

基于改进SuperPoint的空天异源图像匹配算法

刘庚辰^{1,2}, 姜 梁³, 吴国强^{1,2,4}, 黄 坤^{1,2}

(1. 航天时代飞鸿技术有限公司, 北京 100094; 2. 中国航天科技集团有限公司智能无人系统总体技术研发中心, 北京 100094;
3. 中国航天科技集团有限公司第九研究院, 北京 100094; 4. 北京理工大学机电学院, 北京 100081)

摘 要: 空天异源图像特征提取难度较大, 图像匹配精度较低, 给无人机精确目标定位带来了不利影响. SuperPoint-SuperGlue算法由于其自监督, 易训练, 精度高等特性, 近年来在图像匹配领域应用较为广泛, 但针对空天异源图像匹配领域, SuperPoint特征提取能力仍有待提高. 为提高空天异源图像匹配精度, 本文提出基于改进SuperPoint的空天异源图像匹配算法. 首先, 将群智能增强模块(Spatial Group-wise Enhance, SGE)与全局注意力机制(Global Attention Mechanism, GAM)引入到SuperPoint编码器中构成补充编码器, 一定程度上解决了图像特征分布不均匀以及弱纹理图像特征提取较难的问题; 其次, 为进一步增强算法特征提取能力, 将补充编码器与原SuperPoint编码器进行并联构成组合编码器, 结合二者优势, 提取差异性更大的图像特征, 减少对相似区域的特征点误匹配, 提高空天异源图像的匹配精度; 最后, 通过实验验证, 在UAV-VisLoc数据集上80像素误差区间以内可匹配图像数量可达82.14%, 与原SuperPoint算法相比, 80像素误差区间以内可匹配图像数量提高了6.05%, 与其他先进算法相比, 在各像素误差区间以内可匹配图像数量均有提高. 实验表明, 本文提出的算法可以有效地解决空天异源图像匹配中特征提取能力较弱, 特征分布不均匀等问题.

关键词: 空天异源; 图像匹配; 无人机; SuperPoint; SuperGlue; 编码器

中图分类号: TN91

文献标识码: A

文章编号: 0372-2112(2025)04-1201-11

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240724

Aerospace Heterogeneous Image Matching Algorithm Based on Improved SuperPoint

LIU Geng-chen^{1,2}, JIANG Liang³, WU Guo-qiang^{1,2,4}, HUANG Kun^{1,2}

(1. Aerospace Times Feihong Technology Limited Company, Beijing 100094, China;

2. China Aerospace Science and Technology Corporation Limited Intelligent Unmanned System Overall Technology Research and Development Center, Beijing 100094, China;

3. China Aerospace Science and Technology Corporation 9th Institute, Beijing 100094, China;

4. School of Mechanical and Electrical Engineering, Beijing Institute of Technology, Beijing 100081, China)

Abstract: It is quite difficult to extract features from heterogeneous aerospace images, and the image matching accuracy is relatively low, which has a negative impact on the precise target positioning of unmanned aerial vehicles (UAVs). The SuperPoint-SuperGlue algorithm has been widely applied in the field of image matching in recent years due to its characteristics such as self-supervision, easy training, and high accuracy. However, in the field of heterogeneous aerospace image matching, the feature extraction ability of SuperPoint still needs to be improved. In order to improve the matching accuracy of heterogeneous aerospace images, this paper proposes a heterogeneous aerospace image matching algorithm based on the improved SuperPoint. Firstly, the spatial group-wise enhance (SGE) module and the global attention mechanism (GAM) are introduced into the SuperPoint encoder to form a supplementary encoder, which to a certain extent solves the problems of uneven distribution of image features and the difficulty in extracting features from weakly textured images. Secondly, to further enhance the feature extraction ability of the algorithm, the supplementary encoder is connected in parallel with the original SuperPoint encoder to form a combined encoder. By combining the advantages of the two, it can extract image features with greater differences, reduce the false matching of feature points in similar regions, and improve the matching accuracy of heterogeneous aerospace images. Finally, through experimental verification, within the error range of 80 pixels on

the UAV-VisLoc dataset, the number of matchable images can reach 82.14%. Compared with the original SuperPoint algorithm, the number of matchable images within the error range of 80 pixels has increased by 6.05%. Compared with other advanced algorithms, the number of matchable images within each pixel error range has increased. The experiments show that the algorithm proposed in this paper can effectively solve the problems such as weak feature extraction ability and uneven feature distribution in the matching of heterogeneous aerospace images.

Key words: heterogeneous in aerospace domain; image matching; UAV; SuperPoint; SuperGlue; encoder

1 引言

无人机可以快速部署,具备对时敏目标长时跟踪定位能力,但由于无人机侦察图像目标定位主要依靠姿态、载荷信息进行解算,在卫星导航拒止或传感器精度不足时,定位误差较大,难以满足实际应用需求.空天异源图像匹配技术能够快速计算无人机侦察图像中目标位置,有效提高卫星导航拒止或传感器精度不足情况下的无人机目标定位精度.空天异源图像匹配技术具体流程如图1所示.

无人机侦察图像与高精度天基卫星影像经特征提取后得到图像特征点,利用特征匹配算法进行特征点匹配,根据匹配结果计算图像变换单应性矩阵,获取空天异源图像之间的映射关系,结合卫星图片地理信息与无人机侦察图像信息,映射出无人机拍摄图像中待匹配目标的具体坐标.尽管图像特征点匹配作为计算机视觉的核心问题已具有数十年的研究历史,但由于空天异源图像域差异较大、图像畸变较大,手工设计的特征描述子在这些情况下效果较差^[1].

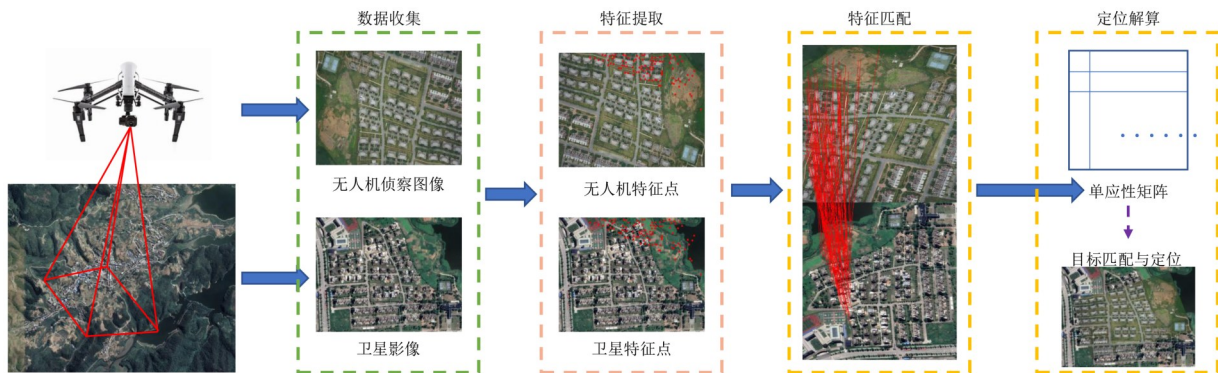


图1 空天异源图像匹配示意图

近年来,基于深度学习的空天异源图像匹配在提升匹配质量和泛化性方面展现出强大的能力,但其在匹配质量上仍有较大的提升空间.刘畅等人^[2]采用知识蒸馏技术,提升了推理效率,降低了匹配时间;任艳等人^[3]采用模糊信息粒技术,对提取到的图像边缘进行相似性匹配,得到无人机航拍图像与卫星图像的匹配位置.然而,这些方面仍然面临着一些亟待解决的问题:(1)由于卷积网络感受野的局限性,无法充分利用全局性息,导致其对于农田、草地等大面积弱纹理区域的匹配效果较差;(2)卷积具有平移不变性,遇到相同的局部特征,就会产生相似的局部响应,无人机拍摄图像与卫星图像均具有较广的视野范围,图像相似区域较多,会产生较多的误匹配情况,相似目标特征差异仍待提高.

为解决以上问题,本文提出基于改进SuperPoint的空天异源图像匹配算法,本算法贡献如下:

(1)本文提出一种基于改进SuperPoint的空天异源图像匹配方法,通过融合SuperPoint编码器^[4]与SGE模

块^[5]、GAM模块^[6]构成补充编码器,有效提升了算法对于弱纹理区域的特征提取效果.

(2)通过将补充编码器与原编码器并联构成组合编码器,结合二者优势,提取差异性更大的图像特征,减少相似区域特征点误匹配,提高空天异源图像匹配精度.

(3)采用UAV-VisLoc数据集^[7]进行实验,在定位精度方面,80像素误差区间以内可匹配图像占比可达到82.14%,与其他先进算法相比,本文算法在匹配精度方面展现出更卓越的性能.

2 相关工作

2.1 图像特征匹配

SuperPoint-SuperGlue^[8]是近年来深度学习中图像匹配的代表算法,SuperPoint负责图像关键点检测并提取关键点描述子,SuperGlue给定图片的多个描述子进行匹配,已在多个实际应用场景取得较好的效果.相较于其他图像匹配算法,SuperPoint是一种自监督学习算

法,减少了对大量标注数据的依赖,同时还具备较高的训练效率与推理效率,在多个图像匹配场景已经取得了较好的应用效果,但是现有算法针对多角度、多时相、多场景的无人机侦察图像与卫星影像的匹配,仍有一定局限.

图像特征点由特征点位置信息与描述子(即特征点周围信息)两部分构成,SuperPoint是一种自监督的特征点检测与描述网络,相较于其他特征提取算法,能够同时提取特征点位置以及描述子,且具有较高精度,已被广泛地应用于图像匹配领域,网络结构如图2所示.

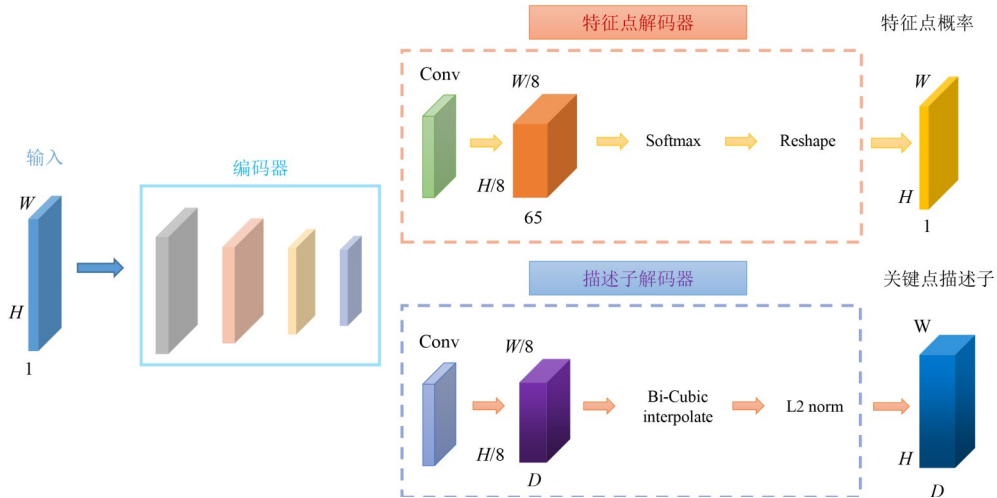


图2 SuperPoint网络结构

SuperPoint由编码器与解码器两部分构成. 编码器为解码器共享网络,输入为 $H \times W$ 的图像,经过4个block块得到编码器输出,生成深度特征图. 前三个block块包括2个卷积层和最大池化层,第四个只包含2个卷积层. 解码器分为特征点解码器和描述子解码器两部分. 特征点解码器输入为解码器输出,通过卷积、Softmax与Reshape操作得到每个像素是特征点的概率. 描述子解码器输入为解码器输出,经过卷积、双

线性插值与L2归一化操作输出关键点描述子,结合特征点解码器与描述子解码器的输出即可得到图像特征点.

SuperGlue是一种特征匹配算法,不同于传统方法计算特征点之间的距离来衡量匹配的概率,SuperGlue将匹配问题看作一个最优运输问题,通过解决最优运输问题得到匹配点最终分布矩阵. SuperGlue流程图如图3所示.

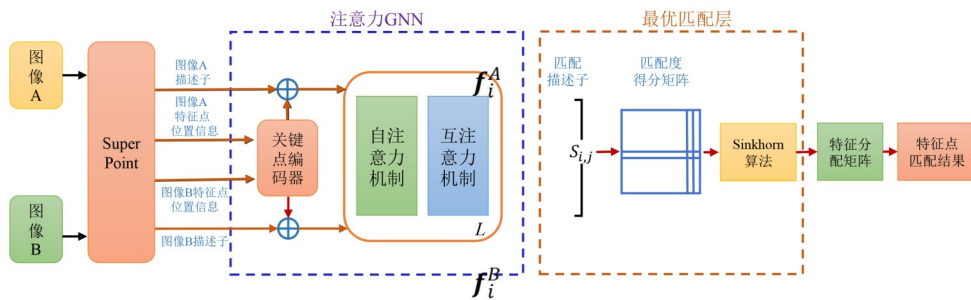


图3 SuperGlue匹配流程

SuperGlue主要由注意力GNN和最优匹配层2个部分组成. 注意力GNN输入为SuperPoint输出的特征点位置信息及描述子,使用关键点编码器将其编码成向量,然后重复使用 L 次自注意力机制与互注意力机制得到特征点的匹配描述子. 在最优匹配层中,通过计算匹配描述子内积得到匹配度得分矩阵,经由Sinkhorn算法得出特征分配矩阵,最后根据特征分配矩阵给出特征点匹配结果.

2.2 空天异源图像特征匹配

无人机侦察图像与卫星影像往往会由于侦察时间、高度、角度不同等因素导致每英寸像素点数(Dots Per Inch, DPI)不同、图像畸变、地貌差异等情况,为特征提取及配准带来较大难度. 空天异源图像匹配作为新兴领域,在多个领域都有较为广泛的应用,也可与其他先进方法结合使用^[9-12]. 国内外研究成果主要分为传统空天异源图像方法和基于深度学习空天异源图像

方法.

传统方法多利用点特征与线特征进行空天异源图像匹配. 赵克龙^[13]对ORB^[14]算法进行了改进,提高了匹配准确率与鲁棒性. 温育杜等人^[15]提出一种可以提高尺度差异较大的航拍图像与卫星影像的匹配准确率的算法,在一定程度上减小了目标尺度差异给匹配精度带来的不利影响. 罗卿莉等人^[16]提出了一种无人机与卫星影像匹配的相似性逼近算法,提高了无人机图像与卫星影像的尺度相似性.

传统方法可以在一定程度上满足空天异源图像匹配的要求,但无法很好应对视角畸变、特征模糊等不利情况,特征提取效果不理想. 基于深度学习的图像匹配方法相较于传统方法而言,具备优秀的图像特征提取能力以及较好的鲁棒性与场景泛用性,已被广泛应用. 基于深度学习特征匹配算法可以分为直接从原始图像中生成对应关系算法与匹配稀疏关键点来建立跨图像对应关系算法. 直接从原始图像中生成对应关系算法以LoFTR^[17]及其改进算法^[18,19]为代表,LoFTR较高的匹配准确度与较快的运算速度,但匹配效率较低,Efficient-LoFTR在LoFTR基础上采用了自适应令牌选择的聚合注意力机制以及两级相关层提高了效率和精度,Aspanformer在LoFTR基础上改善了注意力模块,能够自适应地调整注意力广度. LoFTR及其改进算法在空天异源图像匹配中得到广泛应用. 匹配稀疏关键点来建立跨图像对应关系算法以SuperPoint-SuperGlue、SuperPoint-LightGlue^[20]等算法为代表,相较于LoFTR及其改进算法,SuperPoint-SuperGlue为特征提取-特征匹配两阶段图像匹配方法,便于对特征提取、特征匹配结果进行实验改进,同时也具有较好的匹配精度与效率. 学者不只着眼于提升空天异源图像匹配精度,也从匹配速度,匹配场景泛用性等角度对空天异源图像匹配算法进行改善. 例如陆海刚等人^[21]以ResNet提取稳定特征,以最小欧式距离结合小型Transformer实现精细

匹配,提高了算法实时性. 任艳等人^[3]基于模糊信息粒提出了一种不同季节下无人机航拍图像与卫星图像匹配方法,该方法有效解决了四季变化、光照差异条件下异源图像匹配困难问题. 黄振家^[22]提出了一种针对不同季节无人机航拍图像与卫星影像模板匹配方法,有效应对由于季节变化导致地图覆盖变化的情况.

基于深度学习的空天异源图像匹配虽已取得一定成就,但对于空天异源图像的相似区域与弱纹理区域较多的特点考虑较少,存在相似区域与弱纹理区域特征提取能力较弱的问题,其匹配精度仍然有待提高.

3 改进的SuperPoint模型

SuperPoint-SuperGlue算法虽已在多个实际场景中得到广泛应用,但在空天异源图像匹配领域其匹配精度仍可提高. 制约SuperPoint-SuperGlue在空天异源图像匹配领域精度提高的重要因素之一为空天异源图像存在差异性,图像特征不准确,给后续特征匹配等任务带来了难度.

为减小空天异源图像差异给匹配带来的不利影响,增强SuperPoint的特征提取能力,本文对SuperPoint编码器进行改进,提出基于改进SuperPoint的空天异源图像匹配算法,算法结构如图4所示,由两部分组成:SuperPoint网络提取图像特征点,SuperGlue算法进行特征点匹配. 算法流程如下:SuperPoint编码器通过卷积与池化对待匹配图像对(无人机侦察图像与卫星图像)进行特征提取与降维,生成深度特征图;SuperPoint解码器对深度特征图进行解码,通过特征点解码器与描述子解码器分别输出特征点及其描述子;SuperGlue算法接受SuperPoint网络的输出,通过注意力GNN及最优匹配层得到匹配度得分矩阵,最后通过Sinkhorn算法解算出最优特征分配矩阵. 本文改进了SuperPoint编码器,融合了空间组增强模块、全局注意力机制模块,提出了补充编码器与组合编码器,使其深度特征图中的

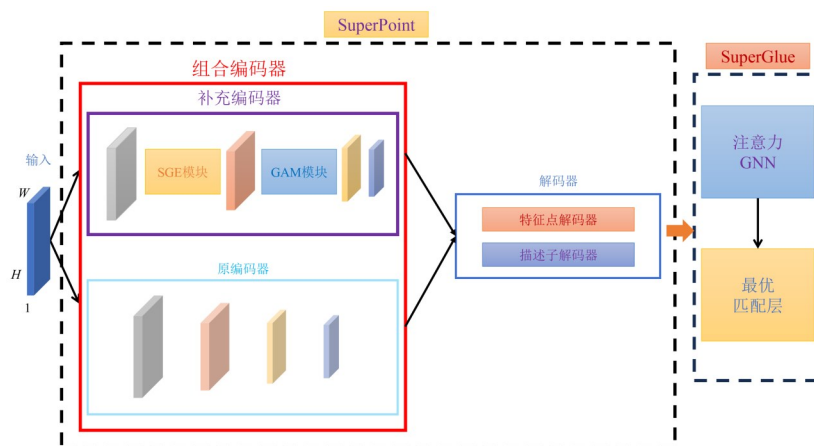


图4 基于改进SuperPoint的空天异源图像匹配算法

特征信息更为全面. 补充编码器在 SuperPoint 编码器的基础上结合 Li 等^[5]提出的 SGE 模块以及 Liu 等^[6]提出的 GAM 模块削弱图像特征由于卷积操作带来的损失, 增加了图像特征信息. 通过将补充编码器与 SuperPoint 原编码器并联构成组合编码器, 提高相似区域特征区分度, 增强弱纹理特征提取能力.

3.1 组合编码器

原始的 SuperPoint 算法主要关注图像中纹理较强的区域, 在通用的图像特征匹配任务中表现较好, 但由于空天异源图像具有相似区域较多, 弱纹理场景出现频繁的特点, 直接使用 SuperPoint 作为特征描述子会造成相似区域特征接近, 弱纹理区域特征提取不足的问题, 从而导致图像匹配精度下降. 针对上述问题, 本文提出了补充编码器与组合编码器. 补充编码器通过融合 SGE 模块与 GAM 模块加强了弱纹理区域特征提取能力; 组合编码器通过并联补充编码器与原编码器, 减少对相似区域的特征点误匹配, 提高了空天异源图像的匹配精度.

本文将补充编码器与原 SuperPoint 编码器并联互补, 将图像同时输入到补充编码器与原编码器中, 将二者最后一个 block 输出的特征点进行解码, 进一步进行

匹配, 利用并联操作能够结合二者优势, 提高网络特征提取能力, 达到在多场景中正确特征匹配点增多, 匹配效果更精确的目的. 基于组合编码器的 SuperPoint 网络结构如图 5 所示.

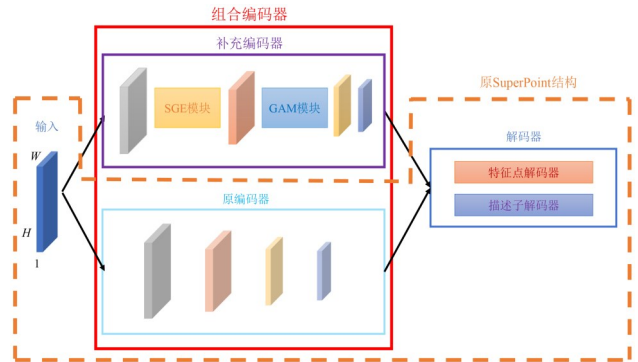


图 5 基于组合编码器的 SuperPoint 网络结构

3.1.1 原编码器

如图 6 所示, SuperPoint 编码器主干网络为 VGG 网络^[23], 作为解码器的共享网络, 由 4 个 block 组成. 前 3 个 block 均包括 2 个卷积层和最大池化层, 最后一个 block 只包括 2 个卷积层. 输入图像经 SuperPoint 编码器处理, 输出深度特征图, 具体结构如图 6 所示.

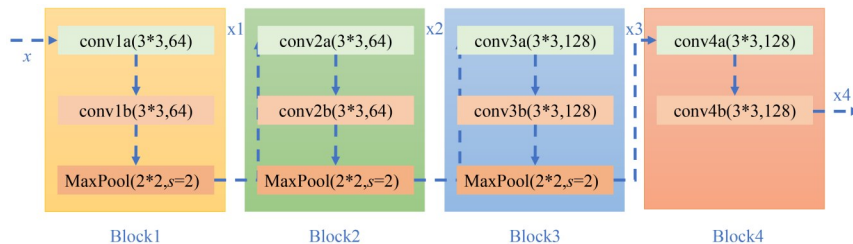


图 6 编码器详细结构

3.1.2 补充编码器

为提高 SuperPoint 对图像弱纹理区域的特征提取能力, 基于 SGE 模块与 GAM 模块原理, 在 SuperPoint 编码器中加入 SGE 模块与 GAM 模块, 补充编码器模块结构如图 7 所示.

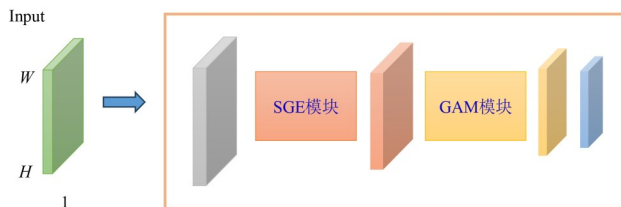


图 7 补充编码器结构

原 SuperPoint 编码器包括四个 block, 图像经每一个 block 处理后均会生成一个特征图, 随着深度加深, 特征图会损失较多图像底层特征. 为提升算法对弱纹理区

域的敏感程度, SGE 输入应尽可能包含较丰富的图像特征, 本文将 SGE 模块放到 SuperPoint 第一个 block 后进行改进, 加强 SuperPoint 对弱纹理区域特征提取能力.

为加强弱纹理区域特征差异性, 补充编码器先引入 SGE 模块处理图像. 经 SGE 模块处理过后的特征图拥有更丰富的特征, 但也会带来一定的特征冗余的问题. 经 SuperPoint 编码器第二个 block 处理后的特征图冗余特征较少, 且保留了较多的图像特征信息. 为了 GAM 模块能够有效融合特征并且减少运算资源的消耗, 本文将 GAM 模块放置在第二个 block 后, 增加了目标特征中的全局信息.

无人机侦察图像范围广, 侦察图像纹理分布不均, 原始 SuperPoint 编码器对整张图像进行特征提取时, 弱纹理区域特征信息由于卷积操作丢失较多, 导致 SuperPoint 对弱纹理区域特征提取精度不足.

为提高 SuperPoint 特征提取能力,突出卷积在正确区域的语义特征,本文引入 Li 等人提出的 SGE 模块^[5]. SGE 模块首先将特征分组,假设每个组都可学习到某一个特定语义,为了利用某个组空间的总体信息来进一步增强语义学习,对于关键区域中的特征,可以使用空间平均函数来近似该组所学的语义向量作为全局特征,如式(1)所示:

$$\mathbf{g} = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i \quad (1)$$

其中, \mathbf{g} 代表某个组的全局特征, \mathbf{x}_i 代表某组特征, m 为特征个数.

再利用全局特征,为每个位置特征生成相应的权重系数,该系数通过简单的点积获得,衡量了全局语义特征 \mathbf{g} 与局部特征 \mathbf{x}_i 之间的相似度,为每个组中的每个空间位置生成一个注意力因子,注意力因子仅由组内全局和局部特征描述符之间的相似性来决定,全局特征由式(2)得出:

$$c_i = \mathbf{g} * \mathbf{x}_i \quad (2)$$

局部特征 \mathbf{x}_i 与全局特征 \mathbf{g} 越相似,则 c_i 越大,通过调整该因子即可调整组内每个子特征的重要性,为防

止不同样本之间系数出现偏差,在空间上由式(3)~式(5)对 c_i 进行归一化处理:

$$\hat{c}_i = \frac{c_i - \mu_c}{\sigma_c + \varepsilon} \quad (3)$$

$$\mu_c = \frac{1}{m} \sum_j c_j \quad (4)$$

$$\sigma_c^2 = \frac{1}{m} \sum_j (c_j - \mu_c)^2 \quad (5)$$

为使归一化操作可被还原,为每个组学习两个缩放偏移参数,经缩放平移后可得:

$$a_i = \gamma \hat{c}_i + \beta \quad (6)$$

最后通过空间上的一个 Sigmoid 门与原始特征向量 \mathbf{x}_i 相乘得到增强后的特征向量:

$$\hat{\mathbf{x}}_i = \mathbf{x}_i * \sigma(a_i) \quad (7)$$

无人机远距离侦察下,地物目标特征不明显,传统的卷积只利用局部信息来计算目标像素而往往忽略全局信息,全局信息缺失可能会给卷积后续操作带来一些偏差.目标特征中的全局信息增多可以减少相似目标误匹配,增加匹配准确率,本文引入 Liu 等人提出的 GAM 模块^[6],GAM 模块网络结构如图 8 所示.

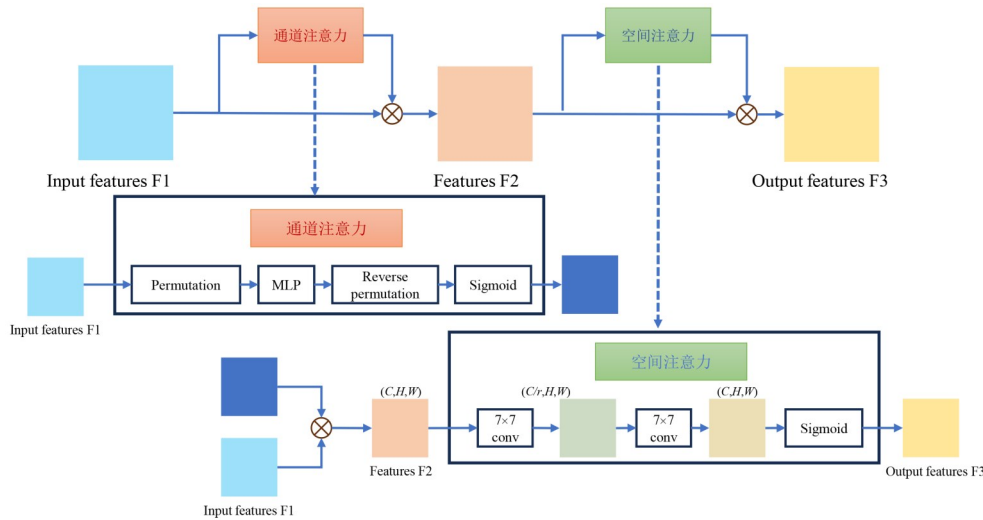


图 8 GAM 模块

GAM 模块采用 CBAM^[24] 的通道-空间注意机制, CBAM 是一种针对卷积神经网络设计的注意力模块. CBAM 机制针对传统卷积操作的不足,提出了通道注意力机制与空间注意力机制, CBAM 结构如图 9 所示.

通道注意力机制通过平均池化和最大池化操作生成两个描述符,然后通过共享的多层感知机生成通道注意力图,如式(8)所示:

$$M_c(\mathbf{F}) = \sigma(\text{MLP}(\text{AvgPool}(\mathbf{F})) + \text{MLP}(\text{MaxPool}(\mathbf{F}))) \quad (8)$$

将其与输入图像相乘得到 \mathbf{F}' , 如式(9)所示:

$$\mathbf{F}' = M_c(\mathbf{F}) \otimes \mathbf{F} \quad (9)$$

空间注意力模块将 \mathbf{F}' 作为输入,通过平均池化和最大池化操作生成两个 2D 特征图,然后通过卷积层生成空间注意力图,公式如式(10)所示:

$$M_s(\mathbf{F}') = \sigma(\text{conv}([\text{AvgPool}(\mathbf{F}'); \text{MaxPool}(\mathbf{F}')])) \quad (10)$$

最后通过空间注意力图与 \mathbf{F}' 相乘得到图像特征:

$$\mathbf{F}'' = M_s(\mathbf{F}') \otimes \mathbf{F}' \quad (11)$$

CBAM 相较于传统卷积操作,从通道和空间两个方面对图像特征进行了处理,目标特征来源不局限于局部信息,从而更好地表征了图像特征.

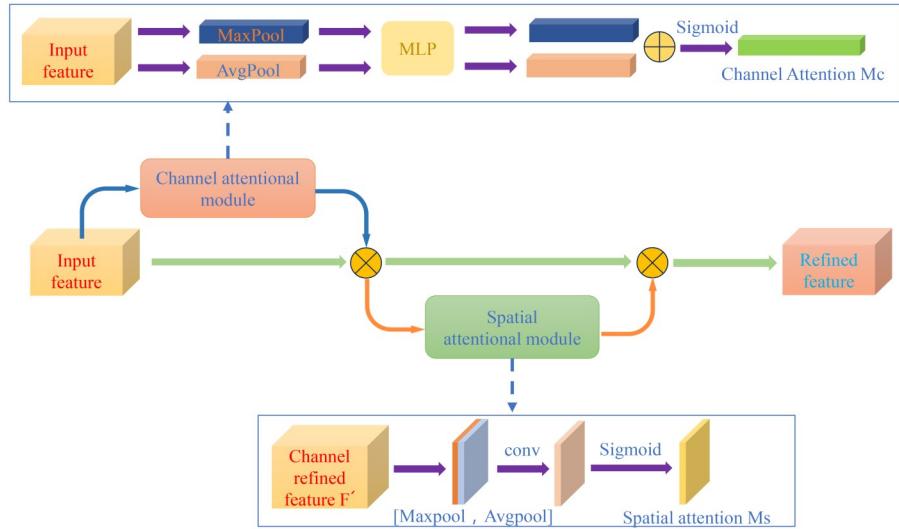


图9 CBAM模块

GAM 采用了 CBAM 的思想,对通道注意力机制与空间注意力机制进行了重构,与原 CBAM 机制相比, GAM 增加了通道与空间的相互作用,在三个维度上捕捉重要特征,目标特征包含的信息更多. 通道注意力机制使用三维排列保留三维信息,再利用多层感知器的学习能力加强跨维度之间依赖关系. 空间注意力机制对通道注意力机制处理后的特征图进行 2 个 7×7 大小卷积操作进行信息融合.

3.2 解码器

SuperPoint 解码器对组合编码器输出进行解码,分为特征点解码器和描述子解码器两部分.

特征点解码器输入为组合解码器输出的深度特征图 \mathbf{x} ,通过卷积得到与原尺寸范围一一对应的张量,再进行 Softmax 与 Reshape 操作得到每个像素是特征点的概率. 具体地,使用卷积层对深度特征图 $H/8 \times W/8 \times 128$ 进行卷积,获得 $H/8 \times W/8 \times 65$ 的张量,再使用 Softmax 进行筛选,最后通过 Reshape 获得 $H \times W \times 1$ 的输出.

$$a = \text{Softmax} \left(\text{cov}_{\text{pb}} \left(\text{relu} \left(\text{cov}_{\text{pa}} (\mathbf{x}) \right) \right) \right) \quad (12)$$

描述子解码器输入为组合解码器输出的深度特征图 \mathbf{x} ,首先求得一个半稠密的描述子,再通过多项式插值得到描述,最后通过归一化统一描述子长度描述. 具体地,深度特征图经过两层卷积得到 $H/8 \times W/8 \times D$ 的张量,经双线性插值扩大张量尺寸为 $H \times W \times 256$,最后对每一个像素的描述子进行 L2 归一化. 可表示为

$$b = \text{L2} \left(\text{cov}_{\text{Db}} \left(\text{relu} \left(\text{cov}_{\text{Da}} (\mathbf{x}) \right) \right) \right) \quad (13)$$

结合特征点解码器输出 a 与描述子解码器输出 b 即可得到图像特征点.

3.3 特征匹配

特征匹配算法采用 SuperGlue 算法, SuperGlue 由注意力 GNN 及最优匹配层两部分组成,结构如图 3 所示.

算法流程如下:通过关键点编码器对两张图片的关键点位置及视觉描述符进行编码:

$${}^{(0)}\mathbf{x}_i = d_i + \text{MLP}_{\text{enc}}(p_i) \quad (14)$$

其中, d_i 代表描述子, p_i 代表关键点位置信息, MLP 代表多层感知机,接着将编码好的输入到注意力聚合模块,注意力聚合模块包括自注意力机制与互注意力机制,迭代 L 次可以得出最终匹配描述子:

$$\mathbf{f}_i^A = W \cdot {}^{(L)}\mathbf{x}_i^A + b \quad (15)$$

$$\mathbf{f}_i^B = W \cdot {}^{(L)}\mathbf{x}_i^B + b \quad (16)$$

其中, \mathbf{f}_i^A 和 \mathbf{f}_i^B 代表匹配描述子,计算匹配描述子的内积 $S_{i,j}$:

$$S_{i,j} = \langle \mathbf{f}_i^A, \mathbf{f}_j^B \rangle \quad (17)$$

再将内积经由 Sinkhorn 算法得出特征分配矩阵,最后根据特征分配矩阵给出特征点匹配结果.

4 实验与分析

本节在公开数据集上验证本文方法的有效性. 首先介绍数据集、评价标准和实现细节;然后进行消融实验研究,最后介绍与主流方法在匹配准确率方面的对比实验.

4.1 数据集和评价标准

本文采用 UAV-VisLoc 数据集对算法匹配精度进行测试,该数据集由多个子数据集构成,每个子数据集包含了无人机侦察图像以及对应地区的卫星影像. 无人机侦察图像涵盖了农村、城镇、农场、城市、河流、丘陵等多种场景. 数据集格式如表 1 所示:

由于卫星影像包含实际地图范围较广,远大于无人机侦察图像包含地图大小,在实际工程应用中为提升匹配效率,一般根据无人机侦察范围,对卫星影像进行预处理,由于已知无人机图像中心点经纬度,根据以

表1 UAV-VisLoc 数据集格式

子数据集名称	无人机图像 像素大小	卫星影像 像素大小
Changjiang-20	3 976×2 652	9 774×26 762
Changjiang-23	3 976×2 652	11 482×34 291
Taizhou-1	3 976×2 652	35 092×24 308
Taizhou-6	3 976×2 652	18 093×38 408
Yunnan	3 000×2 000	9 394×6 144
Zhuxi	3 976×2 652	8 082×9 780
Donghuayuan	3 000×2 000	3 000×170
Huzhou-3	3 976×2 652	43 421×16 294
Huzhou-6	3 976×2 652	44 800×33 280
Huailai	3 000×2 000	6 593×5 077
Shandan	3 976×2 652	29 592×16 582

上信息,本文对数据集中图像进行旋转、裁剪等操作,使卫星子图像中心点像素经纬度与无人机图像中心点像素经纬度保持一致,使卫星图像范围与无人机侦察图像范围大致相同。

本文使用图像中心像素误差评估空天异源图像匹配的效果.计算过程如下:将无人机侦察图像与卫星图像进行匹配,根据匹配结果可得出从无人机图像变换到卫星图像的单应性矩阵.由于无人机图像中心点像素经纬度与卫星子图像中心点像素经纬度保持一致,根据单应性矩阵变换,无人机中心点像素经单应性变换对应卫星图像中心点像素,根据中心点像素误差可以进行匹配精度判别.具体地,使用中心像素误差小于 k 的图像比例评价算法性能,我们设置 $k=80$.

4.2 实现细节

本文模型使用 PyTorch 深度学习框架实现,使用 AID Data Set 等多个遥感数据集^[25-30]进行训练,划分训练集与测试集,最大学习率为 0.000 1,设置批次大小为 8,训练批次为 200 000,取测试集误差最小的参数为模型参数.用 UAV-VisLoc 数据集验证算法匹配精度。

4.3 消融实验

将本文提出的基于改进 SuperPoint 的空天异源图像匹配算法(即组合编码器+SuperGlue),与 SGE 模块+SuperPoint+SuperGlue, GAM 模块+SuperPoint+SuperGlue,补充编码器+SuperGlue,原 SuperPoint+SuperGlue 对比,图片大小统一为 640×480 像素,各方法对比结果如表 2 所示.相较于原 SuperPoint,SGE 模块+SuperPoint、GAM 模块+SuperPoint 80 像素误差区间内图片数量有所提高,但单独采用补充编码器 80 像素误差区间内图片数量有所降低,补充编码器在有的场景表现虽然优于原编码器,但在另一些场景表现劣于原编码器,组合编码器相较于补充编码器与原编码器而言,效果更为理想,如图 10 所示。

表2 各方法图片数量占数据集总数百分比

算法名称	80 像素误差 区间内
原 SuperPoint+SuperGlue	76.09
SGE 模块+SuperPoint+SuperGlue	79.16
GAM 模块+SuperPoint+SuperGlue	81.81
补充编码器+SuperGlue	75.01
组合编码器(补充编码器+SuperPoint)+SuperGlue	82.14

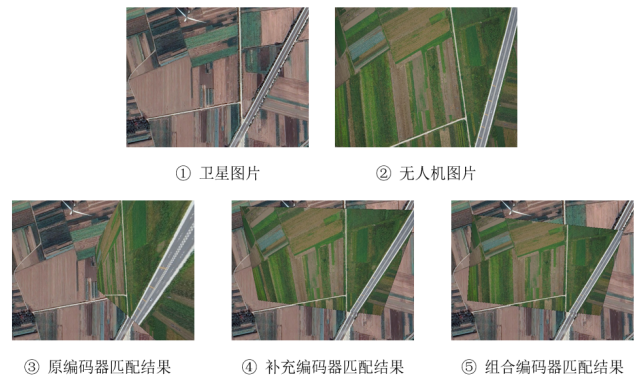


图10 匹配结果对比

4.4 特征可视化

本节对特征进行可视化,使用补充编码器与原编码器处理图像,其特征可视化结果如图 11 所示。

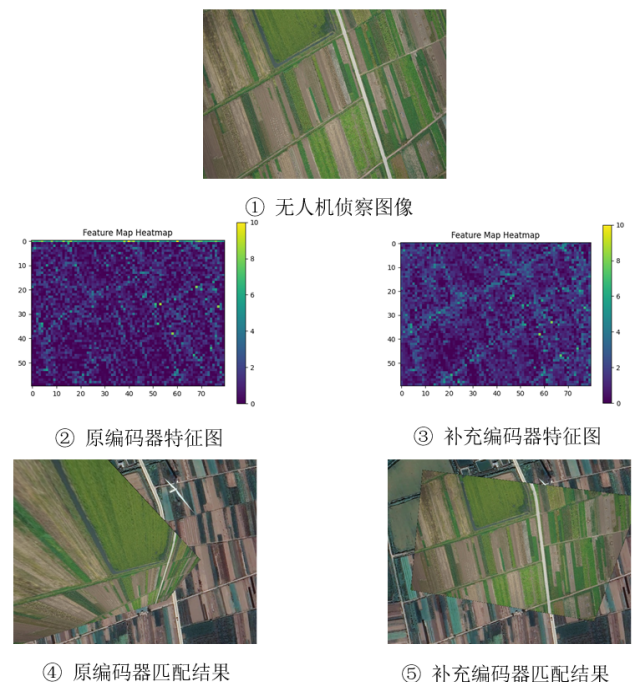


图11 特征可视化

原编码器特征图如图 11②所示,补充编码器特征图如图 11③所示,从图 11②与图 11③可以看出,经过补充编码器处理,在颜色相近的区域,相较于原编码器,可检

测出较不明显的纹理. 对特征点可视化,图 11④展示了原编码器匹配结果,图 11⑤展示了补充编码器匹配结果,相较于原编码器,补充编码器取得了较好的匹配结果.

4.5 对比实验

本文算法与原 SuperPoint-SuperGlue 算法各误差区间图片数量对比如图 12 所示,本文算法相较于原 SuperPoint+SuperGlue 在各误差区间内图片数量更多. 本

文算法与原 SuperPoint+SuperGlue, SuperPoint+Lightglue, LoFTR, Aspanformer, 3DG-STFM^[31] 等图像匹配领域其他先进方法各像素误差区间内图片数量对比如表 3 所示,本文模型在各像素误差区间内均取得了较好的实验效果;整体图片数量随中心误差变化趋势如图 13 所示,本文算法在整体数据集上可以取得更好的匹配效果.

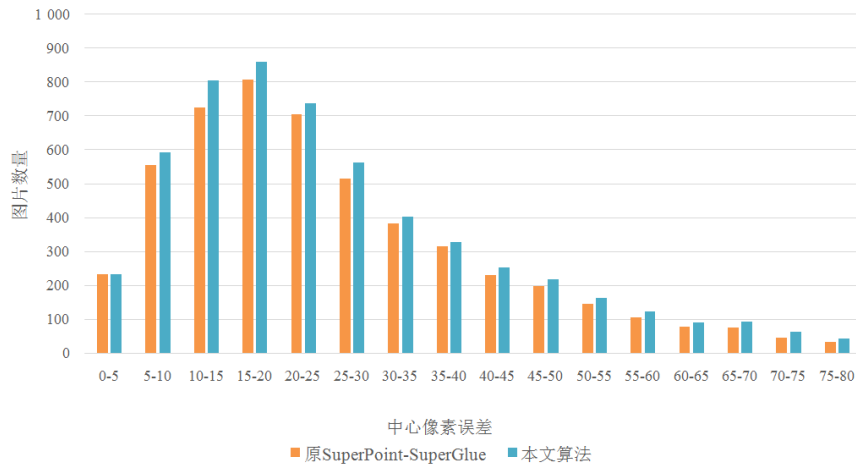


图 12 本文算法与原 SuperPoint-SuperGlue 各误差区间图片数量对比

表 3 各方法误差区间内图片数量对比

算法名称	误差区间内图片数量占数据集总数百分比/%				
	15 像素以内	30 像素以内	45 像素以内	60 像素以内	80 像素以内
本文算法	24.08	55.95	70.45	77.86	82.14
原 SuperPoint+SuperGlue	22.35	52.3	66	72.63	76.09
原 SuperPoint+Lightglue	20.36	48.04	60.97	67.95	72.28
LoFTR	20.42	48.26	61.04	68.29	72.81
3DG-STFM	23.53	54.34	68.53	76.57	81.27
Aspanformer	23.44	54.22	68.14	75.6	79.85

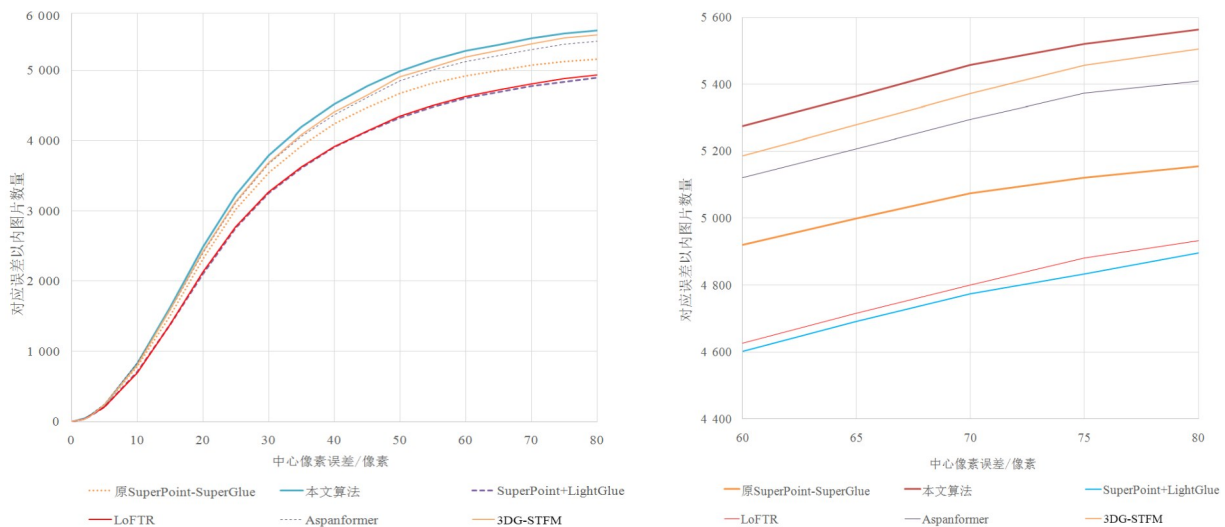


图 13 本文算法与对比算法图片数量随中心误差像素变化趋势

5 结论

针对卫星导航拒止或传感器精度不足时无人机定位需求,本文提出了一种基于改进 SuperPoint 的空天异源图像匹配算法. 该算法在原 SuperPoint 编码器基础上,提出了补充编码器与组合编码器. 补充编码器在 SuperPoint 编码器基础上进行改进,融合了 SGE 模块与 GAM 模块,加强了弱纹理区域特征提取能力;组合编码器通过并联补充编码器与原编码器,减少对相似区域的特征点误匹配,提高了空天异源图像的匹配精度. 经实验表明,基于改进 SuperPoint 的空天异源图像匹配算法有较好的匹配效果,可以有效解决空天异源图像匹配中特征提取能力较弱,特征分布不均匀等问题. 在 UAV-VisLoc 数据集上,80 像素误差区间以内可匹配图像数量可达到 82.14%,与原 SuperPoint 算法相比,80 像素误差区间以内可匹配图像数量提高了 6.05%,与其他先进算法相比,在各像素误差区间以内可匹配图像数量均有提高. 本文算法虽然较现有算法匹配效果有所提升,但本文所作研究仍在实时性与精确性权衡、多场景泛用性等方面有提升空间.

参考文献

- [1] XU S B, CHEN S P, XU R T, et al. Local feature matching using deep learning: A survey[J]. *Information Fusion*, 2024, 107: 102344.
- [2] 刘畅, 李嘉杰, 眭海刚, 等. 基于 FLoFTR 算法的无人机实时在线地理定位[J]. *航空科学技术*, 2023, 34(5): 87-94.
LIU C, LI J J, SUI H G, et al. Real-time geolocation on UAV based on FLoFTR algorithm[J]. *Aeronautical Science & Technology*, 2023, 34(5): 87-94. (in Chinese)
- [3] 任艳, 刘胜男, 陈新禹, 等. 不同季节下无人机航拍图像与卫星图像匹配方法研究[J]. *弹箭与制导学报*, 2023, 43(5): 16-24.
REN Y, LIU S N, CHEN X Y, et al. An image matching method for season-changing UAV aerial images and satellite images[J]. *Journal of Projectiles, Rockets, Missiles and Guidance*, 2023, 43(5): 16-24. (in Chinese)
- [4] 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). Piscataway: IEEE, 2018: 337-33712.
- [5] LI X, HU X L, YANG J. Spatial group-wise enhance: Improving semantic feature learning in convolutional networks[EB/OL]. (2019-05-25) [2024-07-20]. <https://arxiv.org/abs/1905.09646v2>.
- [6] LIU Y C, SHAO Z R, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions[EB/OL]. (2021-11-10) [2024-07-20]. <https://arxiv.org/abs/2112.05561v1>.
- [7] XU W J, YAO Y X, CAO J Q, et al. UAV-VisLoc: A large-scale dataset for UAV visual localization[EB/OL]. (2024-05-20)[2024-07-20]. <https://arxiv.org/abs/2405.11936v1>.
- [8] 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). Piscataway: IEEE, 2020: 4938-4947.
- [9] LI T Q, PANG G S, BAI X, et al. Learning adversarial semantic embeddings for zero-shot recognition in open worlds[J]. *Pattern Recognition*, 2024, 149: 110258.
- [10] ZHANG J, HUANG L, BAI X, et al. Exploring the usage of pre-trained features for stereo matching[J]. *International Journal of Computer Vision*, 2024, 132(10): 1.
- [11] BAI X, ZHANG P, YU X, et al. Learning from human attention for attribute-assisted visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, 46: 10.1109/TPAMI.2024.3458921.
- [12] WU Y C, ZHANG P C, GU M Y, et al. Embodied navigation with multi-modal information: A survey from tasks to methodology[J]. *Information Fusion*, 2024, 112: 102532.
- [13] 赵克龙. 基于改进 ORB 和改进 SURF 的影像匹配算法研究[D]. 西安: 西安科技大学, 2022.
ZHAO K L. Research on Image Matching Algorithm Based on Improved ORB and Improved SURF[D]. Xi'an: Xi'an University of Science and Technology, 2022. (in Chinese)
- [14] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]//2011 International Conference on Computer Vision. Piscataway: IEEE, 2011: 2564-2571.
- [15] 温育杜, 郭士增, 刘晟, 等. 基于尺度配准的航拍图像与卫星影像匹配方法[J]. *哈尔滨工业大学学报*, 2023, 55(10): 19-26.
WEN Y D, GUO S Z, LIU S, et al. Matching method between UAV images and satellite images based on scale registration[J]. *Journal of Harbin Institute of Technology*, 2023, 55(10): 19-26. (in Chinese)
- [16] 罗卿莉, 张书宾, 陈相, 等. 无人机与卫星影像匹配的相似性逼近算法[J]. *测绘通报*, 2023(9): 64-69.
LUO Q L, ZHANG S B, CHEN X, et al. A similarity approximation algorithm for matching UAV and satellite image[J]. *Bulletin of Surveying and Mapping*, 2023(9): 64-69. (in Chinese)
- [17] SUN J M, SHEN Z H, WANG Y A, et al. LoFTR: Detector-free local feature matching with transformers[C]//2021 IEEE/CVF Conference on Computer Vision and Pat-

tern Recognition (CVPR). Piscataway: IEEE, 2021: 8918-8927.

- [18] WANG Y F, HE X Y, PENG S D, et al. Efficient LoFTR: Semi-dense local feature matching with sparse-like speed[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 21666-21675.
- [19] CHEN H K, LUO Z X, ZHOU L, et al. ASpanFormer: Detector-free image matching with adaptive span transformer[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 20-36.
- [20] LINDENBERGER P, SARLIN P E, POLLEFEYS M. LightGlue: Local feature matching at light speed[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 17581-17592.
- [21] 眭海刚, 李嘉杰, 苟国华. 基于异源影像匹配的无人机在线快速定位方法[J]. 系统工程与电子技术, 2023, 45(10): 3008-3015.
- SUI H G, LI J J, GOU G H. Online fast localization method of UAVs based on heterologous image matching[J]. Systems Engineering and Electronics, 2023, 45(10): 3008-3015. (in Chinese)
- [22] 黄振家. 不同季节景象的无人机航拍图像与卫星影像匹配技术研究[D]. 沈阳: 沈阳航空航天大学, 2023.
- HUANG Z J. Research on Matching Technology of UAV aerial Images and Satellite Images in Different Seasons[D]. Shenyang: Shenyang Aerospace University, 2023. (in Chinese)
- [23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-08-10)[2024-07-20]. <https://arxiv.org/abs/1409.1556v6>.

- [24] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block attention module[C]//Computer Vision - ECCV 2018. Cham: Springer International Publishing, 2018: 3-19.
- [25] XIA G S, HU J, HU F, et al. AID: A benchmark dataset for performance evaluation of aerial scene classification.[EB/OL]. (2016-08-18)[2024-07-20]. <https://arxiv.org/abs/1608.05167>.
- [26] CHENG G, HAN J W, LU X Q. Remote sensing image scene classification: Benchmark and state of the art[J]. Proceedings of the IEEE, 2017, 105(10): 1865-1883.
- [27] ZOU Q, NI L H, ZHANG T, et al. Deep learning based feature selection for remote sensing scene classification[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(11): 2321-2325.
- [28] DAI D X, YANG W. Satellite image classification via two-layer sparse coding with biased image representation[J]. IEEE Geoscience and Remote Sensing Letters, 2011, 8(1): 173-176.
- [29] ZHAO B, ZHONG Y F, XIA G S, et al. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(4): 2108-2123.
- [30] ZHENG Z D, WEI Y C, YANG Y. University-1652: A multi-view multi-source benchmark for drone-based geolocalization[EB/OL]. (2020-08-16) [2024-07-20]. <https://arxiv.org/abs/2002.12186>.
- [31] MAO R Y, BAI C, AN Y T, et al. 3DG-STFM: 3D geometric guided student-teacher feature matching[M]//Computer Vision - ECCV 2022. Cham: Springer Nature Switzerland, 2022: 125-142.

作者简介



刘庚辰 男, 2000年2月出生. 现为中国空间技术研究院研究生. 主要研究方向为无人系统信息融合.
E-mail: 1259012790@qq.com



吴国强 男, 1984年9月出生. 现在中国航天科技集团有限公司任职. 主要研究方向为无人系统指挥控制、人工智能.



姜梁 男, 1976年12月出生. 现在中国航天科技集团有限公司任职. 主要研究方向为无人系统总体技术、人工智能.



黄坤 男, 1990年8月出生. 现在中国航天科技集团有限公司任职. 主要研究方向为无人系统情报处理.