

基于FLoFTR算法的无人机实时在线地理定位



刘畅¹, 李嘉杰¹, 眭海刚¹, 雷俊锋¹, 葛亮²

1. 武汉大学, 湖北 武汉 430072

2. 天津市测绘院有限公司, 天津 300381

摘要: 研制高效鲁棒的智能视觉定位方法是解决全球导航卫星系统(GNSS)拒止条件下无人机导航定位的重要途径之一。传统视觉定位方法存在精度较差、容易丢失定位的问题。本文提出一种FLoFTR算法,通过对高精度影像匹配算法LoFTR进行改进优化,在无人机计算平台上实现实时高精度定位。FLoFTR采用知识蒸馏方法压缩模型规模,提升推理效率,并通过改进特征提取模块和应用基于余弦距离的特征匹配方法,进一步降低了匹配时间并维持相当的匹配性能。在研制的软硬一体的平台上试验表明,优化后模型平均定位误差损失维持在0.1m以内,定位平均处理时间为47ms,定位速度提升超过7倍,可满足无人机定位的精度和实时性要求。

关键词: 无人机; 视觉定位; 图像匹配; LoFTR; 知识蒸馏

中图分类号: TP391

文献标识码: A

DOI: 10.19452/j.issn1007-5453.2023.05.012

在现代战争中,无人机成为越来越重要的作战装备,无人机被大量应用于侦察、目标打击等作战任务,对战场有着举足轻重的作用。然而面对复杂的作战环境,目前的无人机装备仍存在性能弱、无法建设成熟的作战体系的问题。特别是随着电磁干扰技术的发展,无人机常用的全球导航卫星系统(GNSS)极易受到干扰,对无人机应用于作战环境存在非常不利的影响。

近年来,计算机视觉发展迅速,基于深度学习的图像匹配技术通过在卫星地图上匹配无人机影像来实现无人机视觉地理定位^[1-4],并成为解决GNSS拒止环境下无人机定位问题的一种有效途径。然而,无人机平台算力和功耗有限,如何在有限的设备环境下保证视觉定位实时性并兼顾定位精度,是实现无人机实时视觉定位的重点。

影像匹配作为视觉地理定位的核心算法^[5-6],也是计算机视觉的重要研究方向。定向快速角点与抗旋转描述(ORB)^[7]、尺度不变特征变换(SIFT)^[8]和加速稳健特征(SURF)^[9]等算法是具有代表性的传统特征匹配方法,常被用于即时定位与地图构建(SLAM)、视觉里程计等无人机

定位系统。然而,传统特征匹配算法的效果并不稳定,在无人机发生较大位姿变换或在纹理稀疏区域时,传统匹配算法无法实现鲁棒匹配效果,导致位姿估计有较大的误差。因此,这类方法常被用于室内小范围的定位感知。对于室外测绘应用,传统特征匹配方法虽然实时性较好,但无法保证定位的精度,鲁棒性较差。

近年来,基于深度学习的匹配算法不断涌现,相比传统算法,深度学习大幅提升了匹配精度。D. Detoe等^[10-11]提出了结合注意力机制^[12]的特征点提取和匹配方法,将基于深度学习的特征匹配性能提升到了新高度。此外,还有Sun等^[13]、Wang等^[14]和Chen Honghai等^[15]基于深度学习的影像匹配方法,分别从匹配精度、模型规模等方面出发,进一步提升了匹配算法的性能。然而,大多数模型的设计都是在不考虑设备限制的情况下追求精度提升的,无法在算力受限的无人机上得到有效应用,以至于深度学习匹配算法难以在无人机视觉定位应用中发挥其高鲁棒性和高精度匹配优势。

寻求高精度和高实时性的图像匹配算法对于实时性要求较高的应用场景具有重要意义^[16]。针对图像匹配算法鲁

收稿日期: 2023-02-03; 退修日期: 2023-03-15; 录用日期: 2023-04-13

基金项目: 航空科学基金(2019460S5001); 广西科技重大专项(AA22068072)

引用格式: Liu Chang, Li Jiajie, Sui Haigang, et al. Real-time geolocation on UAV based on FLoFTR algorithm[J]. Aeronautical Science & Technology, 2023, 34(05): 87-94. 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(05): 87-94.

棒性和实时性的“瓶颈”问题,本文基于无特征检测的局部特征变换器匹配(LoFTR)算法,通过结合模型优化算法,提出了一种在无人机平台应用智能化视觉定位方法的解决方案。优化后的模型在鲁棒性和实时性方面有较好的均衡,能够在研制的软硬一体的嵌入式平台上实现较高精度的实时视觉地理定位。

1 无人机视觉地理定位

无人机视觉地理定位方法有效利用了无人机视觉载荷和预先制备的高分辨卫星底图,通过将无人机影像和卫星底图影像进行匹配,实现无人机高精度、稳定的定位。其基本框架如图1所示。

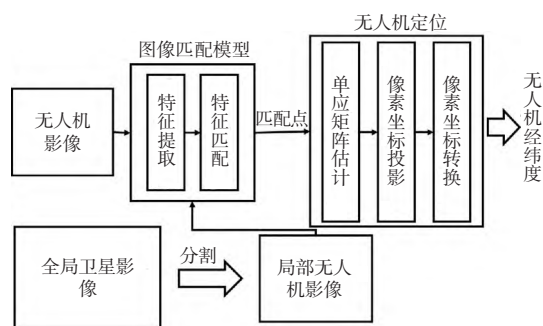


图1 视觉地理定位流程

Fig.1 The framework of visual geo-localization

无人机影像由云台相机拍摄,并实时传输到机载计算机进行数据处理。卫星影像预先在地面制备,根据无人机的飞行高度、飞行区域和飞行环境在起飞前从数据库中提取最合适的卫星影像,并存储在机载计算机中。首先,对卫星影像进行分割,保证定位的精度和效率;其次,将无人机影像和切割的卫星影像进行匹配,输出匹配点;最后,估计单应矩阵,得到影像间坐标映射关系,将无人机影像投影到卫星影像上,并通过卫星影像上存储的地理坐标实现无人机定位,输出无人机的经纬度。

2 LoFTR 影像匹配算法模型

影像匹配算法是视觉地理定位方法的核心,其性能直接影响定位的精度和效率。LoFTR算法是近两年具有代表性的影像匹配算法,具有鲁棒性强、精度高的优点。因此,以LoFTR作为视觉地理定位的核心算法能改善定位精度和鲁棒性,使无人机能在GNSS拒止下保持正常飞行。

LoFTR算法的模型框架如图2所示,模型有4个模块,分别是特征提取模块、Transformer编码模块、粗匹配模块和精匹配模块。将待匹配的影像对输入模型,通过融合特征金字塔的残差网络(ResNet)^[17]骨干网提取多尺度特征,然后利用Transformer模块分别对不同尺度的特征进行自注意力和交叉注意力编码,粗匹配模块通过构建代价矩阵和置信度矩阵输出粗匹配点,精匹配模块通过高分辨率特征微调粗匹配的结果,输出亚像素级的匹配点,精匹配应用了坐标回归^[18]。

3 模型优化方法

由于多模型组合形成的庞大模型无法满足有限的无人机计算和功耗资源,需要压缩模型提高匹配效率。本文将从网络规模压缩和模型推理改进进行优化,知识蒸馏实现对特征提取和Transformer模块的网络规模压缩;余弦距离度量实现粗匹配推理提速。

3.1 知识蒸馏

Gou等^[19]首次提出了知识蒸馏,通过教师网络指导精简学生网络训练,实现知识迁移,使学生网络兼顾效率和性能。知识蒸馏的基本框架如图3所示。

将知识蒸馏引入教师网络,同时采用教师网络输出的软标签和数据集的硬标签对学生模型进行训练。在训练过程中,负样本也带有大量信息,对模型收敛有促进作用。因此,采用软标签能够通过增大信息量加速模型训练,提高模型精度,实现教师模型知识迁移。硬标签则保证模型精度,防止可能受到软标签的错误知识引导。知识蒸馏中的温度

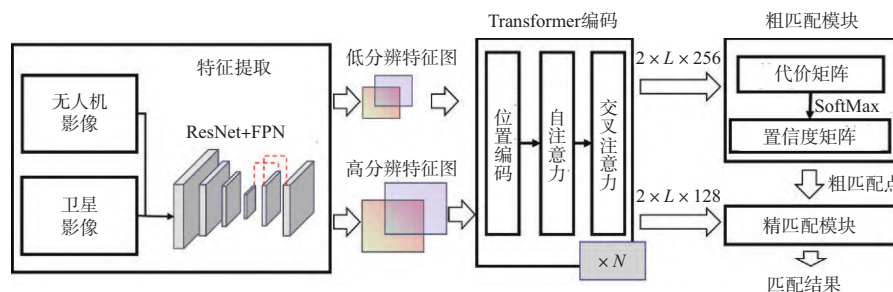


图2 LoFTR模型结构

Fig.2 Network structure diagram of LoFTR

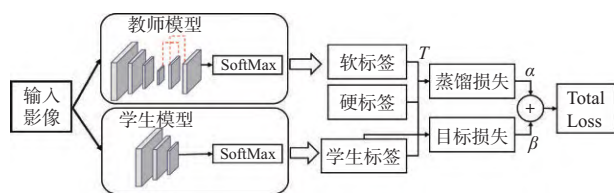


图3 知识蒸馏训练框架

Fig.3 Training process of distillation knowledge

T 是用来控制对负样本的关注程度,温度较低时,对负标签尤其是那些显著低于平均值的负标签的关注较少;而温度较高时,负标签相关的值会相对增大。但由于负标签具有非常大的噪声,并不可靠,因此温度 T 是一个经验值^[20]。目前,知识蒸馏算法对于分类模型具有较好的效果。

LoFTR模型的粗匹配模块实际是通过分类每个像素点实现最相关匹配的。其原理如图4所示。

将提取的特征图展平为 $L \times 256$,在Transformer编码后输入粗匹配模块, L 代表特征图的像素个数。通过计算每两个特征矢量间的点积相似度,输出 $L \times L$ 的代价矩阵。然后利用归化指数函数(SoftMax)分别处理代价矩阵的行和列,将输出的两个结果相乘得到置信度矩阵。显然,该矩阵的元素表示每个像素点的匹配概率。LoFTR的粗匹配模块实质上是对每个像素点的分类,因此可以将置信度矩阵作为知识蒸馏的软标签指导学生模型训练。输出能实现粗匹配的轻量化模型。

3.2 学生模型设计

根据3.1节的描述,粗匹配实质上是分类模型,可以采

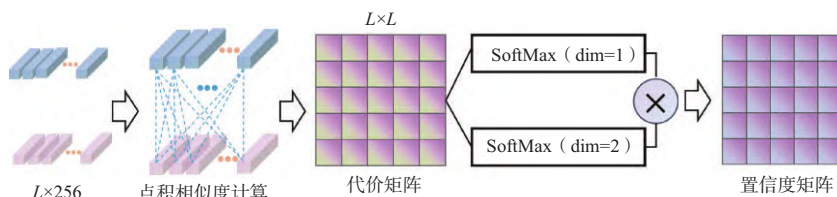


图4 粗匹配模块

Fig.4 The coarse matching module

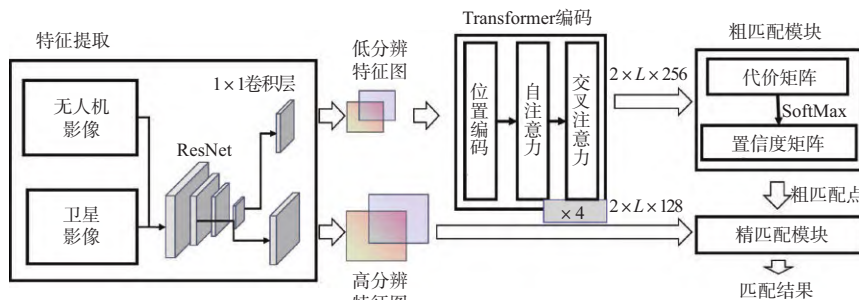


图5 学生模型结构

Fig.5 The structure diagram of student model

用知识蒸馏框架训练学生模型的特征提取、Transformer和粗匹配模块的相关网络参数。

本文还对特征提取模块进行改进,提升推理速度。特征金字塔在无人机影像匹配中的作用并不明显,如果无人机的飞行高度变化显著,特征金字塔无法解决过大尺度变化;LoFTR粗匹配直接采用ResNet提取的低分辨率特征,并未融合高分辨率特征;LoFTR模型的精匹配模块以粗匹配结果为基础,即精匹配只提高粗匹配结果的精度。因此,特征金字塔大大降低了模型推理效率,本文在学生模型中直接采用两个不同的 1×1 卷积层替代特征金字塔输出多尺度特征,加速推理速度。

此外,本文改进LoFTR模型精匹配模块以进一步提升推理速度。该模块分为Transformer编码和坐标回归两部分,Transformer编码参数量较大,推理时间慢;坐标回归基于特征图,效率较高。但回归会导致知识蒸馏无法训练网络参数,该模块需要冻结参数二次训练。本阶段改进策略包括:(1)粗匹配模块已经利用Transformer建立了较精确的匹配关系,精匹配模块采用未编码特征图直接回归实现鲁棒的坐标精细化。(2)精匹配阶段的Transformer编码只面向以匹配点为中心 5×5 的局部特征图,注意力的作用并不显著。在改进阶段中舍弃了该编码部分。

综上,完整的学生模型框架如图5所示。其需要两次训练。知识蒸馏训练特征提取、Transformer编码和粗匹配模块的网络参数。通过降低ResNet和Transformer网络的

通道数和层数等来提升推理效率。原模型和知识蒸馏的模型参数见表1。冻结训练用于单独训练学生模型的精匹配模块。

表1 模型参数

Table 1 Model parameters

网络	模块	LoFTR	知识蒸馏
ResNet	初始卷积层	128	16
	残差卷积层	128, 196, 256	16, 32, 64
Transformer	通道数	256, 128	64, 0
	层数N	8	4
	注意力头	8	1

3.3 基于余弦相似度量的影像匹配

知识蒸馏只能提升模型中的卷积神经网络部分,对于特征间相似度计算、匹配点输出等模块并没有显著作用。实测表明,LoFTR模型中基于SoftMax的粗匹配方案计算较为延时。本文基于余弦距离改进了粗匹配方法,提升了匹配速度。

根据3.1节的描述,LoFTR模型的粗匹配模块实际是为了计算得到特征间的相似性。在训练过程中,由于梯度计算的要求,需要SoftMax输出软标签。而在推理过程中,SoftMax函数不是必须的,只需计算特征相似度并输出最大值即可实现影像匹配。由此,相似度计算由特征点积修改成特征余弦距离,代价矩阵直接输出匹配,效率进一步提高。其原理如图6所示。集合知识蒸馏,形成了FLoFTR模型。

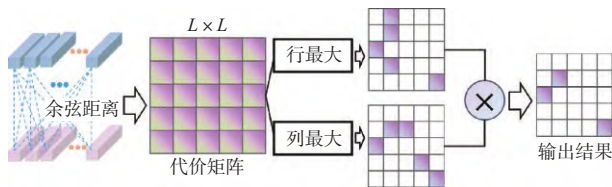


图6 基于余弦距离的粗匹配

Fig.6 The coarse matching module based on cosine distance

3.4 训练

综合模型优化,完整的训练步骤分为两步:首先,利用LoFTR预训练模型指导训练特征提取、Transformer和粗匹配模块的网络参数。其次,输出模型后,冻结部分参数并单独训练精匹配模块,输出完整的轻量化匹配模型。

对于知识蒸馏训练,根据3.1节的描述,训练损失由蒸馏损失和目标损失两部分组成,分别对应相对教师模型和真值的损失。蒸馏损失采用相对熵(KL散度)计算信息损失程度。蒸馏损失利用原模型的代价矩阵生成的概率分布矩阵指导学生模型训练,计算方式如式(1)所示

$$L_d = \sum_i P_{s,i} \log \left(\frac{P_{s,i}}{P_{t,i}} \right) \cdot T^2 \quad (1)$$

式中, $P_{s,i}$ 表示模型对*i*位置像素的匹配概率估计, T 为蒸馏温度,训练中蒸馏温度设置为 x 。

目标损失计算学生模型和地面真值之间的交叉熵损失。目标损失的计算如式(2)所示,同时完整的损失计算如式(3)所示

$$L_t = - \sum_i P_{gt,i} \log P_{s,i} \quad (2)$$

$$L = \alpha L_d + \beta L_t \quad (3)$$

式中, α 和 β 代表两部分损失的权重,训练过程中将蒸馏损失设置为0.7,目标损失设置为0.3,选择这样的值是为了补偿蒸馏损失相对较小的值,使两部分的贡献更加相等^[19]

对于精匹配模块进行单独训练,仅需训练ResNet模型中输出高分辨率特征图的1×1卷积层的参数,其他参数均冻结。本文同时训练了基于余弦距离和LoFTR模型匹配方法的两个模型,用于后续试验对比。

精匹配其中一张影像的匹配点像素位置保持不变,通过调整另一张影像对应匹配点的像素位置,提高匹配点的精度。采用L2范数训练损失对应匹配点像素坐标误差,计算方法如式(4)所示

$$L_f = - \frac{1}{|P_f|} \sum_{(i,j \in P_f)} \|j - j_{gt}\| \quad (4)$$

式中, i 表示参照影像的像素位置, j 表示坐标微调影像的像素位置, j_{gt} 是通过相机参数、深度信息和参比影像的像素位置*i*计算出的地面真值。

4 对比试验

4.1 数据准备和测试环境

试验使用了4个不同的数据集,分别是用于训练的MegaDepth^[21]数据集、BlendedMVS^[22]数据集,用于单应性对比试验的HPatches^[23]数据集和用于定位性能对比的武汉城郊自建数据集。

MegaDepth数据集和BlendedMVS数据集用于训练模型,这两个数据集都是常见的用于训练影像匹配模型的数据集。根据本文3.1节,训练分为两步:首先,使用预训练的LoFTR模型作为教师模型训练出粗匹配的知识蒸馏模型。LoFTR预训练模型在MegaDepth数据集上训练得到,MegaDepth数据集包含了不同拍摄角度、尺度的匹配影像对,具有数据量大、影像差异明显的特点,对训练出鲁棒性强、精度高的模型有明显作用。知识蒸馏模型在BlendedMVS数据集上训练,该数据集相比MegaDepth数据

量较小,能有效缓解数据量大的训练时长和硬件要求高的问题,同时增强模型泛化性。其次,得到粗匹配的知识蒸馏模型后,在MegaDepth数据集上训练精匹配模块。精匹配模块对匹配精度有较大影响,因此采用与预训练模型相同的MegaDepth数据集保持数据一致性。

HPatches数据集包含不同光照和不同角度的匹配影像,同时有对应的单应矩阵真值,用于评估模型估计的单应矩阵的精度。

武汉城郊自建数据集由大疆精灵4拍摄,相机型号为DJI FC6310R鱼眼相机,传感器尺寸为13.2mm×8.8mm,像幅尺寸为5472px×3648px,像元大小为2.41μm,相机焦距为8.8mm,采集过程为下视拍摄正射影像。鱼镜头头的焦距很短、视角很大,其拍摄的影像相较于平面底图存在较大畸变,所以为了减小匹配定位难度和保证定位精度,根据相机参数对鱼眼相机拍摄的影像做了畸变矫正。无人机平台如图7所示。



图7 无人机平台

Fig.7 UAV platform

分别有200m、250m、300m拍摄高度的纹理丰富城区影像1500张,以及300m拍摄高度的纹理稀疏丛林影像260张。同时数据集包含了无人机拍摄时的GNSS数据,作为定位精度的客观真值。卫星影像从谷歌地球上截取,地面分辨率0.5m。

本文分别使用经典传统匹配方法(SIFT)、LoFTR、知识蒸馏后的LoFTR(KD)和FLoFTR4种方法进行对比试验。所有对比试验均在研制的软硬一体的原型系统中进行。该系统以嵌入式开发板Jetson AGX Orin为主体,集成了控制算法、定位算法和卫星影像数据等,形成软硬一体的无人机视觉地理定位系统。

4.2 单应性对比

本文首先在HPatches上评估算法性能。HPatches数据集包含57个光照变化序列、59个视角变化序列,每一个序列包含6张匹配影像。对于每一个序列,取一幅图像作为参考图像,与其他5张图像进行匹配,并估计单应矩阵。通过对比估计单应矩阵和数据集内的真实单应矩阵判断评估

匹配算法的精度,并分析知识蒸馏模型的精度损失。单应矩阵的精度通过影像4个角点的精度评估,计算公式如(5)所示

$$\frac{\|M(H_{pd} - H_{gt})\|}{4} \quad (5)$$

式中, M 为由影像4个角点构成的矩阵。 H_{pd} 和 H_{gt} 分别为估计的单应矩阵和真实的单应矩阵。

试验分别统计了误差在3px以内、3~5px和5px以上的单应矩阵数量占比,结果见表2。

表2 单应矩阵精度

Table 2 Precision of homography matrix

	3px 以内	3~5px	5px 以上
SIFT	61.52	9.45	29.03
LoFTR	80.20	5.68	14.12
KD	60.47	19.48	19.69
FLoFTR	66.51	11.75	21.74

从试验结果可以发现,FLoFTR综合精度较优于传统SIFT方法,但与原模型LoFTR比较有一定的精度下降,这是由于压缩模型后造成的不可避免的精度损失。具体分析来看,误差在三个像素以内的单应矩阵,仅知识蒸馏后损失相对显著,而FLoFTR模型则更好地保留了原模型精度,精度损失维持在15%以内。精度在3~5px的单应矩阵数据占比增加较多。对于像素误差超过5px,不论是知识蒸馏模型还是FLoFTR模型,数量占比增长控制在7%左右。因此,知识蒸馏保持了较好的整体精度。而定位采用地面分辨率为0.5m的卫星影像,5px误差在2.5m左右,足够满足无人机定位要求。

4.3 定位性能对比

本文在武汉郊区自建数据集上进行定位性能评估。无人机影像和卫星影像匹配后,利用随机抽样一致(RANSAC)算法估计单应矩阵并通过其计算无人机影像中心在卫星底图上的投影坐标,然后根据卫星地图的地理坐标转换得到无人机的经纬度坐标。

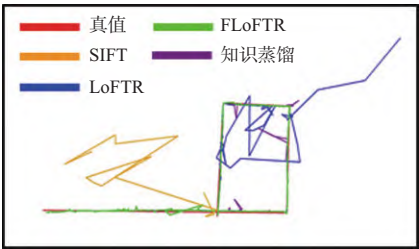
本文设计了不同拍摄高度对比试验,用来比较验证本文算法在无人机不同飞行高度下的定位有效性,试验数据分别为200m、250m和300m拍摄高度的城区影像。试验结果如表3和图8所示,图中红色直线代表真实轨迹,橙色代表SIFT算法预测轨迹,蓝色代表LoFTR算法预测轨迹,紫色代表LoFTR算法知识蒸馏后预测轨迹,绿色代表FLoFTR算法预测轨迹。

从不同高度无人机定位试验结果可以发现,SIFT算法

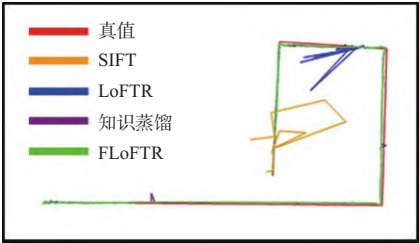
表3 城市场景定位精度

Table 3 Localization precision of city scene

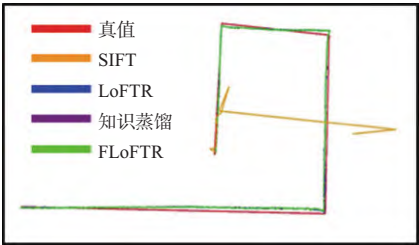
高度	算法模型	平均误差/m	最大误差/m
200m	SIFT	—	—
	LoFTR	—	—
	KD	1.943	56.917
	FLoFTR	1.881	32.345
250m	SIFT	—	—
	LoFTR	—	—
	KD	1.732	13.232
	FLoFTR	1.497	4.983
300m	SIFT	—	—
	LoFTR	2.81	8.49
	KD	2.95	8.51
	FLoFTR	2.91	8.60



(a) 200m



(b) 250m



(c) 300m

图8 城市场景定位轨迹

Fig.8 Localization trajectory of city scene

定位失败,这是由于SIFT算法无法有效匹配无人机与卫星底图这种异源且场景复杂的图像。LoFTR算法在飞行高度300m定位成功,但在250m和200m定位失败,这是因为LoFTR算法具有不抗尺度的局限性,从而对无人机飞行高度变化的定位适应性差。其他两个模型在不同飞行高度均可整体定位,轨迹的完整度和连续性都比较理想,这是因为

在知识蒸馏的过程中舍弃了Transformer编码与回归部分,是算法模型直接学习拟合了训练数据中存在的多尺度特征,提高了算法对不同尺度图像匹配的适应性。同时根据表3数据对比,FLoFTR算法综合适应性和平均定位精度相较于其他算法最高,证明了知识蒸馏和余弦相似度改进的有效性。但随着无人机飞行高度的降低,定位算法的单应性假设也会逐渐失效,从而影响算法稳定性,使部分定位区域误差较大,飞行高度过低甚至会导致定位算法完全失效。

本文同时设计了纹理稀疏场景的试验,试验数据还包括300m高度拍摄的丛林区域影像,定位结果如图9所示,客观评估结果见表4。

表4 丛林场景定位精度

Table 4 Localization precision of wild scene

算法模型	平均误差/m	最大误差/m
SIFT	—	—
LoFTR	1.59	4.74
KD	1.65	6.59
FLoFTR	1.67	5.71

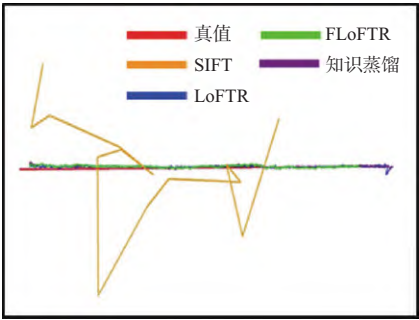


图9 丛林场景定位轨迹

Fig.9 Localization trajectory of wild scene

从图9和表4中可以发现,经典传统方法SIFT同样无法实现无人机视觉定位,三个深度学习模型的定位差异较小,知识蒸馏后的模型保留了原模型的大部分精度,虽然受纹理影响,FLoFTR相比原模型最大定位误差的精度损失接近1m,但平均定位误差的精度损失维持在0.1m以内,证明FLoFTR对于不同场景的无人机定位仍具有一定的鲁棒性。

综合多场景试验可证明,经过知识蒸馏压缩改进后的模型相较于原模型的定位精度未出现显著下降,且增强了算法对飞行高度的适应性,满足无人机不低于200m飞行高度的自主定位的需求。

4.4 效率对比

优化算法的核心目标是提供可应用在无人机平台的高效视觉地理定位方法,因此对三个模型的运行效率进行了

详细测试。在 Jetson AGX Orin 嵌入式载板上推理时间(见表 5), 试验结果多次测试并取平均值, 由于 SIFT 算法无法实现定位, 试验未统计其计算耗时。

表 5 模型推理时间(单位:ms)

Table 5 Model inference time(Unit:ms)

算法模型	特征提取	Transformer	粗匹配	精匹配	总时间
LoFTR	145	67	28	60	300
KD	8	17	34	2	61
FLoFTR	8	17	20	2	47

表 5 中数据表明, 通过知识蒸馏压缩模型, 有效降低了特征提取和 Transformer 编码模块的推理时间, 并进一步提升了精匹配的效率; 通过余弦距离计算相似度, 有效改善了 SoftMax 方法时间复杂度高的问题, 降低了粗匹配推理耗时。需要注意的是, LoFTR 和知识蒸馏模型粗匹配模块的耗时理论上应该是一致的, 此处存在细微差别是由测试存在波动和误差引起的。优化后模型 FLoFTR 的推理时间为 47ms, 相比原模型提升超过 7 倍, 基本达到实时定位。

5 结论

为解决无人机在 GNSS 拒止和计算资源受限条件下的定位问题, 本文基于知识蒸馏提出了面向无人机实时在线视觉地理定位的 FLoFTR 算法。在研制的软硬一体机载嵌入式平台上的对比试验显示, FLoFTR 在定位精度与推理速度之间获得了较好的均衡, 满足无人机应用需求。

FLoFTR 为无人机提供了一种机载高精度和高效的视觉地理定位方法, 但目前的研究测试环境比较理想, 对于真实复杂环境中的场景、天气、天候等变化, 算法的鲁棒性和精确性需要进一步提升, 为复杂环境下的无人机定位提供坚实技术支撑。



参考文献

[1] Gyagenda N, Hatilima J V, Roth H, et al. A review of GNSS-independent UAV navigation techniques[J]. Robotics and Autonomous Systems, 2022, 135: 104069.

[2] Couturier A, Akhloufi M A. A review on absolute visual localization for UAV[J]. Robotics and Autonomous Systems, 2021, 135: 103666.

[3] Sui H, Li J, Lei J, et al. A fast and robust heterologous image matching method for visual Geo-localization of low-altitude UAVs[J]. Remote Sensing, 2022, 14(22): 5879.

[4] 刘飞, 单佳瑶, 熊彬宇, 等. 基于多传感器融合的无人机可降

落区域识别方法研究[J]. 航空科学技术, 2022, 33(4):19-27.

Liu Fei, Shan Jiayao, Xiong Binyu, et al. Research on the identification method of UAV landing area based on multi-sensor fusion[J]. Aeronautical Science & Technology, 2022, 33(4):19-27. (in Chinese)

[5] 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.

[6] 罗世彬, 刘海桥, 胡茂青, 等. 无人飞行器异源图像匹配辅助惯性导航定位技术综述[J]. 国防科技大学学报, 2020, 42(6): 1-10.

Luo Shibin, Liu Haiqiao, Hu Maoqing, et al. Review of multi-modal image matching assisted inertial navigation positioning technology for unmanned aerial vehicle[J]. Journal of National University of Defense Technology, 2020, 42(6):1-10. (in Chinese)

[7] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]. International Conference on Computer Vision. IEEE, 2011: 2564-2571.

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

[9] Bay H, Tuytelaars T, Van Gool L. Surf: speeded up robust features[C]. European Conference on Computer Vision. Springer, 2006: 404-417.

[10] Detone D, Malisiewicz T, Rabinovich A. Superpoint: self-supervised interest point detection and description[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018: 224-236.

[11] Sarlin P E, Detone D, Malisiewicz T, et al. Superglue: Learning feature matching with graph neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4938-4947.

[12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 15: 30.

[13] Sun J, Shen Z, Wang Y, et al. LoFTR: Detector-free local feature matching with transformers[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8922-8931.

[14] Wang Q, Zhang J, Yang K, et al. MatchFormer: Interleaving

- attention in transformers for feature matching[C]. Proceedings of the Asian Conference on Computer Vision, 2022: 2746-2762.
- [15] Chen Honghai, Luo Zixin, Zhou Lei, et al. ASpanFormer: Detector-free image matching with adaptive span transformer [C]. European Conference on Computer Vision. Springer, 2022: 20-36.
- [16] 赵晓冬, 张洵颖, 车军, 等. 精确制导武器末制导目标识别优化算法研究[J]. 航空科学技术, 2022, 33(1): 126-134.
- Zhao Xiaodong, Zhang Xunying, Che Jun, et al. Research on optimization algorithm of terminal guidance target recognition for precision-guided weapons[J]. Aeronautical Science & Technology, 2022, 33(1): 126-134. (in Chinese)
- [17] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [18] Pereira T D, Tabris N, Matsliah A, et al. SLEAP: A deep learning system for multi-animal pose tracking[J]. Nature Methods, 2022, 19(4): 486-495.
- [19] Gou J, Yu B, Maybank S J, et al. Knowledge distillation: A survey[J]. International Journal of Computer Vision, 2021, 129: 1789-1819.
- [20] Gao Y, Zhao L. Coarse TRVO: a robust visual odometry with detector-free local feature[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2022, 26(5): 731-739.
- [21] Li Z, Snavely N. Megadepth: learning single-view depth prediction from internet photos[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2041-2050.
- [22] Yao Y, Luo Z, Li S, et al. Blendedmvs: A large-scale dataset for generalized multi-view stereo networks[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1790-1799.
- [23] Balntas V, Lenc K, Vedaldi A, et al. HPatches: A benchmark and evaluation of handcrafted and learned local descriptors[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 5173-5182.

Real-time Geolocation on UAV Based on FLoFTR Algorithm

Liu Chang¹, Li Jiajie¹, Sui Haigang¹, Lei Junfeng¹, Ge Liang²

1. Wuhan University, Wuhan 430072, China

2. Tianjin Institute of Surveying and Mapping Co., Ltd., Tianjin 300381, China

Abstract: Developing an efficient and robust intelligent visual localization method is an effective means to solve the navigation and positioning of UAV under GNSS denial conditions. However, the accuracy of traditional visual localization method is poor, thus leading to lose location easily. By improving and optimizing the high-precision image matching algorithm LoFTR, this paper propose FLoFTR algorithm to achieve real-time localization on the UAV. FLoFTR uses knowledge distillation method to compress the model size and improve the inference efficiency. By improving the feature extraction module and applying the cosine distance, FLoFTR further reduces the matching time and maintains comparable matching performance. The experimental results show that the average localization error loss of the optimized model is less than 0.1m. And the average localization processing time is 47ms, which is increased by more than 7 times and can meet the precision and real-time requirements of UAV localization.

Key Words: UAV; visual localization; image matching; LoFTR; knowledge distillation

Received: 2023-02-03; **Revised:** 2023-03-15; **Accepted:** 2023-04-13

Foundation item: Aeronautical Science Foundation of China (2019460S5001); Guangxi Science and Technology Major Project (AA22068072)