甘仁德, 李健. 一种无参数的多尺度特征融合跨视图地理定位方法[J]. 智能计算机与应用,2026,16(5):119-125. DOI:10.20169/j.issn.2095-2163.25111701

一种无参数的多尺度特征融合跨视图地理定位方法

甘仁德, 李 健
(四川大学 电子信息学院, 成都 610065)

摘 要: 跨视图地理定位的核心是从不同视角的图像中提取到稳定一致的特征。然而,如何建立鲁棒的跨视图特征表达,仍是一个具有挑战性的关键问题。目前大多数方法主要通过大规模的模型结构或复杂注意力机制增强特征提取能力,但这些方法难以满足模型实时性要求的同时,仅学习到单一粒度的特征表示,无法让模型充分理解场景从整体布局到局部细节的多层次信息。为了满足模型实时性的同时实现更先进的性能,本文构建了无参数多尺度特征融合网络,弥补了单一粒度特征的局限性,使模型能够更全面地理解场景的复杂性。此外,本文提出一种负样本积累策略,在不消耗额外内存的前提下增加了负样本数量,增强模型提取特征的能力。在 University-1652 基准数据集上的实验结果表明,相较于当前先进方法,本文方法不仅实现了更先进的性能,而且模型参数减少了 165.21 M,推理时间缩短了 0.412 s,能够更好地满足跨视图地理定位的实时性需求。

关键词: 跨视图地理定位; 无参数; 多尺度特征融合; 负样本积累策略; 实时性
中图分类号: TP391.41 **文献标志码**: A **文章编号**: 2095-2163(2026)05-0119-07

Parameter-free multi-scale feature fusion method for cross-view geolocation

GAN Rende, LI Jian
(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: The core challenge of cross-view geolocation lies in effectively extracting stable and consistent features from images captured at different viewpoints. However, establishing robust cross-view feature representations remains a critical and challenging problem. Most existing approaches primarily enhance feature extraction via large-scale model architectures or complex attention mechanisms. However, these methods struggle to meet real-time requirements while only learning single-granularity feature representations, failing to enable models to capture multi-level information ranging from global scene layout to local details. To achieve both real-time performance and advanced capabilities, this article proposes a parameter-free multi-scale feature fusion network that overcomes the limitations of single-granularity features, enabling models to comprehensively capture scene complexity. Additionally, this paper proposes a negative sample accumulation strategy that increases the number of negative samples without consuming extra memory, thereby enhancing the model´s feature extraction capability. Experiments on the University-1652 benchmark demonstrate that compared to current state-of-the-art methods, this proposed method not only achieves superior performance but also reduces model parameters by 165.21M and shortens inference time by 0.412s, better meeting the real-time requirements of cross-view geolocation.

Key words: cross-view geolocation; parameter-free; multi-scale feature fusion; negative sample accumulation strategy; real-time

0 引 言

跨视图地理定位(CVGL)旨在通过匹配不同视角的图像,确定目标的地理位置[1]。早期研究主要聚焦于地面与卫星图像的匹配[2],但因视角差异过大,实际应用受限。近年来,随着无人机技术的突破,CVGL 迎来了新机遇:无人机提供的视图与卫星视图的角度差异较小,显著降低了跨视图匹配的难度[3]。然而,跨视图视角差异导致的语义鸿沟仍是核心挑战[4]。如何提取鲁棒的跨视图特征表达,仍是目前亟待突破的关键问题。

CVGL 主要有两大典型应用:无人机视图目标

作者简介: 甘仁德(1999—),男,硕士研究生。主要研究方向:拒止环境下无人机视觉定位技术。
通信作者: 李 健(1969—),女,博士,副教授。主要研究方向:信号检测,频谱感知,强化学习。Email:1446935022@qq.com。

收稿日期: 2025-11-17

定位和无人机导航。目前解决上述两项任务的方法主要围绕基于深度学习的图像检索[5]。在跨视图图像检索中,其核心是有效从不同视角的图像中提取稳定一致的特征[6]。早期,大多数方法基于卷积神经网络(CNN)提取特征[7]。然而,基于 CNN 的方法受限于高斯先验假设下的局部感受野设计,其特征提取多集中于局部区域的判别性特征[8],导致模型难以捕捉全局信息。主流方法[9]则通过引入注意力机制增强挖掘全局信息的能力。过度使用注意力模块可能显著增加计算资源的消耗[10],无法满足实时定位的要求。同时引入注意力也无法使模型全面获取场景从整体布局到局部细节的多层次信息。目前,Transformer 网络因其强大的上下文信息提取能力被广泛地应用于 CVGL[11]。但是,基于 Transformer 网络的方法参数量大,计算复杂度高,难以满足实时定位的要求[12]。为了更好地满足实时定位的要求,同时提取到更多细节特征,本文构建了一个多尺度特征融合网络,旨在融合全局(粗粒度)和局部(细粒度)等多粒度特征信息,弥补单一粒度特征提取的局限性。值得一提的是,本文构建的多尺度特征融合网络是一个无参数的网络,不会增加模型参数的数量。

在比较提取到的不同源特征相似性时,三重态损失被广泛使用[13]。然而,该损失函数每次仅使用一个负样本,使得模型在面对难负样本时容易出现坍塌现象[14]。Deuser F[15]等已经证明,增加负样本的数量可以让模型更好地学习到特征表示。增加批次样本的数量会消耗额外内存,因此本文提出一种基于 In-foNCE[16] 损失的负样本积累策略,该策略在没有消耗额外内存的基础上增加了负样本的数量,进而增强了模型对相似性场景的区分能力。

1 多尺度特征融合跨视图地理定位方法

多尺度特征融合可以实现图像全局与局部信息的融合,改善网络提取特征的能力[17]。对于不同视角的卫星与无人机图像,即使在同一区域,图像的内容也会受到天气、阴影和季节变化的影响[6]。为克服卫星图像和无人机图像之间的差异,有效提取跨视图特征信息,本文以 ConvNeXt-T 网络为基础,构建一个孪生网络模型(如图 1 所示)不同视角的图像经过 ConvNeXt-T 网络处理后传输到无参数的多尺度特征融合网络做进一步处理,提取融合多粒度特征信息的多层次特征,便于后续相似性计算,从而确定地理位置。

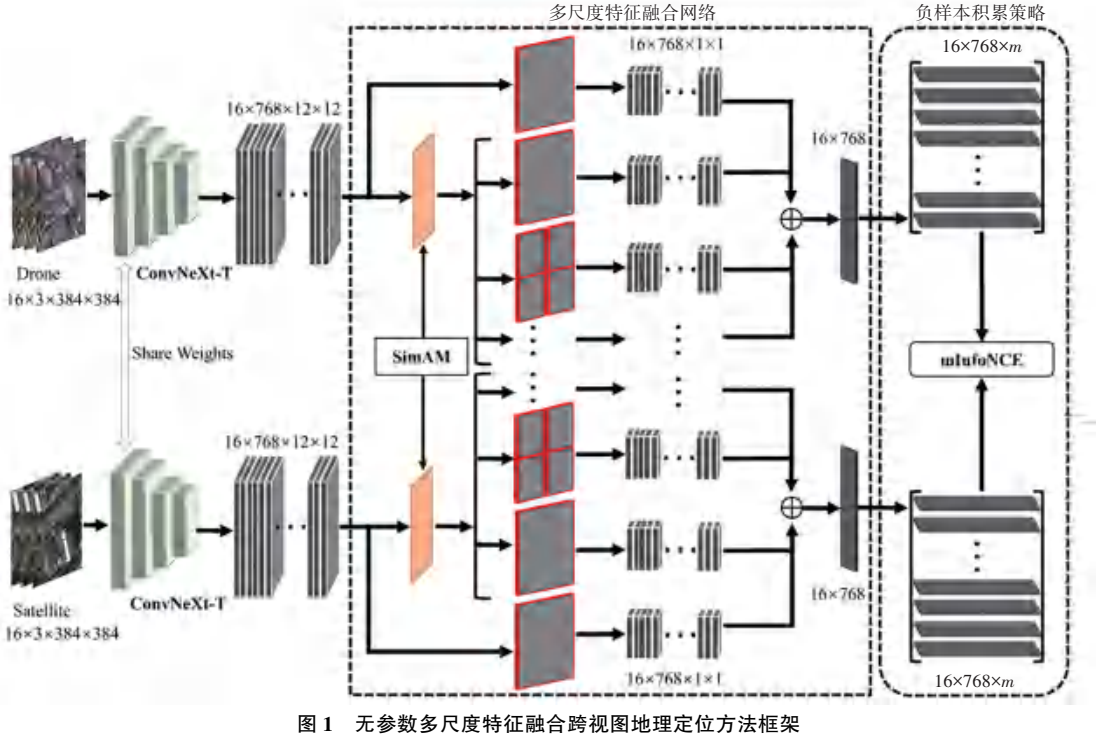


图 1 无参数多尺度特征融合跨视图地理定位方法框架

Fig. 1 Parameter-free multi-scale feature fusion framework for cross-view geolocation

1.1 多尺度特征融合网络

在 CVGL 任务中,现实世界场景的复杂性要求模型要能够捕捉到从整体布局到局部细节的多层次信息,以确保定位的准确性。ConvNeXt-T 在提取图像的粗粒度特征方面表现优异,但可能会丢失一些对精确定位至关重要的细粒度特征。细粒度特征丢

失会加剧特征匹配和定位的难度。因此，本文构建了一个多尺度特征融合网络。如图 2 所示，本文对主干网络输出的特征进行不同尺度的划分，从而提取到从粗粒度到细粒度的特征。粗粒度特征能够揭示场景的整体布局和结构，为模型提供宏观层面的指导。细粒度特征则深入挖掘场景中物体的局部细节和纹理，为模型提供精确的局部细节信息。多尺度特征融合网络不仅有助于保留深度卷积过程中可能遗漏的细粒度特征，还可以让模型能够更全面地理解场景的复杂性，准确识别场景中的各个组成部分及其相互关系，从而在定位和导航任务中实现更高的精度。

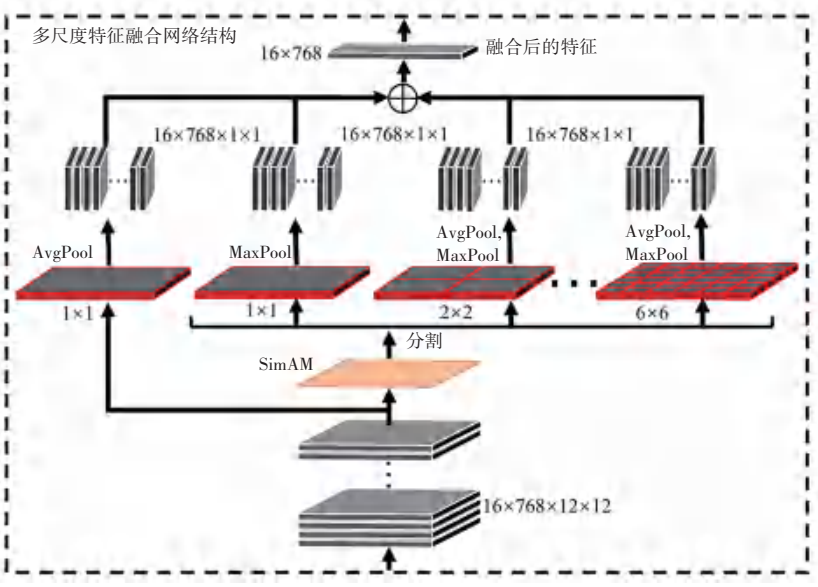


图 2　多尺度特征融合网络的结构

Fig. 2　Structure of multi-scale feature fusion networks

为了让模型可以更好地捕捉不同视角图像中的关键信息，本文在构建无参数多尺度特征融合网络的时候引入无参数注意力 SimAM。与传统的通道注意力和空间注意力模块不同，SimAM 能够为特征图的每个神经元分配一个三维的注意力权重，从而更好地捕捉关键特征信息[18]。简单来说，SimAM 的核心思想是基于图像的局部自相似性，对特征进行增强处理，其增强方式定义如下：

$$\tilde{F} = \mathrm{Sigmoid}\left(\frac{1}{E}\right) \odot F \tag{1}$$

其中，$\tilde{F}$ 是增强后的特征；F 是增强前的特征；E 是经过计算的能量张量。

经过 ConvNeXt-T 网络后，每一个特征图的大小为 12×12。特征会分为两条线进行下一步处理，一条是直接进行平均值池化操作，以关注图中的全局特征。如下式所示：

$$\tilde{F}_0 = \mathrm{AvgPool}(F) \tag{2}$$

其中 F 表示经过主干网络处理后的特征。

另一条是先将特征 F 经过 SimAM 注意力，得到进一步增强以后的特征，具体见式(1)。本文对增强后的特征进行 1×1，2×2，3×3，4×4，6×6 尺度的划分。划分块的大小是特征图大小的一个因数，避免了提取到过多的边界信息。经过划分的特征会分为两种处理方式，对于 1 × 1 的尺度直接进行最大值池化，以关注图中最为突出的特征。如下式所示：

$$\tilde{F}_1 = \mathrm{MaxPool}(F^1_{\mathrm{SimAM}}) \tag{3}$$

其中，F^1_{SimAM} 表示经过 SimAM 注意力处理以后被划分成 1 × 1 尺度大小的特征。

对于 2×2，3×3，4×4，6×6 的尺度特征，对其进行如式(4)所示的池化操作，先对分块的区域进行平均值池化，然后再对平均值池化后的区域特征进行最大值池化，以关注图中信息最为突出区域的特征。对于分块的增加，其实就是不断地缩小突出特征区域的范围，以关注到特征最为突出的区域。4 种尺度的特征融合表达式如下：

$$\begin{cases} \tilde{F}_2 = [\mathrm{AvgPool}(F^2_{\mathrm{SimAM}}), \mathrm{MaxPool}(F^2_{Avg})] \\ \tilde{F}_3 = [\mathrm{AvgPool}(F^3_{\mathrm{SimAM}}), \mathrm{MaxPool}(F^3_{\mathrm{Avg}})] \\ \tilde{F}_4 = [\mathrm{AvgPool}(F^4_{\mathrm{SimAM}}), \mathrm{MaxPool}(F^4_{\mathrm{Avg}})] \\ \tilde{F}_6 = [\mathrm{AvgPool}(F^6_{\mathrm{SimAM}}), \mathrm{MaxPool}(F^6_{\mathrm{Avg}})] \end{cases} \tag{4}$$

其中，F^2_{SimAM}，F^3_{SimAM}，F^4_{SimAM}，F^6_{SimAM} 表示经过

SimAM 注意力处理以后被划分成 2×2,3×3,4×4,6×6 尺度的特征; $F^2_{\text{Avg}}, F^3_{\text{Avg}}, F^4_{\text{Avg}}, F^6_{\text{Avg}}$ 表示经过平均池化以后对应尺度的特征。

经过上述处理以后,将得到的多个特征进行融合,最终的特征 $\tilde{F}_f$ 如下式:

$$\tilde{F}_f = \sum_{i=0}^{5} \tilde{F}_i \tag{5}$$

本文提出的特征融合网络全部是池化操作,SimAM 也是一个无参数的注意力机制模块。因此,该特征融合网络在改善模型性能的同时不增加模型的参数量。

1.2 负样本积累策略

三重态损失函数被广泛应用于 CVGL 任务中,进行损失计算时,每个锚点样本通常与一个正样本和一个负样本进行对比。Deuser F[15] 等证明了增加负样本的数量可以让模型更好地学习到特征表示。直接增加批次的样本数量会显著增加内存使用量。为解决计算机内存受限的情况下负样本扩充问题,本文提出了负样本积累策略,具体实现方式如下式:

$$x = \{x_1 \cup x_2 \cup \cdots \cup x_m\} \tag{6}$$

其中, m 为指定的负样本积累批次数量; x_1 是第 1 个批次样本的特征集合; x 是 m 个批次样本特征的集合。

InfoNCE 是一种通过比较正样本和多个负样本之间的相似性,从而驱动模型学习更具判别性特征的损失函数。本文基于 InfoNCE 构建了适配负样本积累策略的损失函数,具体形式如下:

$$L_{\text{InfoNCE}}(q,R,m) = -\log \frac{\exp(q.\frac{r_+}{\tau})}{\sum_{i=1}^{m}\sum_{j=0}^{R}\exp(q.\frac{r_{ij}}{\tau})} \tag{7}$$

其中, q 表示一个编码后的查询图像; R 表示一组编码后的参考图像; m 为指定的样本批次数量。分子部分 r_+ 表示唯一与查询 q 匹配的正样本,分母部分的 r_{ij} 为式(6)中的 x 集合的元素,其中 $i \in \{1,2,\cdots,m\}$, $j \in \{0,1,2,\cdots,R\}$。当 $\sum_{i=1}^{m}\sum_{j=0}^{R} r_{ij}$ 中的 $m=1$ 时,该损失函数退化为标准 InfoNCE 损失函数。τ 是超参数,也称为温度系数。

2 实验结果与分析

本文在 University-1652 数据集上对提出的方法进行评估,并在 $R@K$ 以及 AP 得分方面与目前先进的方法进行比较,验证提出方法的有效性。

2.1 实验数据集以及评价指标

University-1652[3] 是覆盖全球 72 所大学 1 652 个地点的多视图、多源数据集,其提供了卫星视图、无人机视图和地面视图数据。数据集包含用于训练的 33 所大学的 701 个采样点和用于测试的 39 所大学的 951 个采样点,两者之间没有重叠。University-1652 数据集支持两个新的跨视图任务:无人机视图目标定位(Drone→Satellite)和无人机导航(Satellite→Drone),为 CVGL 研究提供了丰富的资源,并带来了视点变化等独特挑战。

实验采用了 Recall@K($R@K$) 和平均精度(AP) 两个关键指标来衡量模型的性能:$R@K$ 用于衡量在模型生成排名的前 K 位结果中正确匹配的图像所占全部图像的比例,比例越高,表明模型的性能越优。同时,本文还计算了 AP,该指标能够综合反映检索性能中的查准率和查全率,全面评估模型效果。

2.2 实验实现细节

在实验中,本文采用在 ImageNet 上预训练的 ConvNeXt-T 作为主干网络。参数初始化方面,由于多尺度特征融合网络不带参数,因此不需要对其初始化。优化器方面,本文采用 AdamW 优化器,初始学习率设为 0.001,同时采用具有 n epoch(n 是实数) 预热期的余弦学习率调度器。损失函数方面,本文提出了一个负样本积累策略来指导模型优化参数的损失函数,其中超参数 label_smoothing=0.1,温度参数 τ 为可学习参数。训练阶段,输入图像的大小为 384×384,并采用图像增强策略以提升模型的稳健性。本文模型基于 PyTorch 框架实现,所有实验均在内存为 11 GB 的 Nvidia GTX 1080 GPU 上完成。

2.3 实验验证

在 University-1652 基准上对本文提出的方法进行评估,结果见表 1,表 1 给出了在无人机视角目标定位(Drone→Satellite)和无人机导航应用(Satellite→Drone)任务中与先进方法进行对比的结果。与先进的 SRLN[19] 进行比较,在 Drone→Satellite 任务中, $R@1$ 和 AP 分别增长了 0.26% 和 0.31%,在 Satellite→Drone 任务中, $R@1$ 和 AP,分别增长了 0.87% 和 0.34%。同时相较于 SRLN 方法,本文提出的方法模型参数少了 165.21 M,推理时间快了 0.412 s。与 FSRA[11]、Safe-Net[20]、MCCG[21]、Sample4Geo[15]、CCR[22] 等几种方法相比,

本文提出的方法不仅参数量小、推理时间快，而且性能也是最先进的。进一步证明了该方法的有效性及实时性。

表 1　在 University-1652 数据集上与不同方法的比较

Table 1　Comparison with different methods on the University-1652 dataset

方法	主干	模型参数/ M	推理时间/ s	Drone→Satellite		Satellite→Drone	
				R@1/ %	AP/%	*R*@1/ %	AP/%
FSRA	Vit-S	52.01	0.316	84.51	86.71	88.45	83.37
Safe-Net	Vit-S	53.05	0.321	89.12	91.10	92.87	88.52
MCCG	ConvNext-T	37.47	0.313	89.64	91.32	94.30	89.39
Sample4Geo	ConvNext-B	87.57	0.363	92.65	93.81	95.14	91.39
CCR	ConvNext-B	169.09	0.416	92.54	93.78	95.15	91.80
SRLN	Swin-B	193.03	0.573	92.70	93.77	95.14	91.97
本文方法	ConvNext-T	**27.82**	**0.161**	**92.96**	**94.08**	**96.01**	**92.31**

如图 3 所示，将本文方法在 Satellite→Drone 任务中的 Top-1 检索定位结果与同结构但参数量不同的模型 Sample4Geo 和 CCR 进行对比，本文方法在 Sample4Geo 与 CCR 模型都检索定位出错的地点上实现了正确的检索。从图 3 中的热力图可以看出，相较于 Sample4Geo 和 CCR 方法，本文的方法更加关注于图中整体场景布局和局部区域细节的多层次信息，这一特征对本文方法具有卓越检索定位能力提供了有力证明。

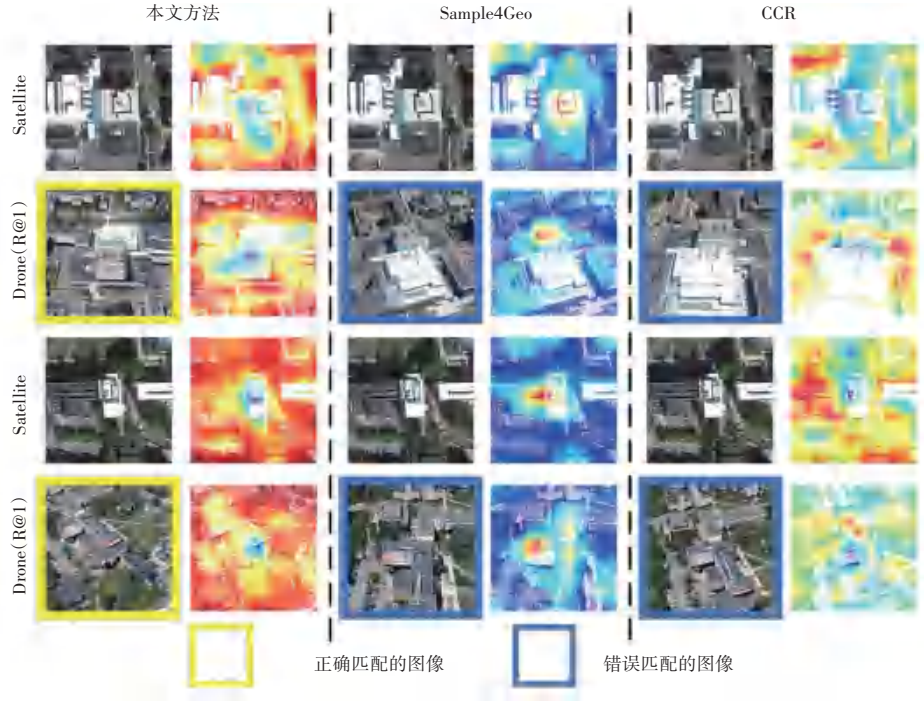


图 3　Satellite→Drone 任务中不同方法的 Top-1 检索结果对比

Fig. 3　Comparison of Top-1 retrieval results of different methods in Satellite→Drone tasks

2.4　消融实验

2.4.1 多尺度特征划分组合消融实验

尺度的划分是多尺度特征融合网络中的一个关键参数，为了确定最佳的尺度划分组合，本文进行了相关的消融实验。在实验中，本文依次从 1×1 尺度增加到 4×4 尺度，然后在增加到 6×6 尺度。从表 2 中可以看出，随着增加划分尺度的组合，在 Drone→Satellite 和 Satellite→Drone 任务中，*R*@1 和 AP 都呈现一定的上升趋势，增加到 4×4 时达到多个指标的最优，相较于基础模型，在 Drone→Satellite 任务中，

R@1 和 *AP* 中分别提升了 10.65% 和 9.29%。在 Satellite→Drone 任务中，*R*@1 和 AP 分别提升了 8.84% 和 11.12%。虽然继续将尺度增加到 6×6 时，Drone→Satellite 任务的 *R*@1 和 AP 略优于 4×4 的结果，但从整体指标来说，增加到 4×4 时，1×1，2×2，3×3，4×4 已经是最佳的尺度划分组合。

表 2 最佳尺度划分组合的消融实验

Table 2 Ablation experiments for optimal scale partitioning combinations

基础模型	1×1	2×2	3×3	4×4	6×6	Drone→Satellite		Satellite→Drone	
						R@1/%	AP/%	*R*@1/%	AP/%
√						76.84	80.15	85.31	75.69
√	√					86.43	88.44	91.87	84.82
√	√	√				86.22	88.24	92.72	84.48
√	√	√	√			86.40	88.46	93.15	85.66
√	√	√	√	√		87.49	89.44	**94.15**	**86.81**
√	√	√	√	√	√	**87.59**	**89.45**	93.44	86.27

2.4.2 加入不同模块的消融实验

为了验证所提方法各个组成部分的有效性，本文开展对应的消融实验。结果见表 3，在基础模型的基础上依次添加多尺度特征融合网络以及负样本积累策略模块，每增加一个模块，网络的性能都有一定的提升。实验结果表明通过叠加本文提出来的模块，可以有效提高跨视图地理定位的结果。

表 3 不同模块的消融实验

Table 3 Ablation experiments in different modules

基础模型	多尺度特征融合	负样本积累策略	Drone→Satellite		Satellite→Drone	
			R@1/%	AP/%	*R*@1/%	AP/%
√			76.84	80.15	85.31	75.69
√	√		87.49	89.44	94.15	86.81
√		√	90.87	92.35	93.87	89.34
√	√	√	**92.96**	**94.08**	**96.01**	**92.31**

3 结束语

本文提出了一种简单、实时性高且适用于跨视图地理定位的方法。该方法结合了多尺度特征融合网络与负样本积累策略。其中，多尺度特征融合网络旨在不增加模型参数量的同时融合不同粒度的特征，将细粒度结构特征与粗粒度全局特征相结合，使模型能够更全面地理解场景的复杂性，准确识别场景中的各个组成部分及其相互关系，从而在定位和导航任务中实现更高的精度。负样本积累策略在不消耗额外内存的基础上增加了样本的数量，增强了模型区分相似性场景的能力。实验结果表明，与现有先进的方法相比，本文方法对硬件资源的依赖更低，可更好地适应资源受限的设备，同时具有更快的推理速度。未来，将进一步探索缩小模型的参数量，在不降低模型性能的同时实现模型在资源受限的设备上的应用。

参考文献

[1] 石浚茹. 面向无人机视觉定位的图像匹配算法研究[D]. 成都：电子科技大学，2025.

[2] TIAN Y, CHEN C, SHAH M. Cross-view image matching for geo-localization in urban environments[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 3608-3616.

[3] ZHENG Z, WEI Y, YANG Y. University-1652: A multi-view multi-source benchmark for drone-based geo-localization[C]// Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 1395-1403.

[4] 张天博. GNSS 拒止环境下无人机视觉定位技术研究[D]. 沈阳：沈阳航空航天大学，2024.

[5] ZHU S, YANG T, CHEN C. Vigor: Cross-view image geo-localization beyond one-to-one retrieval[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 3640-3649.

[6] LIU G, LI C, ZHANG S, et al. VL-MFL: UAV visual localization based on multisource image feature learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-12.

[7] WANG T, ZHENG Z, ZHU Z, et al. Learning cross-view geo-localization embeddings via dynamic weighted decorrelation regularization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024,62:1-12.

[8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of Advances in Neural Information Processing Systems. NeurIPS,2017:5998-6008.

[9] YUE P, XIN J, ZHANG Y, et al. Semantic-driven autonomous visual navigation for unmanned aerial vehicles [J]. IEEE Transactions on Industrial Electronics, 2024, 71(11): 14853-14863.

[10] 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. Piscataway, NJ: IEEE, 2016: 770-778.

[11] DAI M, HU J, ZHUANG J, et al. A transformer-based feature segmentation and region alignment method for UAV-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(7): 4376-4389.

[12] 孙子文,钱立志,杨传栋,等. 基于 Transformer 的视觉目标跟踪方法综述[J]. 计算机应用,2024,44(5):1644-1654.

[13] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway,NJ:IEEE, 2015: 815-823.

[14] WU C Y, MANMATHA R, SMOLA A J, et al. Sampling matters in deep embedding learning[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 2840-2848.

[15] DEUSER F, HABEL K, OSWALD N. Sample4geo: Hard negative sampling for cross-view geo-localisation [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway,NJ:IEEE,2023: 16847-16856.

[16] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from naturallanguage supervision[C]// Proceedings of International Conference on Machine Learning. PMLR, 2021: 8748-8763.

[17] 周晓吉. 基于多尺度特征融合的单目深度估计算法[J]. 智能计算机与应用,2024,14(9):34-40.

[18] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]// Proceedings of International Conference on Machine Learning. PMLR, 2021: 11863-11874.

[19] LV H, ZHU H, ZHU R, et al. Direction-guided multiscale feature fusion network for geo-localization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-13.

[20] LIN J, LUO Z, LIN D, et al. A self-adaptive feature extraction method for aerial-view geo-localization[J]. IEEE Transactions on Image Processing, 2025, 34: 126-139.

[21] SHEN T, WEI Y, KANG L, et al. MCCG:A ConvNeXt-based multiple-classifier method for cross-view geo-localization[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(3): 1456-1468.

[22] DU H, HE J, ZHAO Y. CCR: A counterfactual causal reasoning-based method for cross-view geo-localization [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 11630-11643.