DOI:10.19651/j.cnki.emt.2519081

融合双域特征与跨维门控注意力的遥感图像配准*

魏庭旭[1] 陈 颖[1] 李铖昊[2] 马文浩[1]
(1.上海应用技术大学智能技术学部 上海 201418;2.同济大学中德工程学院 上海 201804)

摘 要: 针对遥感图像配准中复杂环境带来的特征提取困难、多尺度几何变形造成的配准精度受限等挑战,本文提出一种融合双域特征与跨维门控注意力的遥感图像配准模型。首先,在特征提取阶段设计多尺度傅里叶模块改进StarNet网络结构通过融合多尺度的空间特征与频域特征进行特征提取,增强模型的特征提取能力;接着,设计跨维门控注意力,使得模型能够在不牺牲全局感受野的情况下,高效地捕捉图像中的上下文信息;其次,在特征匹配阶段,采用基于部分分配矩阵的双向匹配方法获取双向参数,最终通过仿射变换实现图像配准。在使用Aerial Image数据集进行实验时,研究结果表明,当正确估计的关键点比例系数分别设置为0.01、0.03和0.05时,配准精度分别达到了42.8%、85.7%和96.9%,且平均配准时间为0.87 s,有效提升了遥感图像配准的精度和速度。

关键词: 遥感图像配准;多尺度傅里叶提取;多尺度特征;跨维门控注意力;双向匹配
中图分类号: TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Remote sensing image registration integrating dual-domain feature and cross-dimensional gated attention

Wei Tingxu[1] Chen Ying[1] Li Chenghao[2] Ma Wenhao[1]
(1. Faculty of Intelligence Technology, Shanghai Institute of Technology, Shanghai 201418, China;
2. College of Sino-German Engineering, Tongji University, Shanghai 201804, China)

Abstract: Aiming at the challenges of remote sensing image registration such as feature extraction difficulties caused by complex environment and registration accuracy limitations caused by multi-scale geometric deformation, this paper proposes a remote sensing image registration model that integrates dual-domain features and cross-dimensional gated attention. Firstly, the multi-scale Fourier module is designed in the feature extraction stage to improve the StarNet network structure to enhance the feature extraction capability of the model by fusing the multi-scale spatial features with the frequency domain features; then, the cross-dimensional gated attention is designed so that the model can efficiently capture the contextual information in the image without sacrificing the global sensing field; secondly, the feature matching stage bidirectional parameters are obtained by applying bidirectional matching based on partial assignment matrix, and finally, the registration is completed by affine transformation. In the experiments using the aerial image dataset, the results show that when the correctly estimated keypoint scale factor is set to 0.01, 0.03 and 0.05, the registration accuracy reaches 42.8%, 85.7% and 96.9%, respectively, and the average registration time is 0.87 s, which significantly improves the accuracy and speed of remote sensing image registration.

Keywords: remote sensing image registration; multi-scale Fourier extraction; multi-scale features; cross-dimensional gated attention; bi-directional matching

0 引 言

遥感图像配准作为空间信息处理的关键环节,其核心任务在于实现不同时空条件下采集的多元图像数据的空间一致性校准。近年来,随着对地观测技术的突飞猛进,海量遥感数据的获取使得这项技术在国土监测、智慧城市建设、生态评估及国防安全等领域的应用价值愈发显著[1]。然而,值得注意的是,当前遥感图像配准工作仍存在若干技术瓶颈,包括影像尺度差异显著、信噪比波动、辐射条件变化以及观测角度差异等复杂因素,这些变量都会对配准结果

收稿日期:2025-06-12
*基金项目:国家自然科学基金面上项目(61976140)、上海应用技术大学协同创新基金项目(XTCX2022-25)资助

的准确性和计算性能产生直接影响。

遥感图像配准领域中，传统的方法主要划分为灰度相关配准和特征点配准两大类别。灰度相关的方法以 Ma 等[2]的研究为代表，通过直接比对像素亮度值实现影像对齐。这类方法对几何形变具有较好的适应性，但容易受到噪声干扰和光照条件的影响。特征点配准技术则依赖图像中的关键点或显著区域完成配准，如 Zhang 等[3]改进的 SIFT 算法，展现了优异的尺度不变性和抗干扰特性，但也存在计算耗时长、人工干预过多等缺点。在处理高分辨率、多源异构的遥感数据时，上述方法在运算效率和匹配精度方面往往难以满足实际工程需求。

当前深度学习技术的突飞猛进使得端到端配准方法逐步替代了传统方法。Park 等[4]提出了双向集成对称的双流网络(two-stream)，有效改善了配准结果不对称的缺陷，但其特征提取模块未能充分建模远距离依赖关系，在复杂场景下的配准效果仍有提升空间。陈颖等[5]使用参合成空间变换网络(parameter-weighted synthesis spatial transformation network，PWSTNet)通过参数化空间变换学习，在遥感影像配准任务中展现出优异的非线性变换建模能力，显著提升了配准质量。Tang 等[6]提出 MU-Net 利用注意力机制来提升配准难题，但在复杂的光学遥感图像中配准性能不佳。Yao 等[7]提出双维特征融合网络(spatial-frequency dual-domain feature fusion network，SFDDFF-Net)，融合空间域和频域对遥感图像进行特征提取，性能得到提升，但没有考虑多尺度特征，导致特征提取不充分。唐海林等[8]将 RTformer 与注意力机制相结合，有效提升了模型提取遥感图像特征的表征能力。Geng 等[9]将 Transformer 与 CNN 架构相融合，在遥感语义分割任务中取得了精度突破，但传统 Transformer 的平方级计算复杂度制约了运算效率。Liang 等[10]提出一种基于 Transformer 的粗精特征匹配网络(coarse-to-fine feature match network，CFFMNet)，在弱纹理条件下提高了遥感图像配准的性能，然而粗精匹配策略的运算效率不高。

为了提高遥感图像配准的精度和效率。本文提出一种融合双域特征与跨维门控注意力的端到端遥感图像配准算法。该算法增强了模型的特征表示能力，有效提高了遥感图形配准的精确性和鲁棒性。本文的主要贡献为：1)设计多尺度傅里叶提取模块(multi-scale Fourier extraction module，MSFM)融合多尺度空间域与频域特征，嵌入到 StarNet[11]特征骨干网络中，提升网络的特征提取能力和效率；2)提出跨维门控注意力(cross-dimensional gated attention，CDGA)应用到特征提取阶段，高效地聚合上下文信息；3)设计基于部分分配矩阵的双向匹配层，以提升配准精度。

1　本文方法

1.1　模型结构

CNNGeo[12]是当前遥感图像配准领域的一个经典基准模型，然而其主要依赖空间域特征，未能充分处理图像中的多尺度几何变形和复杂背景信息，并且单向匹配导致匹配不充分。在复杂场景下，其配准精度受限。倪力政等[13]对 CNNgeo 进行改进，提出注意力多尺度配准网络(attention and multi-scale features registration network，AMR-Net)，利用自注意力机制从空间和通道两个维度对图像进行特征提取，提升了配准性能，但自注意力机制的缺乏局部特征提取能力，导致模型提取的上下文信息不完整，且计算复杂度较高；融合多尺度的思想提取图像在不同感受野下的空间特征，但普通卷积提取到的感受野有限，且仅考虑空间维度，忽略了特征之间的长距离依赖关系。本文的改进策略区别于 AMR-Net，在运用注意力机制时不仅考虑其全局信息提取而且加入局部信息提取能力是模型更全面的提取上下文信息；本文也将运用多尺度思想，但不同于 AMR-Net 的是，本文使用更大的卷积核提取感受野，并且结合频域信息特征使模型提取更鲁棒的信息。

本文提出了一种基于深度学习的新型端到端遥感图像配准框架，以 CNNGeo 为基准模型，如图 1 所示，本研究的创新之处在于通过双域特征和跨维门控注意力机制的结合，全面提升了遥感图像配准的精度和效率。

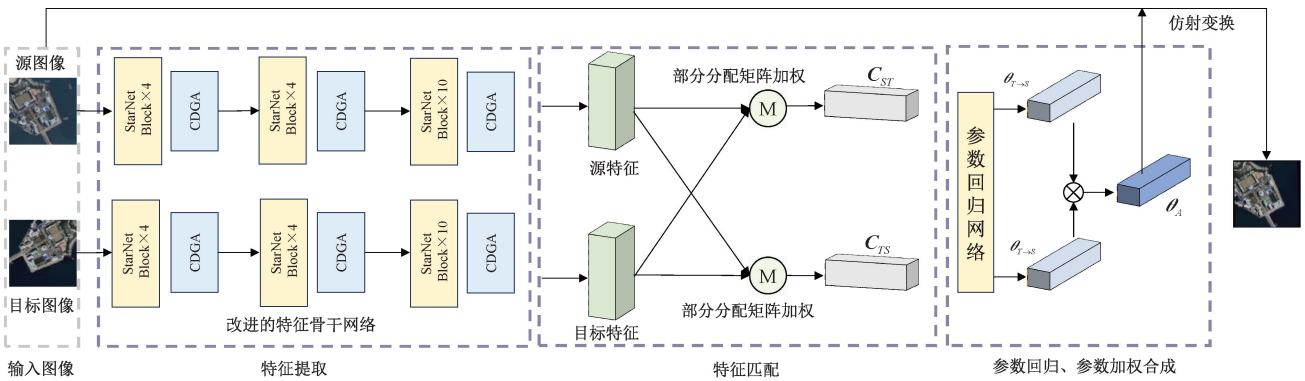

图 1　整体配准框架

Fig. 1　Overall registration framework

首先,在特征提取阶段选用轻量且高效的 StarNet 网络作为特征骨干网络。由于遥感图像具有复杂的背景干扰和多尺度几何形变的特点,本文设计多尺度傅里叶提取模块(MSFM)来改进 StarNet 网络,并嵌入跨维门控注意力(CDGA),同时接收源图像与目标图像,生成密集特征图,作为图像的特征描述符。接着,在匹配阶段,采用基于部分分配矩阵加权的双向匹配策略,以确保特征匹配的高效性和准确性。通过这一策略,生成了双向密集对应关系,随后这些匹配关系被输入到参数回归网络中,进一步回归出双向变换参数 $\boldsymbol{\theta}_{T\to S}$ 和 $\boldsymbol{\theta}_{S\to T}$ 并加权合成得到最后的仿射参数 $\boldsymbol{\theta}_A$,最终,通过仿射参数对原始图像进行仿射变换,从而得到最终的配准图像。目的是在遥感图像复杂背景干扰、多尺度几何形变的情况下提升配准的精确性和稳定性。

1.2 特征提取

1)多尺度傅里叶模块(MSFM)

由于成像机制,遥感图像配准存在目标尺度变化大和图像信息冗余等问题,这给配准任务带来了挑战。现有的方法大多通过大核卷积扩大空间感受野来解决,但是大核卷积通常会引入相当大的背景噪声,而且单一尺度的特征提取往往不能充分提取图像的不同阶特征之间的信息交互[14]。此外,傅里叶变换能够将图像从空间域转换到频域,其中每个像素对应于所有空间像素,使网络能够更高效地捕捉图像的整体结构并建立长程相关性[15]。因此,本文设计多尺度傅里叶提取模块来改进 StarNet 网络的网络结构,旨在高效提取遥感图像的局部和全局特征,过滤冗余信息,提高配准的鲁棒性。该模块通过使用多尺度空间域-频域级联结构进行双域特征提取,以改善复杂场景中的去噪,从而提升网络的特征表征能力。具体结构如图 2 所示。

具体而言,MSFM 首先在空间域对输入特征进行特征提取,生成空间域特征图。使用 3×3 小核深度卷积对凝练输入特征的局部信息,随后是一组并行深度卷积以捕获跨多个尺度的空间特征。然后,通过大小为 1×1 的逐点卷积融合局部和全局特征,表征各种通道之间的相互关系。1×1 逐点卷积作为通道融合机制,将不同感受野大小的特征融合在一起,MSFM 就可以在不影响局部纹理特征完整性的情况下,捕获更广泛的空间特征。空间域特征的计算流程如式(1)所示。

$$\boldsymbol{X}_{\text{spatial}} = PConv\left(GeLU\left(x + \sum_{m=1}^{4} \boldsymbol{X}^{(m)}\right)\right) \tag{1}$$

其中,$PConv(\cdot)$ 为 1×1 逐点卷积。

随后,在频域运算中,MSFM 对空间特征 $\boldsymbol{X}_{\text{spatial}}$ 进行离散傅里叶变换(discrete Fourier transform, DFT)[16],将其转换为实部和虚部,其包含了图像的全局信息。在频域中 MSFM 对实部和虚部应用卷积操作,对其进行空间交互,进一步增强和提取全局特征。然后,将处理后的实部和虚部通过逆离散傅里叶变换转换回空间域。计算流程如

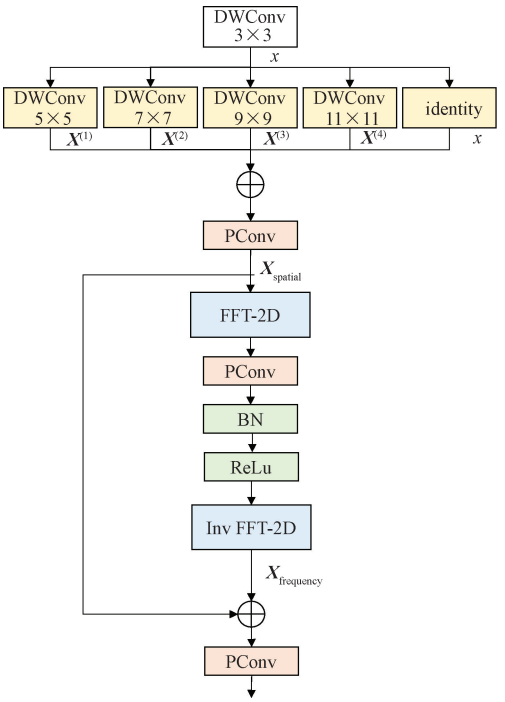

图 2 MSFM 模块结构

Fig. 2 MSFM module framework

式(2)~(5)所示。

$$\mathcal{F}(u,v) = \frac{1}{\sqrt{HW}} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} X(h,w) \mathrm{e}^{-\mathrm{j}2\pi\left(\frac{hu}{H}+\frac{wv}{W}\right)} \tag{2}$$

$$\boldsymbol{X}_R, \boldsymbol{X}_L = \mathcal{F}(\boldsymbol{X}_{\text{spatial}}) \tag{3}$$

$$\hat{\boldsymbol{X}}_R, \hat{\boldsymbol{X}}_L = ReLU \cdot BN(PConv(Concat(\boldsymbol{X}_R, \boldsymbol{X}_L))) \tag{4}$$

$$\boldsymbol{X}_{\text{frequency}} = \mathcal{F}^{-1}(\hat{\boldsymbol{X}}_R, \hat{\boldsymbol{X}}_L) \tag{5}$$

式(2)为 DFT 的计算公式,其中 $\mathcal{F}(u,v)$ 基于傅里叶空间作为复分量,u 和 v 是傅里叶空间的坐标。

最后,MSFM 将处理后的频域特征与原始的空间域特征通过残差连接进行融合,并通过 1×1 卷积进行通道交互,生成包含丰富全局和局部信息的输出特征图。如式(6)所示。

$$\boldsymbol{X}_{\text{out}} = PConv(\boldsymbol{X}_{\text{spatial}} + X_{\text{frequency}}) \tag{6}$$

MSFM 通过频域全局建模和多尺度空间特征提取的结合,不仅能够高效地捕捉全局信息,还能保留局部细节,从而为遥感图像配准任务提供更全面的特征表达。这种设计在降低计算成本的同时,提高了配准精度,特别是在处理高分辨率遥感图像时更具优势。

改进的 StarNet block 如图 3 所示,通过嵌入 MSFM,将多尺度空间域和频率域的特征信息结合起来,利用两者的互补性,提高图像配准的效果,聚焦图像的关键区域。

2)跨维门控注意力(CDGA)

高分辨率的遥感图像包含丰富的地理结构信息以及复杂的背景信息,所以,高效地提取遥感图像中的上下文

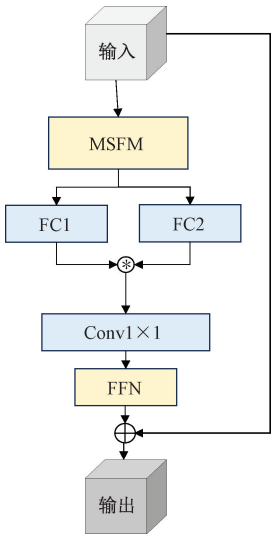

图 3　改进的 StarNet block 结构

Fig. 3　Improved StarNet block structure

信息尤为关键[17]。传统的自注意力具有良好的全局上下文建模能力，通常在空间维度进行点积运算，但这样往往忽略了特征通道之间的复杂关系，而且建立局部上下文信息关系的能力不足，导致特征表达不足，影响配准精度。此外，传统的全局建模方法由于其二次复杂度，在处理高分辨率遥感图像时计算成本过高，且容易引入冗余信息。针对以上问题，本文改进传统自注意力机制，并设计门控单元，组合为跨维门控注意力（CDGA）应用在特征提取阶段，旨在通过在通道维度计算自注意力和门控特征转换高效地聚合上下文信息，同时降低计算复杂度，提高图像配准的全面性。CDGA 的结构如图 4 所示。

改进的自注意力。为了弥补传统自注意力局部上下文信息建模能力不足的问题，本文在 1×1 卷积层之后利用卷积核大小为 3×3 的深度卷积生成 $\boldsymbol{Q}$，$\boldsymbol{K}$，$\boldsymbol{V}$ 替代传统的直接使用 1×1 卷积的方式，以强调局部上下文之间的特征交互，同时参考协方差注意力（covariance attention，

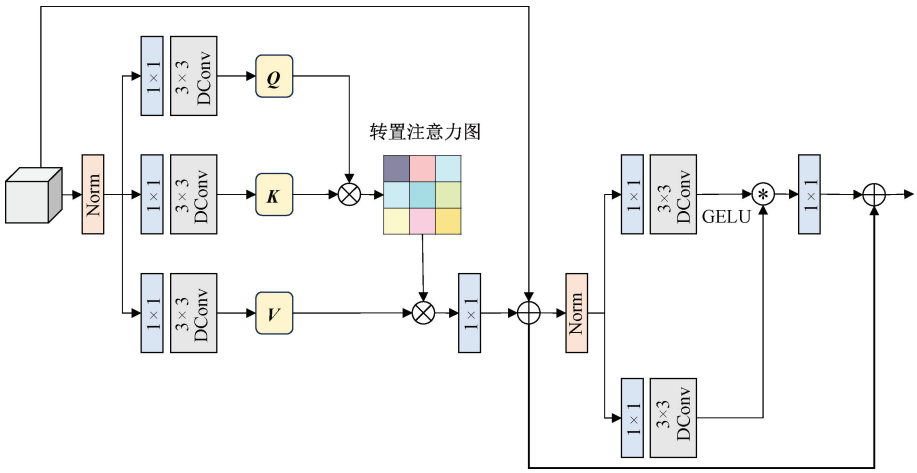

图 4　CDGA 模块结构

Fig. 4　CDGA module framework

CA）[18]，跨通道维度计算自注意力，即计算跨通道的互协方差以生成隐式编码全局上下文的注意力图，增强图像通道之间的交互关系。

初始化输入特征 $\boldsymbol{Y} \in \mathbb{R}^{H\times W\times C}$，复制为 3 个分支，首先利用 1×1 卷积融合每个像素的跨通道上下文信息，然后通过 3×3 深度卷积来计算逐通道的空间上下文信息，从而生成包含丰富的局部上下文信息的查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 投影，然后，对查询和关键投影进行重塑，通过点积运算在通道维度生成大小为 $\mathbb{R}^{C\times C}$ 的转置注意力图 $\boldsymbol{A}$，$\boldsymbol{A}$ 的大小远小于传统注意力图的大小 $\mathbb{R}^{HW\times HW}$。转置注意力图隐式地编码了全局上下文信息。改进的自注意力计算过程如式(7)所示。

$$\hat{\boldsymbol{Y}} = PConv(Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})) + \boldsymbol{Y} \tag{7}$$

通过跨通道协方差计算，隐式地捕捉全局上下文信息，同时利用深度卷积增强局部上下文，确保全局建模时保留局部细节。这种设计不仅降低了计算复杂度，还提高了特征表达的一致性和鲁棒性。

门控单元。常规前馈神经网络为了转换特征分别对每个像素位置进行相同操作，然后连续使用两个 1×1 卷积对特征图进行升维和降维操作，本文所设计的门控单元对其进行修改以接收来自改进自注意力的输出并增强特征表示能力。首先是设计两个并行分支，利用线性层的逐元素乘积实现门控机制，同时每个分支同样利用 3×3 深度卷积以编码来自空间相邻像素位置的信息，其中的一条分支使用 GELU 激活函数，与另一条分支的输出进行逐元素相乘并进行通道融合，最后利用残差链接与转置注意力图进行相加得到最终输出。计算流程如式(8)～(9)所示。

$$Gating(\cdot) = GeLU(\boldsymbol{W}_d^1\boldsymbol{W}_p^1(LN(\cdot)))\odot\boldsymbol{W}_d^2\boldsymbol{W}_p^2(LN(\cdot)) \tag{8}$$

$$\boldsymbol{Y}_{\text{out}} = PConv(Gating(\hat{\boldsymbol{Y}}) + \hat{\boldsymbol{Y}}) \tag{9}$$

式(8)中，$\boldsymbol{W}_p^{(\cdot)}$ 为 1×1 逐点卷积，$\boldsymbol{W}_d^{(\cdot)}$ 为 3×3 深度卷积，$\odot$ 为逐元素相乘。

门控单元专注于利用上下文信息来丰富图像特征，并通过门控单元控制信息流，确保关键特征在层次间高效传递。这种特征选择与信息优化机制特别适合处理遥感图像中的噪声和非目标区域干扰，从而提升配准精度。

1.3　基于部分分配矩阵的双向匹配

特征匹配层旨在通过匹配函数计算由特征提取网络提取的两幅图像特征之间的对应关系。传统的仅在源图像的语义信息到目标图像的语义信息单方向的匹配方式会使得匹配性能受限[12]。而且没有充分交互两幅图像特征之间的信息。为解决以上问题，本文设计了基于部分分配矩阵的双向匹配层，引入部分分配矩阵[19]，通过计算两幅图像特征描述符之间的成对得分矩阵和匹配性分数，随后根据二者计算出部分分配矩阵，更精确地筛选出正确的匹配点对，同时剔除不可匹配的点。同时参考 Two-strame 的设计，引入双向匹配策略，构建从源图像到目标图像以及目标图像到源图像的双向密集对应关系来缓解匹配不对称问题。计算公式如式(10)所示。

$$\boldsymbol{S}_{ij} = Linear(\boldsymbol{f}_T(\boldsymbol{i},\boldsymbol{j}))^{\top} Linear(\boldsymbol{f}_S(\boldsymbol{i}_k,\boldsymbol{j}_k)) \quad (10)$$

其中，$\boldsymbol{S}_{ij}$ 为源图像 I_S 与目标图像 I_T 之间的成对得分矩阵，$\boldsymbol{f}_T(\boldsymbol{i},\boldsymbol{j})$ 表示在坐标 (i,j) 处的目标特征描述符，$\boldsymbol{f}_S(\boldsymbol{i}_k,\boldsymbol{j}_k)$ 表示在坐标 (i_k,j_k) 处的源特征描述符。

$$\sigma = Sigmoid(Linear(\cdot)) \in [0,1] \quad (11)$$

式(11)为每个点计算匹配性分数。

$$\boldsymbol{P}_{ij} = \sigma_i^{\mathrm{T}}\sigma_j^{S} \sum Sorft\max(\boldsymbol{S}_{kj})_i \sum Sorft\max(\boldsymbol{S}_{ik})_j \quad (12)$$

$$\boldsymbol{C}_{ST}(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) = C(f_T(i,j)P_{i,j}, f_S(i_k,j_k)P_{i,j}) \quad (13)$$

将成对得分矩阵与匹配性分数组合成为部分分配矩阵 $\boldsymbol{P}_{ij}$，用于加权计算最终从源图像到目标图像之间的对应关系 $\boldsymbol{C}_{ST}(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})$。

以上过程公式为计算从源图像到目标图像之间的匹配对应关系，随后以同样的方式计算从目标图像到源图像的对应关系，最后，将双向密集对应关系传入到参数回归网络。

1.4　参数回归

将双向匹配层传来的双向密集对应关系 $\boldsymbol{C}_{ST}$ 和 $\boldsymbol{C}_{TS}$ 依次送入两个卷积核大小分别为 7×7 和 5×5 的卷积层，在每次卷积之后执行批归一化和 RELU 激活函数处理，估计双向参数 $\boldsymbol{\theta}_{S\to T}$ 和 $\boldsymbol{\theta}_{T\to S}$ 的值。随后，使用网格距离损失函数来计算参数变换前后的损失，网格距离损失函数如式(14)所示。

$$l(\hat{\theta},\theta^{gt}) = \frac{1}{N}\sum_{i,j=1}^{N} d(\boldsymbol{T}_{\hat{\theta}}(\boldsymbol{x}_i,\boldsymbol{y}_j),\boldsymbol{T}_{\theta^{gt}}(\boldsymbol{x}_i,\boldsymbol{y}_j))^2 \quad (14)$$

其中，分母部分：N 为总共的网格点数量；分子部分：θ^{gt} 和 $\hat{\theta}$ 分别为真值参数和变换后的参数，d 为变换前图像的真值点 $\boldsymbol{T}_{\theta^{gt}}(\boldsymbol{x}_i,\boldsymbol{y}_j)$ 与变换后图像上点 $\boldsymbol{T}_{\hat{\theta}}(\boldsymbol{x}_i,\boldsymbol{y}_j)$ 之间的平方差。

1.5　参数加权合成

对参数回归网络得到的双向变换参数进行加权合成，参考 Tow-stream，并通过对 $\boldsymbol{\theta}_{T\to S}$ 取逆以保持参数的方向一致性。该过程涉及仿射变换所需的 6 个参数 $[\partial_1,\partial_2,\partial_3,\partial_4,\partial_5,\partial_6]$，即平移、旋转等，确保了变换过程的准确性和一致性。取逆的计算过程如式(15)所示。

$$(\boldsymbol{\theta}_{T\to S})^{-1} = \begin{pmatrix} \partial_1 & \partial_3 & \partial_5 \\ \partial_2 & \partial_4 & \partial_6 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \quad (15)$$

在加权合成过程中，正向变换参数 $\boldsymbol{\theta}_{S\to T}$ 与逆运算后的 $\boldsymbol{\theta}_{T\to S}$ 进行加权，生成最终的合成变换参数 $\boldsymbol{\theta}_A$。在加权合成中，采用了算术均值函数进行融合，如式(16)所示。

$$\boldsymbol{\theta}_A = \mu(\boldsymbol{\theta}_{S\to T},(\boldsymbol{\theta}_{T\to S})^{-1}) \quad (16)$$

其中，$\mu(\cdot)$ 是融合两个参数的算术均值函数。

2　实验与分析

2.1　实验数据集与环境

训练集使用 Aerial Image 数据集，该数据集来自 Google Earth，包含多种类型的遥感航空图像由不同类型传感器在不同日期的不同视角拍摄。包括城市、沙漠、海洋、山脉等不同类型环境，共有 9 000 对图像。为全面评估模型性能，测试集整合了 3 个基准数据集：Aerial Image 数据集、SUIRD[20]数据集和 MRSI[21]数据集作为测试集来评估模型的性能。SUIRD 数据集包括超过 1 200 幅小型无人机航拍图像。MRSI 数据集是各种场景下的多模态图像数据集。

实验的开发环境方面，实验在 Windows 11 操作系统下运行，PyTorch 作为深度学习框架，使用 Python3.9 和 CUDA 11.7。训练网络方面，学习率设定为 1×10^{-4}，批处理大小设置为 4，权重衰减系数设为 0，动量参数为 0.8，进行 100 个 epoch 的训练。硬件配置方面，使用了 NVIDIA GeForce RTX 3080 显卡，其运行内存为 12 GB。

2.2　评价指标

本研究通过定性和定量评估的方式，全面验证了配准算法的效果，确保了算法在实际应用中的精确性与高效性。

在定性评估方面，研究利用棋盘格图和配准结果图来分析配准效果。棋盘格图用于直观展示特征点的对齐情况，而配准结果图则能够展示配准后图像之间的空间关系和变形情况，从而提供配准质量的可视化评估。

在定量评估方面，研究使用了 4 个关键指标来量化配准效果：

关键点正确估计比例（percentage of correct keypoints，PCK）：此指标用于衡量配准过程中，正确估计的关键点比例，能够反映特征匹配的准确性。其计算如

式(17)所示。

$$PCK = \frac{\sum_{i=1}^{N}[d(\mathcal{T}(p_{Si}), p_{Ti}) < \tau \cdot \max(h, w)]}{N} \tag{17}$$

其中，分子部分表示检测正确关键点的数量，$\mathcal{T}(\cdot)$ 代表仿射变换，p_{Si} 是源图像第 i 个关键点经过仿射变换后的坐标，p_{Ti} 是第 i 个图像对基准真值点。$\tau \cdot \max(h, w)$ 中的尺寸参数 h，w 分别表示图像的高度和宽度，τ 与定位精度要求呈负相关，τ 越小，以真值点为中心的允许偏差半径同步缩小，评估标准随之严格化。分母 N 表示图像关键点的总数。PCK 有效衡量模型在定位和对齐关键点上的准确性，值越高表明图像对对齐得越好，通常将 τ 设置为 0.05。

平均绝对误差(mean absolute error，MAE)：MAE 量化了配准结果与真实值之间的平均差异，帮助评估配准的几何精度。

均方根误差(root mean square error，RMSE)：RMSE 作为误差的标准化度量，提供了配准结果的准确性分析，尤其适用于衡量较大误差的影响。

MAE 和 RMSE 如式(18)～(19)所示。

$$MAE = \frac{1}{N}\sum_{i=1}^{N} \| (x_i, y_i) - (\hat{x}_i, \hat{y}_i) \| \tag{18}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \| (x_i, y_i) - (\hat{x}_i, \hat{y}_i) \|^2} \tag{19}$$

其中，分母 N 表示检测到的所有关键点的数量，(x_i, y_i) 和 $(\hat{x}_i, \hat{y}_i)$ 分别是模型检测到的关键点的坐标和真值关键点的坐标。

配准所用时间(Time)：该指标衡量配准算法的计算效率，帮助评估其在实际应用中的实时性。

2.3　实验结果与分析

本文选取了 Aerial Image、SUIRD 和 MRSI 数据集上的四组遥感图像对进行图像配准测试，目的是验证算法在不同数据集和场景下的表现。图 5 展示了源图像、目标图像、配准结果图以及目标图像与配准结果图叠加形成的棋盘格图。能够清晰地看到配准效果以及配准后图像的对齐质量。

在测试中，第 1 组图像来自 Aerial Image 数据集，图像拍摄于不同时间和角度的海港码头，拍摄于不同时间和角度的海港码头，且光照强度变化大。配准结果显示，防波堤轮廓基本对齐，陆地建筑配准良好；第 2 组图像来自 Aerial Image 数据集的公路与农田交织的场景，周边散布着建筑物，且拍摄角度与时间不同。通过变换，公路曲线与农田轮廓基本对齐；第 3 组图像源于 SUIRD 数据集的复杂城市场景，包含密集的房屋和道路信息，配准后边界几乎重叠；第 4 组图像来自 MRSI 数据集，拍摄于不同传感器的山脉深度图与光学图像，通过配准后，棋盘格图显示山脉的轮廓大体对齐。

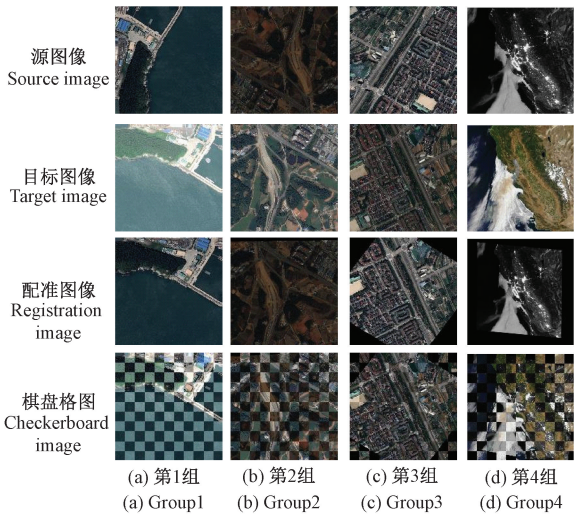

图 5　算法配准结果

Fig. 5　Algorithm registration results

实验结果表明本文提出的算法在各类复杂场景图像中，配准结果均与目标图像基本对齐，验证了算法在 Aerial Image、SUIRD 和 MRSI 数据集上的有效性。

2.4　对比实验

为全面验证本研究算法的综合性能，设计了系统的对比实验方案。在实验设计上，选用 SIFT、CNNGeo、CFFMNet、SFDDFF-Net 和 CIRSM-Net[22] 五种具有代表性的配准算法进行对比分析：首先选取 SIFT，为经典的传统配准算法；选取 CFFMNet、SFDDFF-Net 和 CIRSM-Net 这 3 个近年来在遥感图像配准领域备受关注的端到端深度学习方法，以此构建完整的对比实验体系。

1)定性评估

选取的测试集图像对包括：第 1 组来自不同时间拍摄的海岸图；第 2 组来自不同角度拍摄的港口码头；第 3 组为乡村复杂场景图；第 4 组为来自不同传感器的鸟瞰城市图。棋盘格图是将配准图像与目标图像分成若干个，每个方块交替拼接的图像。本文采用棋盘格可视化分析方法对 6 种算法的配准效果进行系统性评估。棋盘格图中采用红色方框作为配准误差的视觉标识符，方框越大表明对应区域误差越大。与上述 5 种算法进行定性对比分析，6 种算法配准结果对比图与棋盘格对比图如图 6～7 所示。

图 7 为 6 种算法在上述测试集下配准后的棋盘格对比图，实验结果显示，传统算法的棋盘格图误差极大。配准效果较好的 CFFMNet、SFDDFF-Net 和 CIRSM-Net 的棋盘格图在图像中目标对象的轮廓细节上还存在偏差。本文方法所展现的实验结果误差最小，配准图像与目标图像基本对齐，仅有少量的误差。经过分析实验结果，本文算法与其他 5 种算法相比配准效果最佳。

本文提出的算法通过融合多尺度空间域与频域特征，配准结果在整体轮廓与局部细节上的高度对齐。同时，算

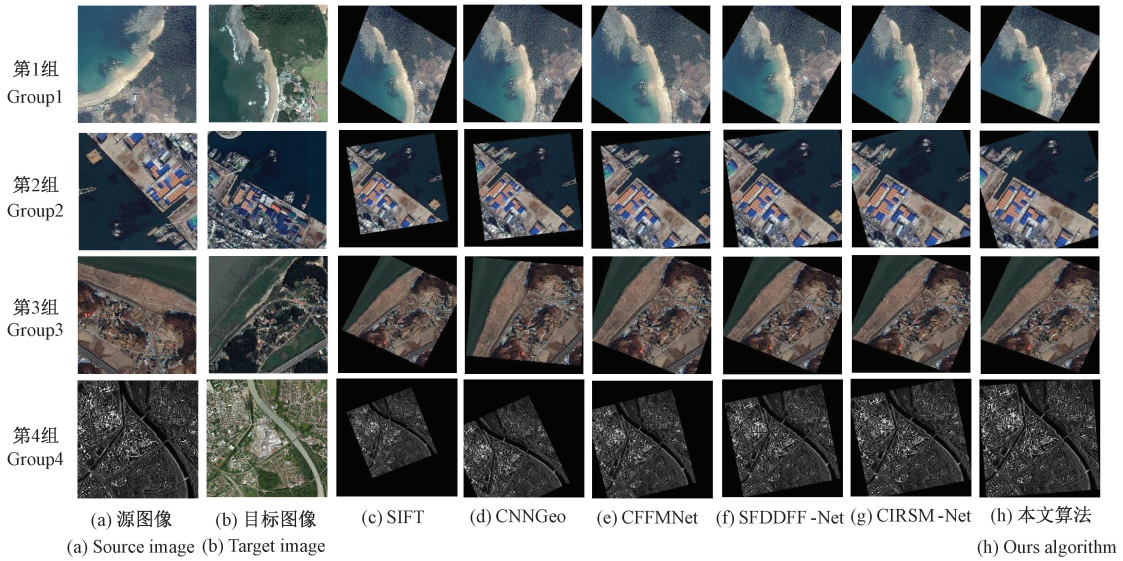

图 6　6 种算法配准结果图对比

Fig. 6　Comparison of the registration results of the six algorithms

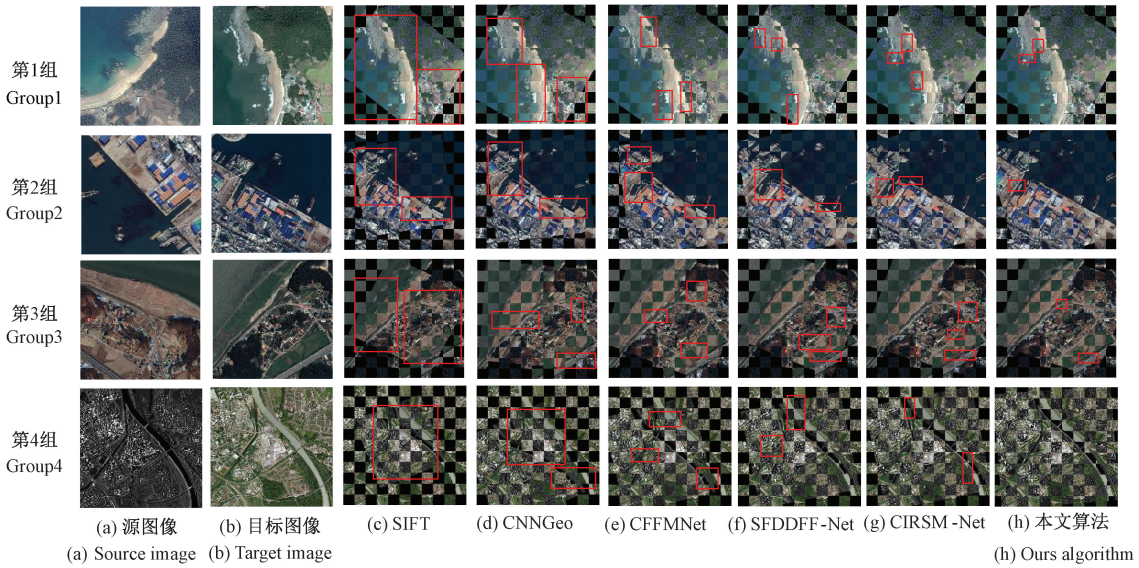

图 7　6 种算法棋盘格图对比

Fig. 7　Comparison chart of the six algorithmic checkerboards

法注重关键细节特征提取,使得在纹理复杂的遥感图像中也能取得优异表现。

2)定量评估

定量分析的测试集如图 8 所示,包含多时相、多视角、多模态和复杂情况 4 种类型的遥感图像对。并使用 MAE、RMSE 和 Time 3 个评估指标对 6 种算法进行评估,MAE 和 RMSE 的比较结果如表 1 所示,最佳数据已用粗体标出。从表 1 中可以看出,传统算法由于缺乏深度学习的特征提取和自动优化能力,在复杂的遥感图像配准任务中表现不佳,导致配准误差较大,在 4 种不同场景下的配准误差显著高于其他端到端深度学习遥感图像配准方法。CFFMNet、SFDDFF-Net 和 CIRSM-Net 方法虽然误差相对较低,但在多模态和复杂情况下表现不佳,表明它们在由复杂背景干扰的情况下配准性能欠佳,无法有效地处理多变的场景和干扰因素。相比之下,本文提出的算法在 MAE 和 RMSE 方面的配准误差最低,展现了优异的配准性能,通过融合双域特征与注意力机制的优化,本文算法能够更好地应对复杂背景、不同模态和几何变形问题,从而提供更加精准和稳定的遥感图像配准结果。

图 9 展示了 6 种不同算法在遥感图像配准任务中的配准时间评价结果,从中可以直观地比较各算法的计算效率。结果表明,本文提出的算法在配准时间方面明显优于传统 SIFT 方法。这说明所提方法不仅在精度上有所突破,同时在运行效率上也取得了显著提升。

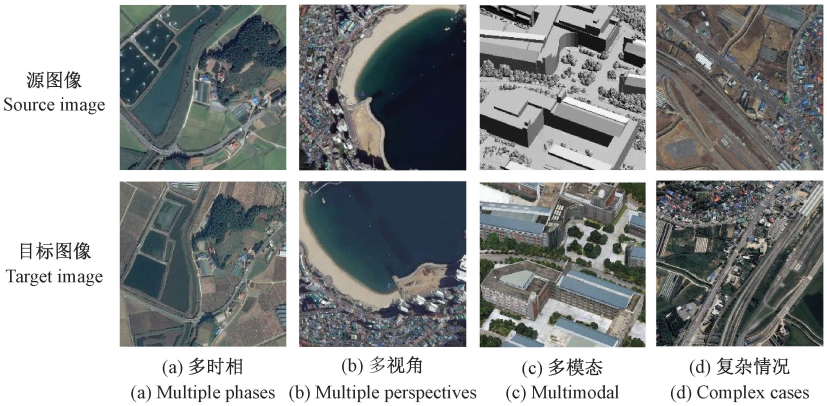

图 8　4 种类别测试图像对

Fig. 8　Test image pairs of four categories

表 1　6 种算法在 4 种情况下 MAE 和 RMSE 值对比

Table 1　Comparison of MAE and RMSE values in four cases by six algorithms

方法	多时相		多视角		多模态		复杂情况	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
SIFT	14.73	8.26	14.63	5.061	—	—	—	—
CNNGeo	3.51	4.05	4.37	3.89	5.11	5.32	5.26	4.71
CFFMNet	1.73	2.04	2.32	2.63	2.96	3.01	1.71	2.32
SFDDFF-Net	1.64	1.95	1.75	2.37	2.56	2.88	1.89	2.71
CIRSM-Net	1.58	1.74	1.62	2.23	2.38	2.63	1.76	2.39
本文	**1.32**	**1.47**	**1.53**	**2.08**	**2.12**	**2.37**	**1.59**	**2.14**

注："—"表示配准结果误差过大不纳入计算。

进一步对比配准性能较好的 CFFMNet、SFDDFF-Net 和 CIRSM-Net 方法，可以看出，本文算法在多个实验设置中均展现出更高的时间效率，在第 1 组多时相实验中分别缩短了 0.85、0.77 和 0.34 s，在第 2 组多视角实验中分别缩短了 0.53、0.33 和 0.39 s，在多模态实验组中分别缩短了 0.9、1.17 和 1.4 s，在第 4 组纷杂情况实验中分别缩短了 0.86、1.04 和 0.92 s，其平均配准时间为 0.87 s。证明该方法显著提升了配准速度，满足了实际遥感应用中对高效性的要求。

随后，利用 PCK 指标对上述测试集的关键点匹配性能进行了评估，测试阈值 τ 设定为 0.05、0.03 和 0.01，最优结果加粗表示。评估结果如表 2 不同算法 PCK 值对比所示。结果表明，本文方法在阈值为 0.05 时的 PCK 值达到 0.969，远高于 SIFT，分别比 CFFMNet、SFDDFF-Net 和 CIRSM-Net 高出 3.2%、3.4%和 1.2%。结果表明，本文提出的算法不仅在配准精度上表现优异，且相比于其他方法，它在保持较高配准精度的同时，还能保持较快的执行速度。这使得该算法能够更好地适应实际应用场景中的需求，特别是在需要高效和精确图像配准的遥感图像处理任务中。

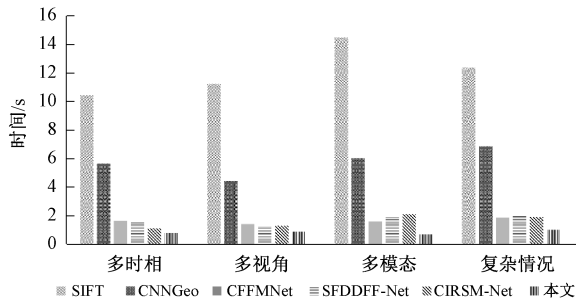

图 9　6 种算法配准时间对比图

Fig. 9　Time comparison chart of the six algorithms

表 2　6 种算法 PCK 值对比

Table 2　Comparison of PCK values of algorithms

方法	$\tau=0.05$	$\tau=0.03$	$\tau=0.01$
SIFT	0.306	0.227	0.036
CNNGeo	0.601	0.447	0.227
CFFMNet	0.937	0.817	0.363
SFDDFF-Net	0.935	0.831	0.372
CIRSM-Net	0.957	0.803	0.409
本文	**0.969**	**0.857**	**0.428**

2.5 消融实验

为了进一步验证本文所提出的创新点,进行了消融实验,创新点归纳为3部分:1)设计多尺度傅里叶模块(MSFM)模块嵌入到StarNet块中提高特征提取的性能;2)提出跨维门控注意力(CDGA),提升模型的特征表示能力;3)设计了基于部分分配矩阵的双向匹配层,进一步提升配准精度。

本文以CNNGeo为基准框架,逐一将上述3个创新点添加到框架中,并在9 000对Aerial Image数据集上进行训练,使用PCK(关键点正确估计比例)、MAE和RMSE指标在500组图像验证集上进行定量分析,评估各创新对配准性能的提升效果。

测试结果通过表3中的消融实验PCK值分析以及表4中的MAE和RMSE值分析呈现,实验结果表明,逐步添加创新点后,配准精度逐步提升,误差逐步减小,验证了每个创新点对配准性能的重要贡献。

表3 消融实验PCK值分析

Table 3 The results of PCK value analysis of ablation experiment

方法	PCK(τ=0.05)
CNNGeo	0.601
CNNGeo+(1)	0.754
CNNGeo+(2)	0.802
CNNGeo+(3)	0.727
CNNGeo+(1)+(2)	0.889
CNNGeo+(2)+(3)	0.854
CIRSM-Net	0.957
本文	**0.969**

表4 消融实验MAE和RMSE值分析

Table 4 The results of MAE and RMSE value analysis of ablation experiment

方法	MAE	RMSE
CNNGeo	4.62	5.38
CNNGeo+(1)	2.12	2.45
CNNGeo+(2)	2.19	2.63
CNNGeo+(3)	2.23	2.78
CNNGeo+(1)+(2)	1.79	1.93
CNNGeo+(2)+(3)	1.94	2.28
CIRSM-Net	1.82	2.16
本文	**1.46**	**1.68**

具体而言,在基准模型CNNGeo的基础上嵌入MFCM模块,使整个网络融入MFCM融合多尺度空间域与频域的双域特征,PCK值提升了15.3%,MAE和RMSE分别下降了2.50和2.93,证明了该创新点能够有效提升模型的特征提取能力并增强遥感图像配准的准确性。提出跨维门控注意力(CDGA)应用于CNNGeo,使网络能够高效捕获图像中的上下文信息,PCK值提升了20.1%,MAE和RMSE分别下降了2.43和2.75,说明该创新点对于遥感图像配准精度的重要性。在基准模型CNNGeo的特征匹配阶段设计双向匹配策略并引入部分分配矩阵来加权特征匹配点,从数据上进行分析,创新点(3)提升了遥感图像配准的性能,PCK值有明显提升。

3 结 论

针对遥感图像配准中复杂环境带来的特征提取困难、多尺度几何变形造成的配准精度受限等挑战,本文提出一种融合双域特征与跨维门控注意力的遥感图像配准模型,通过融合多尺度空间域与频域特征来增强模型特征表示能力,并设计跨维空间注意力高效聚合图像的上下文信息。此外,设计基于部分分配矩阵的双向匹配策略过滤冗余匹配点,以提升配准精度。实验结果表明,该算法与近年来提出的遥感图像配准算法相比可以提高复杂遥感图像对的图像配准精度和效率。未来打算将研究重心放在轻量级网络的设计上,注重遥感图像配准性能和速度的共同提升。

参考文献

[1] SOMMERVOLD O, GAZZEA M, ARGHANDEH R. A survey on SAR and optical satellite image registration[J]. Remote Sensing, 2023, 15(3): 850.

[2] MA J, JIANG X, FAN A, et al. Image matching from handcrafted to deep features: A survey [J]. International Journal of Computer Vision, 2021, 129(1): 23-79.

[3] ZHANG W, ZHAO Y. An improved SIFT algorithm for registration between SAR and optical images[J]. Scientific reports, 2023, 13(1): 6346.

[4] PARK J H, NAM W J, LEE S W. A two-stream symmetric network with bidirectional ensemble for aerial image matching[J]. Remote Sensing, 2020, 12(3): 465.

[5] 陈颖,张祺,李文举,等.参数合成空间变换网络的遥感图像一致性配准[J].中国图象图形学报,2021,26(12):2964-2980.
CHEN Y, ZHANG Q, LI W J, et al. Consistent registration of remote sensing images in parametric synthesis' spatial transformation network[J]. Journal of Image and Graphics, 2021,26(12): 2964-2980.

[6] TANG T, CHEN T, ZHU B, et al. MU-NET: A multiscale unsupervised network for remote sensing image registration[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial

Information Sciences, 2022, 43: 537-544.

[7] YAO Z S, FAN G D, FAN J F, et al. Spatial-frequency dual-domain feature fusion network for low-light remote sensing image enhancement[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024(62): 1-16.

[8] 唐海林,张俊,李屹旭,等.基于注意力改进 RTformer 的滑坡遥感图像语义分割[J].电子测量技术,2024,47(21):150-158.
TANG H L, ZHANG J, LI Y X, et al. Semantic segmentation of landslide remote sensing image based on improved attention RTformer [J]. Electronic Measurement Technology, 2024, 47(21):150-158.

[9] GENG J, SONG S A, JIANG W. Dual-path feature aware network for remote sensing image semantic segmentation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 34 (5): 3674-3686.

[10] LIANG C B, DONG Y Y, ZHAO C J, et al. A coarse-to-fine feature match network using transformers for remote sensing image registration[J]. Remote Sensing, 2023, 15(13):3243.

[11] MA X, DAI X, BAI Y, et al. Rewrite the stars[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 5694-5703.

[12] ROCCO I, ARANDJELOVIC R, SIVIC J. Convolutional neural network architecture for geometric matching [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6148-6157.

[13] 倪力政,陈颖,李翔,等.融合注意力与多尺度特征的遥感图像配准[J].计算机工程与应用,2025,61(3):275-285.
NI L ZH, CHEN Y, LI X, et al. Remote sensing image registration integrating attention and multi-scale features[J]. Computer Engineering and Applications, 2025,61(3):275-285.

[14] 闫祎巧,王宏生,赵怀慈,等.融合多尺度卷积和注意力机制的场景提取方法[J].电子测量技术,2023,46(16):172-178.
YAN Y Q, WANG H SH, ZHAO H C, et al. Scene extraction methods incorporating multi-scale convolution and attention mechanisms[J]. Electronic Measurement Technology, 2023,46(16):172-178.

[15] KONG L S, DONG J X, GE J J, et al. Efficient frequency domain-based transformers for high-quality image deblurring [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 5886-5895.

[16] ZHOU M, HUANG J, GUO C L, et al. Fourmer: An efficient global modeling paradigm for image restoration[C]. International Conference on Machine Learning. PMLR, 2023: 42589-42601.

[17] 王伟,陈颖,王嘉浩,等.基于注意力和特征融合的光学遥感图像配准[J].激光杂志,2023,44(5): 174-181.
WANG W, CHEN Y, WANG J H, et al. Optical remote sensing image registration based on attention and feature fusion[J]. Laser Journal, 2023,44(5): 174-181.

[18] LIU Y K, KANG X D, HUANG Y W, et al. Unsupervised domain adaptation semantic segmentation for remote-sensing images via covariance attention[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.

[19] LINDENBEGER P, SARLIN P-E, POLLEFYES M. LightGlue: Local feature matching at light speed[C]. IEEE/CVF International Conference on Computer Vision, 2023: 17627-17638.

[20] GONG X Y, LIU Y Y, YANG Y, Robust stepwise correspondence refinement for low-altitude remote sensing image registration[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(10): 1736-1740.

[21] YAO Y X, ZHANG Y J, WAN Y, et al. Multi-modal remote sensing image matching considering co-occurrence filter[J]. IEEE Transactions on Image Processing, 2022, 31: 2584-2597.

[22] WANG P, LIU Y, LIANG X, et al. CIRSM-Net: A cyclic registration network for SAR and optical images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-19.

作者简介

魏庭旭,硕士,主要研究方向为图像处理。
E-mail:weitingxu0705@163.com

陈颖(通信作者),博士,教授,主要研究方向为人工智能、图像处理。
E-mail:chy@sit.edu.cn

李铖昊,本科,主要研究方向为图像处理。
E-mail:1608447067@qq.com

马文浩,硕士,主要研究方向为图像处理。
E-mail:1390637807@qq.com