)}\|\boldsymbol{m}_{\varepsilon\to i}\right),\tag{2}$$

式中：l为网络层数；ε为自注意力机制模式参数；$[\cdot\|\cdot]$为串联操作。

网络总共L层，l为奇数时代表该网络层为自注意力层（$\varepsilon=\varepsilon_{\mathrm{self}}$），表示$\boldsymbol{A}$中的特征点$i$与$\boldsymbol{A}$中其他所有特征点进行信息的注意力聚合，而$l$为偶数时代表该网络层为交叉注意力层（$\varepsilon=\varepsilon_{\mathrm{cross}}$），表示$\boldsymbol{A}$中的特征点$i$与$\boldsymbol{B}$中所有的特征点进行信息的注意力聚合。其中，$\boldsymbol{m}_{\varepsilon\to i}$表示特征点$i$与其他特征点聚合信息的线性加权和，即

$$\boldsymbol{m}_{\varepsilon\to i}=\sum\boldsymbol{\alpha}_{i,j}\boldsymbol{v}_j,\tag{3}$$

其中网络每一层的注意力权重$\boldsymbol{\alpha}_{i,j}$与信息$\boldsymbol{v}_j$的获取方式类似数据库检索，首先将特征点$i$视为查询点，获得属性值$\boldsymbol{q}_i$，对其他所有相关的特征点$j$进行检索，特征点$j$的属性键值为$\boldsymbol{k}_j$，检索信息值为$\boldsymbol{v}_j$，这里的$\boldsymbol{q}_i$、$\boldsymbol{k}_j$和$\boldsymbol{v}_j$都是通过特征的线性映射得到的，即

$$\boldsymbol{q}_i=\boldsymbol{W}_1^{(l)}\boldsymbol{f}_i^{(l)}+\boldsymbol{b}_i^{(l)},\tag{4}$$

$$\begin{bmatrix}\boldsymbol{k}_j\\ \boldsymbol{v}_j\end{bmatrix}=\begin{bmatrix}\boldsymbol{W}_2\\ \boldsymbol{W}_3\end{bmatrix}^{(l)}\boldsymbol{f}_j^{(l)}+\begin{bmatrix}\boldsymbol{b}_2\\ \boldsymbol{b}_3\end{bmatrix},\tag{5}$$

式中：$\boldsymbol{W}$和$\boldsymbol{b}$是每层网络的权重参数和偏置参数。注意力权重$\boldsymbol{\alpha}_{i,j}$是通过对查询点$i$与所有检索对象$j$的属性值内积进行Softmax操作得到的，即

$$\boldsymbol{\alpha}_{i,j}=\mathrm{Softmax}\left(\left\langle\boldsymbol{q}_i,\boldsymbol{k}_j\right\rangle\right),\tag{6}$$

式中：Softmax(•)表示进行Softmax操作。

利用自注意力层与交叉注意力层重复多次进行特征描述符的几何上下文信息增强，学习潜在的几何变换与特征投影。

3）最优传输

将GNN输出的所有特征向量$\boldsymbol{f}_i$与$\boldsymbol{f}_j$分别进行内积，即$\boldsymbol{P}_{i,j}=\left\langle\boldsymbol{f}_i,\boldsymbol{f}_j\right\rangle$，获得特征的内积矩阵，将内积矩阵$\boldsymbol{P}$通过可微分的最优传输算法Sinkhorn[22-23]迭代T次，近似解算出特征点匹配矩阵$\boldsymbol{M}'$，与传统的最近邻匹配算法相比，最优传输算法运算可导，适用于反向传播的深度学习网络。

3　匹配矩阵与损失函数

CGNet直接采用SuperPoint进行特征点检测与描述，需要训练的部分是特征点匹配网络，目的是弥补现有局部特征描述符的不足，进一步增强描述符的几何上下文信息。

由于存在视角变化的SAR与可见光图像难以通过标注匹配特征点的方式获得训练真值，故本文采用预先配准对齐的SAR与可见光数据集进行训练，通过仿射变换矩阵随机生成几何变换，模拟视角变化，其中主要的空间变换类型包括：平移、旋转、缩放和复合的刚体变换，如图4所示。

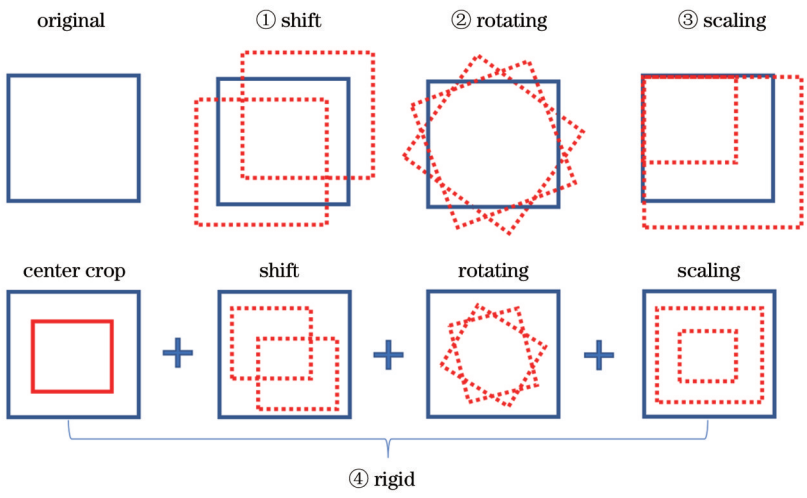

图4　随机几何变换

Fig. 4　Random geometric transformation

3.1　随机几何变换与匹配矩阵

在训练过程中，对图片I_{a}和I_{b}施加随机几何变换$\boldsymbol{H}$后，输入SuperPoint网络中，获取特征点集合$\boldsymbol{A}$、$\boldsymbol{B}$。如图5(a)所示，假定图像I_{a}与I_{b}检测特征点的个数分别为$m=4$与$n=4$，存在3对正确匹配对，且每张图都有一个未匹配点。为了表示所有特征点的匹配关系，采用维度为$(m+1)\times(n+1)$的匹配矩阵$\boldsymbol{M}$来表达点与点之间的匹配关系，匹配矩阵$\boldsymbol{M}$具备两条基本性质：1）$\boldsymbol{M}$中元素都为0或1，元素为1时代表正确匹配，元素为0时代表未匹配；2）$\boldsymbol{M}$每行和每列的和都为1，代表每个特征点在另一张图上有且仅有一个匹配点。

3.2　损失函数

如图5(a)所示，匹配矩阵$\boldsymbol{M}$中值为1的元素分为三个部分，其中$\boldsymbol{C}$区域代表同名点对的集合，$\boldsymbol{\Phi}_{m+1}$、

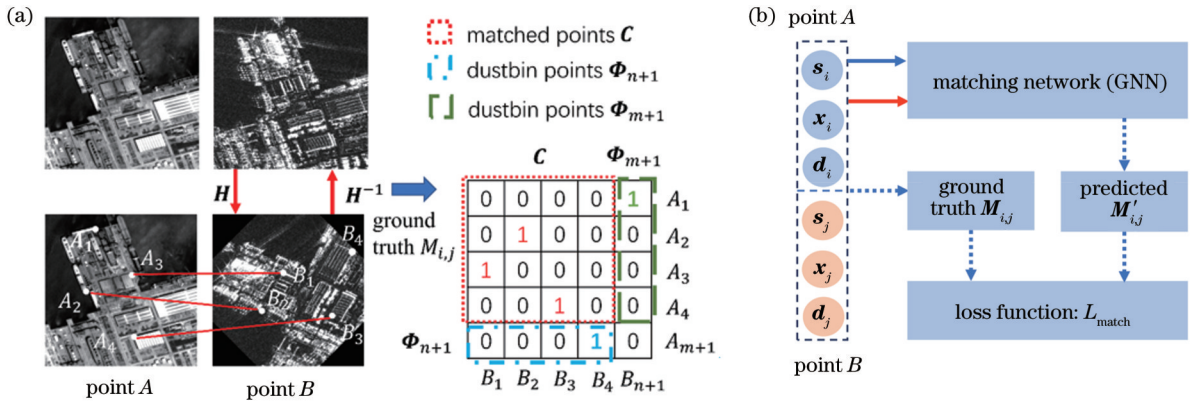

图 5　匹配矩阵。(a)真值生成;(b)预测与损失计算

Fig. 5　Matching matrix. (a) Ground truth generation; (b) prediction and loss calculation

$\boldsymbol{\Phi}_{n+1}$区域代表垃圾箱列,表示非同名特征点集合。从图5(b)可知,损失是通过匹配矩阵的真值$\boldsymbol{M}$与预测矩阵$\boldsymbol{M}'$计算获得的,本文采用负对数似然作为损失函数,由于匹配矩阵的值均为0或1,0的部分交叉熵也为0,故只需要计算值为1的部分。损失函数的公式为

$$L_{\text{match}}(I_{\text{a}},I_{\text{b}})=-\sum_{(i,j)\in C}\ln \boldsymbol{M}'_{i,j}-\left(\sum_{i\in\boldsymbol{\Phi}_{n+1}}\ln \boldsymbol{M}'_{i,n+1}+\sum_{j\in\boldsymbol{\Phi}_{m+1}}\ln \boldsymbol{M}'_{m+1,j}\right)。\tag{7}$$

由于SAR与可见光图像差异明显,特征点检测重复率低,同名点数量稀少,故本文需要特别将$\boldsymbol{C}$中值为1的元素视为正样本,表示正确的同名点对,而将$\boldsymbol{\Phi}_{m+1}$、$\boldsymbol{\Phi}_{n+1}$中值为1的元素视为负样本,表示非同名特征点对。因为在实验中同名点对数量远少于非同名点对,所以需要引入平衡因子λ,防止网络只进行非同名点对的匹配学习,而未能正确学习同名点对的匹配。同时,由于特征点匹配网络在刚开始训练时,局部描述符的差异过大,会导致预测匹配矩阵$\boldsymbol{M}'$中的值(描述符内积)过小,故为防止损失函数出现无穷大的情况,需要引入小量常数τ,最后将损失函数修正为

$$L_{\text{match, correction}}(I_{\text{a}},I_{\text{b}})=-\lambda\sum_{(i,j)\in C}\ln(\boldsymbol{M}'_{i,j}+\tau)-\left[\sum_{i\in\boldsymbol{\Phi}_{n+1}}\ln(\boldsymbol{M}'_{i,n+1}+\tau)+\sum_{j\in\boldsymbol{\Phi}_{m+1}}\ln(\boldsymbol{M}'_{m+1,j}+\tau)\right]。\tag{8}$$

4　实验结果与分析

4.1　实验数据与参数设置

实验中的SAR与可见光图像数据集来源于高分三号与资源三号卫星,覆盖包括港口、城市、河流、机场、岛屿和平原共6个场景。图像需要预先进行配准对齐,经过切片后,分辨率为512 pixel×512 pixel,选择其中1497张图片作为训练样本,350张图片用于测试。本文的CGNET方法基于PyTorch框架实现。特征点检测与描述网络SuperPoint的置信度阈值s_{th}设为0.0001,该网络在训练时需要冻结参数。在特征点匹配网络中,用于位置信息编码的MLP网络共4层,通道维度为32、64、128和256。GNN共9层,即$L=9$。最优传输算法Sinkhorn的迭代次数为$T=50$。将损失函数中用于平衡正负样本比例的参数λ设为20,将防止损失无穷大的小量常数τ设为10^{-6}。训练采用Adam优化器,初始学习率设定为10^{-5},批尺寸设为1,训练300个epoch,每75个epoch后学习率衰减为原来的1/2。

4.2　特征点检测性能评估

为验证特征点检测器对NRD的鲁棒性,本文在对比实验中采用的方法包括:经典的点特征匹配算法(SIFT[3]、ORB[4])、基于相位一致性特征的异源图像匹配算法(RIFT[2])和基于CNN的深度学习匹配算法(SuperPoint[16]和D2Net[17])。图6显示了一组SAR与可见光图像的特征点检测结果,每组方法均采用置信度前1000的特征点进行显示。可以看出:ORB检测的特征点分布较为集中;SIFT和D2Net方法易受噪声干扰,大量特征点分布在亮斑区域;RIFT、SuperPoint的特征点主要分布在边缘、角点上,在SAR与可见光图像上的分布均匀且近似,检测效果较为理想。本文采用特征点重复率对检测器的效果进行定量分析,其计算公式为

$$R_{\text{rep}}=\frac{N^{\text{c}}}{(m+n)/2}=\frac{\left|\left\{\left\|\boldsymbol{x}_i^{\text{a}}-\boldsymbol{H}\boldsymbol{x}_j^{\text{b}}\right\|_2<3\right\}_{i=1}^{m}\right|}{(m+n)/2},\tag{9}$$

式中:$\boldsymbol{x}_i^{\text{a}}$和$\boldsymbol{x}_j^{\text{b}}$表示特征点的位置坐标;$\left\|\mathbf{x}_i^{\text{a}}-\boldsymbol{H}\boldsymbol{x}_j^{\text{b}}\right\|_2<3$表示将待匹配特征点之间的像素误差小于3时视为正确匹配对;$\left|\left\{\left\|\boldsymbol{x}_i^{\text{a}}-\boldsymbol{H}\boldsymbol{x}_j^{\text{b}}\right\|_2<3\right\}_{i=1}^{m}\right|$表示该组图片中正确匹配对的数量$N^{\text{c}}$。

本文使用整个数据集(1847对图片)进行测试评估,计算每个检测器的N^{c}与R_{rep}平均值,评估结果如

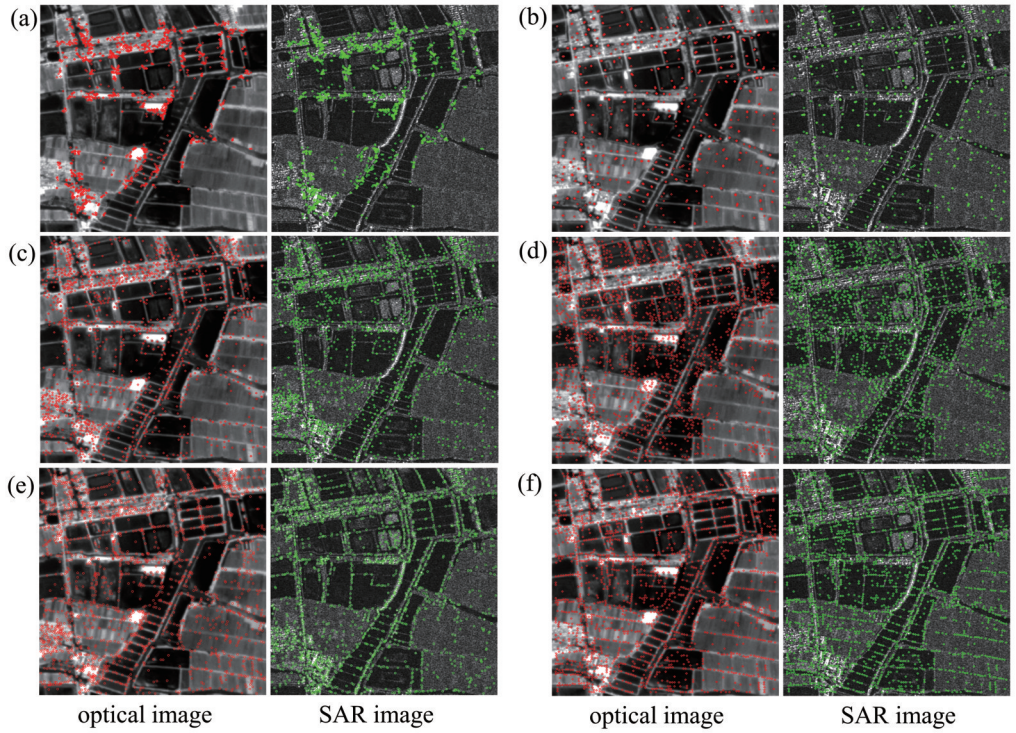

图 6 特征点检测结果。(a)原图;(b) ORB;(c) SIFT;(d) D2Net;(e) RIFT;(f) SuperPoint

Fig. 6 Results of feature point detection. (a) Original image; (b) ORB; (c) SIFT; (d) D2Net; (e) RIFT; (f) SuperPoint

表1所示。综合定性与定量分析,SuperPoint检测效果最佳,对NRD较为鲁棒。值得注意的是,以上特征检测器在同源图像数据集中的R_{rep}通常超过50%,但在本文的SAR与可见光数据集上,R_{rep}均低于25%,这将导致特征点正负样本不均衡,是造成特征点匹配网络训练困难的主要原因之一。

表1 特征检测器性能评估

Table 1 Performance evaluation of feature detector

Index	SIFT	ORB	RIFT	D2Net	SuperPoint
N^c	158.2	196.9	178.1	158.8	221.4
R_{rep}/%	17.5	21.3	18.1	16.0	22.6

4.3 特征点匹配性能评估

特征点匹配性能的评价指标包括:匹配成功率R_S、正确匹配点数量N_{CM}和匹配点的均方根误差R_{MSE}。在特征点匹配阶段,将正确匹配点对的像素误差阈值设为3,当每对图片中的正确匹配对数量大于4时,认为该组图像匹配成功,R_S由匹配成功的图像对数量除以整个数据集中图像对的总量得到。为验证CGNet的匹配性能,本文对平移、旋转、尺度缩放和刚体变换下的待配准图像进行了定性与定量的评估,对比实验包括两种传统方法(OS-SIFT[6]和RIFT[2])和三种深度学习方法(SuperPoint[16]、D2Net[17]和CMM-Net[18])。其中,OS-SIFT、RIFT和CMM-Net是专门为SAR与可见光图像配准而设计的,而SuperPoint和D2Net主要用于自然图像配准任务,但对SAR与可见光配准任务也具有一定的泛化性能,具有参考价值。

首先验证CGNet的平移不变性,本文根据已有的350幅对齐的测试集图像,利用空间变换矩阵生成具有平移变换的SAR与可见光图像数据集,每组图像在纵横两个方向的相对平移量范围为−128~128 pixel。其中三组平移情况下的特征点匹配效果如图7所示。可以看出,在平移量较大的情况下,OS-SIFT和SuperPoint匹配性能相对较差,存在失效的情况,D2Net获取的匹配数量相对较少,而所提CGNet方法可获得最丰富的匹配特征点。

对具有平移变化的350对测试集图片进行定量的评估,计算匹配成功率R_S,并统计所有匹配成功图片的N_{CM}与R_{MSE}均值,其结果如表2所示。可以看出:OS-SIFT匹配性能最差,说明基于SIFT改进的方法受NRD影响十分明显;SuperPoint的匹配成功率较低,说明该方法的局部特征描述符在异源图像匹配任务中并不鲁棒;D2Net的R_{MSE}较高,匹配精度较差;CMM-Net采用了与D2Net相同的网络架构,故在图7中匹配数量较为丰富,但特征图为原图大小的1/4,精度较低,进而会导致实际正确匹配数量N_{CM}较少;CGNet的正确匹配数量超过RIFT,与SuperPoint相比获得了显著的性能提升,表明本文利用GNN进行特征匹配,替代经典的最近邻匹配算法,将显著提升SAR与可见光图像配准的成功率。

与平移不变性实验相同,为验证CGNet具有较好的旋转与尺度不变性,利用空间变换矩阵随机生成具有旋转、缩放和刚体变换的测试集,其中旋转变换的角度范围为−180°~180°,尺度比例的变化范围为原图宽

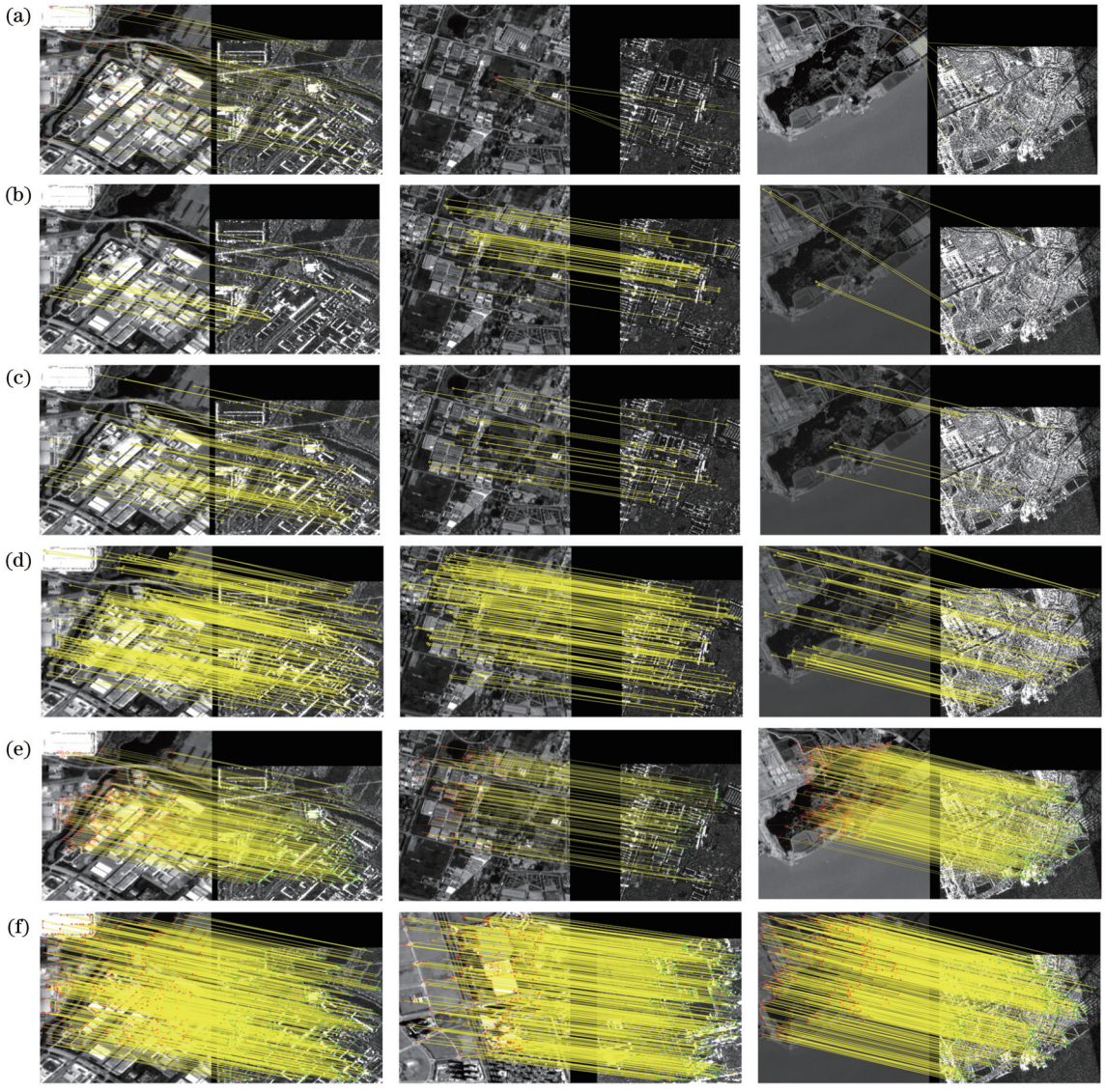

图 7 平移不变性实验匹配结果。(a) OS-SIFT; (b) SuperPoint; (c) D2Net; (d) CMM-Net; (e) RIFT; (f) 所提CGNet

Fig. 7 Matching results of shift invariance experiment. (a) OS-SIFT; (b) SuperPoint; (c) D2Net; (d) CMM-Net; (e) RIFT; (f) proposed CGNet

表2 平移不变性实验中的特征匹配性能

Table 2 Performance of feature matching in shift invariance experiment

Method	R_S	N_{CM}	R_{MSE}
OS-SIFT	0.097	8.9	1.688
SuperPoint	0.454	25.9	1.705
D2Net	0.663	15.9	1.895
CMM-Net	0.880	23.3	1.928
RIFT	0.969	86.7	1.733
CGNet	0.977	171.6	1.711

高的0.6～1.0,刚体变换由中心裁剪(原图宽高的0.75)、旋转(−180°～180°)、平移(原图宽高的0.1)和缩放(原图宽高的0.75～1.25)随机组合而成,评估结果如表3所示。实验结果表明:大范围旋转与尺度变换将造成除CGNet以外的所有对比方法的匹配成功率R_S大幅度下降,如未加入旋转不变性的RIFT,在大尺度旋转变化情况下,配准成功率降低91.5个百分点,而由于RIFT旋转不变性算法需要消耗10倍以上的计算代价,故本文不进行进一步讨论与验证;与SuperPoint相比,所提CGNet效果提升显著,说明GNN有效地利用了特征点位置与几何上下文信息,对大范围的旋转与尺度变换具有较强的适应能力。

最后,图8展示了两组图片旋转150°与210°后的匹配与棋盘格配准结果。图9展示了两组图片缩放尺度分别为0.6和0.75下的匹配结果,以及两组随机刚体变换下的匹配结果。

4.4 消融实验

为了验证CGNet网络中各部分组件的必要性,进行了消融实验。所有实验设置相同的参数,在相同的训练数据上训练300个epoch后,在平移情况下的配准测试集上进行性能评估,结果如表4所示。从前4组实

表 3 旋转、缩放和刚体变换实验中的特征匹配性能

Table 3 Performance of feature matching in rotation, scaling and rigid transformation experiments

Transformation	Index	OS-SIFT	SuperPoint	D2Net	CMM-Net	RIFT	CGNet
Rotation	R_S	0	0.060	0.063	0.097	0.054	0.951
	N_{CM}		29.1	16.5	19.7	85.7	98.7
	R_{MSE}		1.729	1.900	1.931	1.912	1.753
Scaling	R_S	0.240	0.523	0.600	0.917	0.620	0.957
	N_{CM}	31.7	26.9	15.6	22.1	69.8	173.6
	R_{MSE}	1.748	1.877	1.885	1.953	1.869	1.811
Rigid	R_S	0.040	0.137	0.166	0.251	0.220	0.940
	N_{CM}	37.5	27.4	12.2	20.4	31.9	84.2
	R_{MSE}	1.708	1.829	1.945	1.969	1.848	1.827

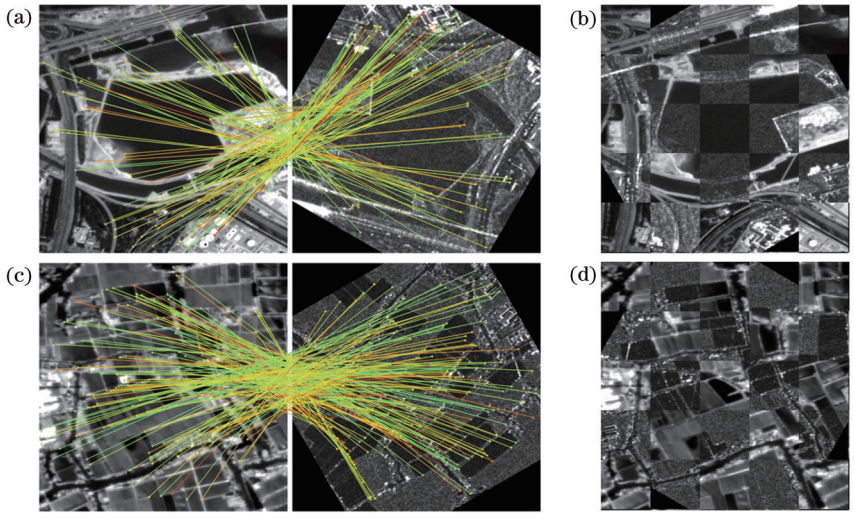

图 8 旋转不变性实验的匹配和配准结果。(a)150°旋转的匹配结果;(b)150°旋转的配准结果;(c)210°旋转的匹配结果;(d)210°旋转的配准结果

Fig. 8 Matching and registration results in rotation invariance experiment. (a) Matching result of 150° rotation; (b) registration result of 150° rotation; (c) matching result of 210° rotation; (d) registration result of 210° rotation

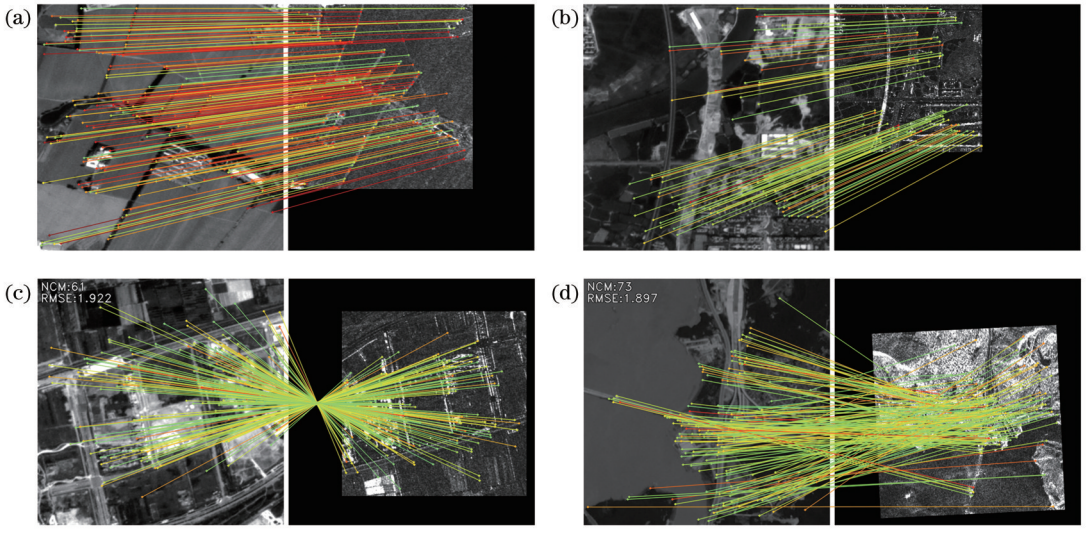

图 9 尺度与刚体变换实验的匹配结果。(a)缩放尺度为 0.75;(b)缩放尺度为 0.6;(c)随机刚体变换 1;(d)随机刚体变换 2

Fig. 9 Matching results in scaling and rigid transformation. (a) Scaling scale of 0.75; (b) scaling scale of 0.6; (c) random rigid transformation 1; (d) random rigid transformation 2

验可以看出，位置编码器、自注意力和交叉注意力层均为关键组件，剔除任何一个部分都会显著降低性能。后三组实验证明，加深注意力模块的层数可提高网络的拟合能力。

表 4　消融实验结果

Tabel 4　Results of ablation study

SuperPoint	Positional encoding	Number of layers of self-attention module	Number of layers of cross-attention module	R_S	N_{CM}	R_{MSE}
√				0.454	25.9	1.705
√	√			0		
√	√	9		0.649	54.7	1.614
√		9	9	0.386	18.3	1.524
√	√	3	3	0.274	8.8	1.440
√	√	6	6	0.800	71.1	1.621
√	√	9	9	0.977	171.6	1.711

4.5　正负样本均衡与计算效率

由于特征点 R_{rep} 低，将使得构建匹配矩阵损失函数时，出现显著的正负样本不均衡问题，故为研究平衡因子 λ 对匹配性能的影响，选择不同的 λ 值进行训练。在训练过程中，匹配成功率 R_S 随 epoch 的变化如图 10 所示。可以发现，平衡因子 λ 会影响损失函数的收敛速度，并影响最终的匹配成功率，$\lambda=1$ 时的匹配成功率非常低。在增大 λ 后，匹配成功率的增长量如表 5 所示。可以看出，当 $\lambda=20$ 时，数据训练 300 个 epoch 后，匹配成功率提高了 87.5%，λ 显著影响最终的匹配性能。

最后，进行算法运行时间与参数量比较，结果如表 6 所示。本文实验的计算机硬件资源为：中央处理器（CPU）型号为 Intel(R) Xeon(R) W-2150B，显卡型号为 GTX 2080ti。在检测阶段中，无法使用图形处理单元（GPU）加速的 RIFT 算法需要近 6 s 的时间进行特征检测与描述，而 CGNet 的检测速度是 RIFT 算法的 300 多倍。在匹配阶段中，SuperPoint、D2Net 利用最近邻匹配算法进行特征描述符匹配，并结合随机抽样一致（RANSAC）算法与误匹配点筛除，但由于 SAR 与可见光图像同名特征点的描述符差异较大，故通常需要迭代 1000 次以上才能有良好的匹配性能，匹配时间相对较长。CMM-Net 采用了快速最近邻匹配方法，匹配速度相对较快。CGNet 引入了 GNN，虽然参数量比 SuperPoint 网络多了约 4.1×10^7，但是匹配速度和精度均得到了显著提升。总之，CGNet 在提升匹配精度的同时，计算速度比 CMM-Net 提高了 10 倍，比 RIFT 提高了 50 倍以上。

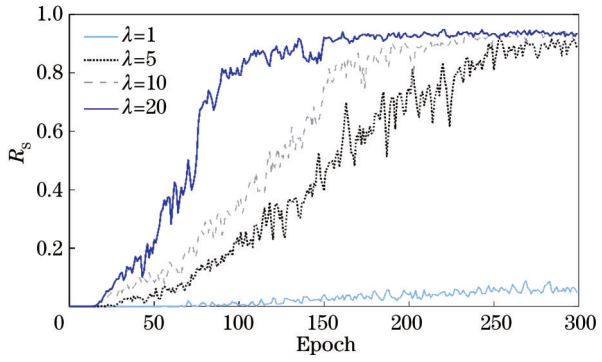

图 10　R_S 随 epoch 的变化

Fig. 10　R_S varying with epoch

表 5　不同 epoch 下的匹配成功率增长量

Table 5　Growth value of matching success rate R_S under different epochs

λ	50	100	150	200	250	300
5	0.029	0.183	0.317	0.729	0.694	0.849
10	0.071	0.297	0.732	0.854	0.853	0.871
20	0.254	0.780	0.889	0.897	0.860	0.875

表 6　运行时间与参数量比较

Table 6　Running time and number of parameters comparison

Index	OS-SIFT	SuperPoint	D2net	CMM-Net	RIFT	CGNet
Number of parameters /10^6		4.96	29.1	29.1		45.8
Detection time /s	12.133	0.023	0.231	0.269	5.885	0.017
Matching time /s	0.685	2.284	2.152	0.821	1.416	0.086

5　结　　论

提出了一种基于卷积与 GNN 的 SAR 与可见光图像配准方法 CGNet。由于采用了 CNN 的特征点检测方法，故 CGNet 在计算速度上显著优于基于相位一致性特征的遥感图像匹配方法。在 SAR 与可见光

配准任务中，通过传统方法和CNN方法构建的局部特征描述符匹配性能不佳，而CGNet利用GNN学习特征点之间的位置关系，丰富了局部描述符的上下文信息，使得描述符具有良好的旋转与尺度不变性，显著提高了匹配性能。同时，针对难以通过标注获得匹配特征点真值的问题，CGNet利用预先对齐的配准数据集设计了一套弱自监督的训练方法，可以端到端直接输出两幅图像的特征点匹配关系，具有很高的计算效率。CGNet可拓展到其他类型的异源图像配准任务中，如红外与可见光图像配准、医学图像配准等。

参 考 文 献

[1] 谢志华, 刘晶红, 孙辉, 等. 可见光图像与合成孔径雷达图像的快速配准[J]. 激光与光电子学进展, 2020, 57(6): 062803.
Xie Z H, Liu J H, Sun H, et al. Fast registration of visible light and synthetic aperture radar images[J]. Laser & Optoelectronics Progress, 2020, 57(6): 062803.

[2] Li J Y, Hu Q W, Ai M Y. RIFT: multi-modal image matching based on radiation-variation insensitive feature transform[J]. IEEE Transactions on Image Processing, 2020, 29: 3296-3310.

[3] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[4] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 2564-2571.

[5] 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.

[6] Xiang Y M, Wang F, You H J. 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.

[7] Kovesi P. Phase congruency detects corners and edges [C]//The Australian Pattern Recognition Society Conference, December 10-12, 2003, Sydney. [S.l.: s.n.], 2003: 309-318.

[8] Ye Y X, Shan J, 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.

[9] Li Z Y, Zhang H T, Huang Y H. A rotation-invariant optical and SAR image registration algorithm based on deep and Gaussian features[J]. Remote Sensing, 2021, 13(13): 2628.

[10] 孙明超, 马天翔, 宋悦铭, 等. 基于相位特征的可见光和SAR遥感图像自动配准[J]. 光学 精密工程, 2021, 29(3): 616-627.
Sun M C, Ma T X, Song Y M, et al. Automatic registration of optical and SAR remote sensing image based on phase feature[J]. Optics and Precision Engineering, 2021, 29(3): 616-627.

[11] 李泽一, 赵薇薇, 喻夏琼, 等. 基于最大相位索引图的异源影像配准方法[J]. 中国激光, 2021, 48(15): 1509002.
Li Z Y, Zhao W W, Yu X Q, et al. Registration of heterologous images based on maximum phase index map [J]. Chinese Journal of Lasers, 2021, 48(15): 1509002.

[12] Zhang J, Ma W P, Wu Y, et al. Multimodal remote sensing image registration based on image transfer and local features[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(8): 1210-1214.

[13] Toriya H, Dewan A, Kitahara I. SAR2OPT: image alignment between multi-modal images using generative adversarial networks[C]//IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, July 28-August 2, 2019, Yokohama, Japan. New York: IEEE Press, 2019: 923-926.

[14] Hughes L H, Schmitt M, Mou L C, et al. Identifying corresponding patches in SAR and optical images with a pseudo-siamese CNN[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5): 784-788.

[15] Zhang H, Ni W P, Yan W D, 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.

[16] DeTone D, Malisiewicz T, Rabinovich A. SuperPoint: self-supervised interest point detection and description [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 18-22, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 337-349.

[17] Dusmanu M, Rocco I, Pajdla T, et al. D2-net: a trainable CNN for joint description and detection of local features[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 8084-8093.

[18] 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 C Z, Lu W J, Yu J M, 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.

[19] Luo Z X, Shen T W, Zhou L, et al. ContextDesc: local descriptor augmentation with cross-modality context[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 2522-2531.

[20] Sarlin P E, DeTone D, Malisiewicz T, et al. SuperGlue: learning feature matching with graph neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June

13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 4937-4946.

[21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04) [2021-05-06]. https://arxiv. org/abs/1409.1556.

[22] Sinkhorn R, Knopp P. Concerning nonnegative matrices and doubly stochastic matrices[J]. Pacific Journal of Mathematics, 1967, 21(2): 343-348.

[23] WangR Z, YanJ C, YangX K. Learning combinatorial embedding networks for deep graph matching[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3056-3065.