埃菲尔铁塔：用于长期视觉定位的深海水下数据集

Cl´ementin Boittiaux[1,2,3], Claire Dune[2], Maxime Ferrera[1], Aur´elien Arnaubec[1], Ricard Marxer[3], Marjolaine Matabos[4], Lo¨ıc Van Audenhaege[4] 和 Vincent Hugel[2]

摘要
视觉定位在机器人系统于先前访问过的环境中进行位置确定和导航方面发挥着重要作用。当访问间隔较长时，与季节或昼夜周期相关的环境变化会带来重大挑战。在水下，变化来源则是水况或海洋生物生长等其他因素。然而，这仍然是一个主要障碍，也是一个研究较少的问题，部分原因在于数据匮乏。本文提出了一个新的深海数据集，用于对水下长期视觉定位进行基准测试。该数据集由在五年期间对同一热液喷口建筑体进行四次访问所获取的图像组成。利用导航数据和运动恢复结构，估计了相机位姿以及场景的统一几何结构。这些信息可作为评估视觉定位技术时的参考。对数据的分析揭示了多年间观察到的主要变化。此外，本文还在该数据集上评估了几种成熟的视觉定位方法，结果表明，水下长期视觉定位仍有改进空间。相关数据已公开发布于 seanoe.org/data/00810/92226/。

关键词
水下数据集　　t，长期视觉定位，深海，视觉定位基准，埃菲尔铁塔热液喷口建筑体。

arXiv:2305.05301v1 [cs.CV] 9 May 2023

引言

随着自主水下航行器（AUVs）和遥控潜水器（ROVs）的出现，人们需要使这些设备能够在水下环境中实现精确的自主定位。本文研究的是视觉定位问题，即根据相机拍摄的图像以及该区域的先前观测结果，估计相机的六自由度（6DOF）位姿。随着自动驾驶汽车的发展，这一问题在过去十年中受到了广泛关注。当环境发生重大变化时，这项任务会变得更加困难，例如在深海连续多次探访期间采集的图像便面临这一问题。长期定位方法旨在应对环境中的重大变化，例如，冬季的降雪（Sattler 等人 2018年）。

用于评估这些方法的大多数数据库都由陆地数据组成。它们涵盖了广泛的环境变化，例如昼夜、季节和天气状况（Griffith 等人 2017年；Sattler 等人 2018年；Burnett 等人 2023）。

由于水下图像具有不同来源的技术和环境变化，现有数据集并不适合评估此类场景中的长期定位性能。事实上，水下介质的特性会导致许多与光线和颜色吸收、浑浊度及后向散射相关的视觉扰动。此外，在深海场景中，水下航行器必须配备人工照明系统，以照亮环境中的绝对黑暗。虽然这使得使用摄像机记录海床上的景象成为可能，但也会造成照明方面的显著差异，而这种差异取决于机器人与场景之间的距离。

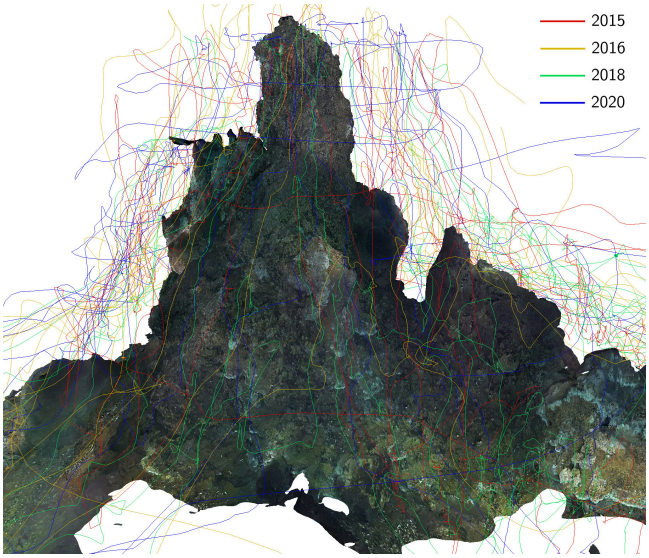


图1。 沿着埃菲尔铁塔 热液喷口的轨迹。使用 COLMAP（Sch¨onberger 和 Frahm 2016年）恢复了相机位姿。使用 OpenMVS（Cernea 2020年）对 3D 模型进行了网格化和纹理处理。

[1]Ifremer，布雷加永港区，拉塞讷叙尔梅尔，法国[2]土伦大学，COSMER，土伦，法国[3]土伦大学，艾克斯-马赛大学，法国国家科学研究中心，LIS，土伦，法国[4]布雷斯特大学，法国国家科学研究中心，Ifremer，UMR6197 BEEP，F-29280 普卢扎内，法国

通讯作者： Cl´ementin Boittiaux，Ifremer 地中海中心，布雷加永港区，83500 拉塞讷叙尔梅尔，法国 电子邮件：boittiauxclementin@gmail.com

除了这些短期扰动之外，这些环境中还会发生长期变化，例如微生物和动物群落分布的变化或地形变化。

一些定位方法依赖深度检索和特征匹配网络，以应对大范围的变化来源。然而，这些网络，例如，NetVLAD（Arandjelovic 等人 2016年），仅使用陆地数据进行预训练。由于水下成像的特殊性所导致的领域偏移，在水下环境中应用这些模型可能并不简单。

此前发布的面向在线定位的水下数据集（Ferrera 等人，2019 年；Mallios 等人，2017 年）所涵盖的时间跨度都非常短，即，数据均在同一天内采集。因此，这些数据集无法涵盖此类环境中可能出现的长期变化，导致目前缺乏能够处理多年期深海图像序列的方法和数据集。

<escape>本文提出了一个用于深海环境下长期视觉定位的新数据集。该数据集由五年间对同一热液地貌进行的四次不同探访所构成（Girard 等人，2020年）（图 1）。更具体地说，它提供了以下数据：</escape>

- 四次访问期间热液喷口的图像。
- 以纬度、经度和海拔信息形式呈现的导航数据。
- 3使用运动恢复结构（SfM）为每次访问年份估算的场景三维模型。
- 包含所有图像并采用统一参考坐标系的全球三维模型。

该数据集呈现了与上述所有水下成像因素相关的随时间变化情况，还展现了一些独特特征，例如会喷出高温热液流体的黑烟囱和白烟囱。此外，由于该数据集中存在众多变化来源，它还可用于检测深海环境中发生的长期变化。本文的贡献如下：i）提供了一个新的公开数据集，用于深海环境中的长期视觉定位；ii）分析了不同访问期间的环境和地形变化；iii）在该数据集上对多种视觉定位方法进行了基准测试。

相关工作

用于对视觉定位算法进行基准测试的Sattler 等人(2018年)以及 Cambridge（Kendall 等人 2015年）、7-Scenes（Shotton 等人 2013年）和 12-Scenes（Valentin 等人 2016年）介绍的 Aachen Day-Night、RobotCar Seasons 和 CMU Seasons 等数据集大多为陆地环境。7-Scenes 和 12-Scenes 采集于室内环境，而其他数据集则由室外环境构成。Sattler 等人(2018年) 数据集展现了一些具有挑战性的定位场景，例如昼夜观测。由于数据采集成本高昂，用于对水下视觉定位任务进行基准测试的类似数据集非常少见。

现有水下数据集（Mallios 等人，2017年；Ferrera 等人，2019年）侧重于为以下领域的开发提供数据：

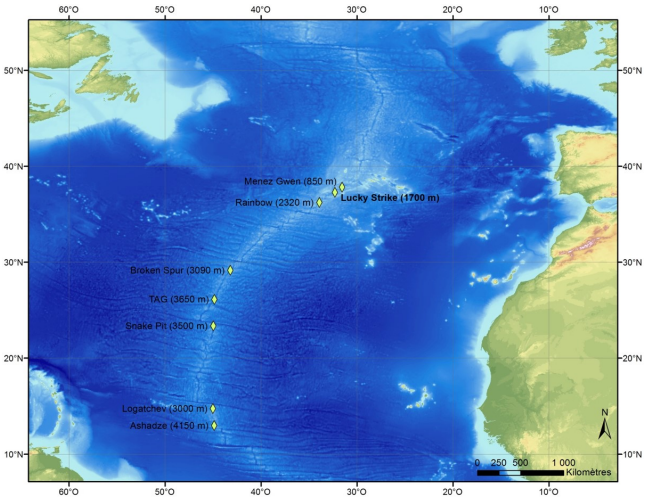
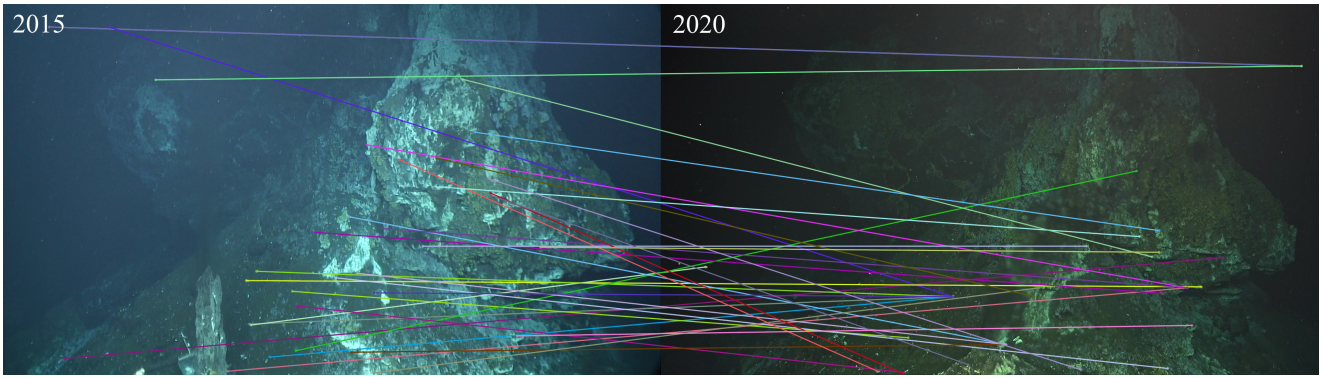


图 2. 大西洋中脊上 Lucky Strike 热液喷口场的位置（来源：Esri、GEBCO、NOAA、National Geographic、DeLorme、HERE、Geonames.org）。

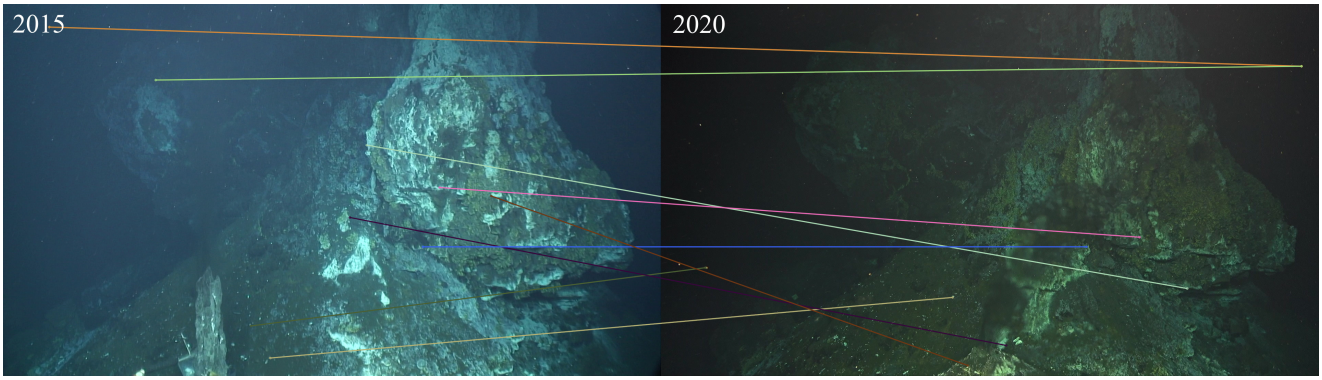
水下SLAM算法。AQUALOC数据集（Ferrera 等人，2019年）提供了与惯性数据和深度数据同步的水下单色图像，采集自科西嘉岛附近的三个不同地点。其中一个地点是水深3至4米的港口，另外两个是分别位于水深270米和380米处的考古遗址。尽管各段序列沿不同轨迹采集，但所有不同的访问都发生在同一天，未能涵盖这一环境中可能发生的所有变化，例如，可能改变针孔模型的盐度变化、浑浊度增加、沉积或海洋生物群落变化。Nielsen 等人（2019年）在一个游泳池中采集的水下数据集上评估了 PoseNet（Kendall 等人，2015年），这是一种端到端视觉定位神经网络；其中的相机位姿由水下动作捕捉系统获取。其他侧重于不同任务的水下数据集，例如去雾（Akkaynak 和 Treibitz，2019年；Li 等人，2019年；Berman 等人，2021年），并未提供同一地点长期拍摄的图像。

Campos等人(2015年) 已经发布了2015年对埃菲尔铁塔热液喷口进行探访时采集的图像数据。作者提出了一种用于水下结构表面重建的新方法。为评估其算法，他们使用了根据2015年潜水作业图像通过运动恢复结构生成的点云。本研究也利用运动恢复结构为该数据集生成参考位姿和场景几何结构，并补充了同一地点在不同年份进行的另外三次潜水作业所采集的类似数据，以记录长期变化。

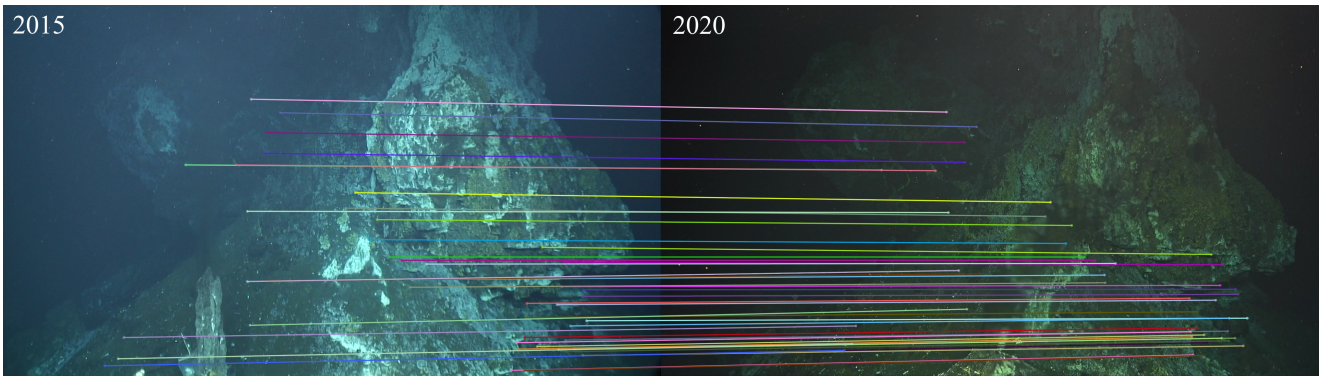
视觉定位基准数据集需要为每张图像提供参考相机位姿，这些位姿可以通过不同方式构建。获取这类信息和场景几何结构最常见的方法依赖于运动恢复结构或基于深度的SLAM（Brachmann 等人 2021）。在大多数深海任务中，由于部署此类系统十分困难，因此不会采用动作捕捉。标准RGB-D传感器无法直接在水下使用，因为红外光会被水吸收，这使得基于深度的SLAM更难部署。因此，运动恢复结构为估计水下环境中的相机位姿和场景几何结构提供了一种实用方案。然而，Brachmann 等人



(a)使用暴力匹配的 SIFT。



(b)使用暴力匹配的 SIFT。匹配 然后在估计基础矩阵后进行筛选 在 RANSAC 框架内。



（c）SuperPoint（DeTone等，2018年），随后是SuperGlue（Sarlin等，2020年）。

图3。 使用不同方法进行跨年度图像之间的特征匹配。

(2021年)表明，定位方法在给定数据集上的性能会受到构建该数据集“真实值”所用方法的显著影响。事实上，采用与估计真实姿态的算法相同误差进行最小化的方法，具有收敛到相同局部极小值的优势。因此，倾向于采用基于运动恢复结构的真实值的方法，可能在我们的数据集上表现更好。

表1. 摄像机设置。

Year	分辨率	帧率
2015	1920x1080 像素	25 帧/秒
2016	1920x1080 像素	25 帧/秒
2018	1920x1080 像素	25 帧/秒
2020	3840x2160 像素	30 帧/秒

数据采集

位于大西洋中脊的EMSO-亚速尔群岛深海观测站，自2010年以来一直对*LuckyStrike* 热液喷口区（图2）进行长期监测。在年度维护航次期间（Cannat andSarradin2010），法国国家海洋科学研究所（Ifremer）使用遥控潜水器研究热液循环的演变及其相关的

数年间的动物群落（Matabos 等，2022年）。在这一领域中，热液喷口建筑体 埃菲尔铁塔位于海面下1700米处，自1992年发现以来一直受到广泛研究（Langmuir 等，1993年）。2015年、2016年、2018年和2020年进行了四次潜水作业，旨在对该结构进行三维重建，从而实现对热液喷口群落分布和动态变化的定量监测（Girard 等，2020年）。

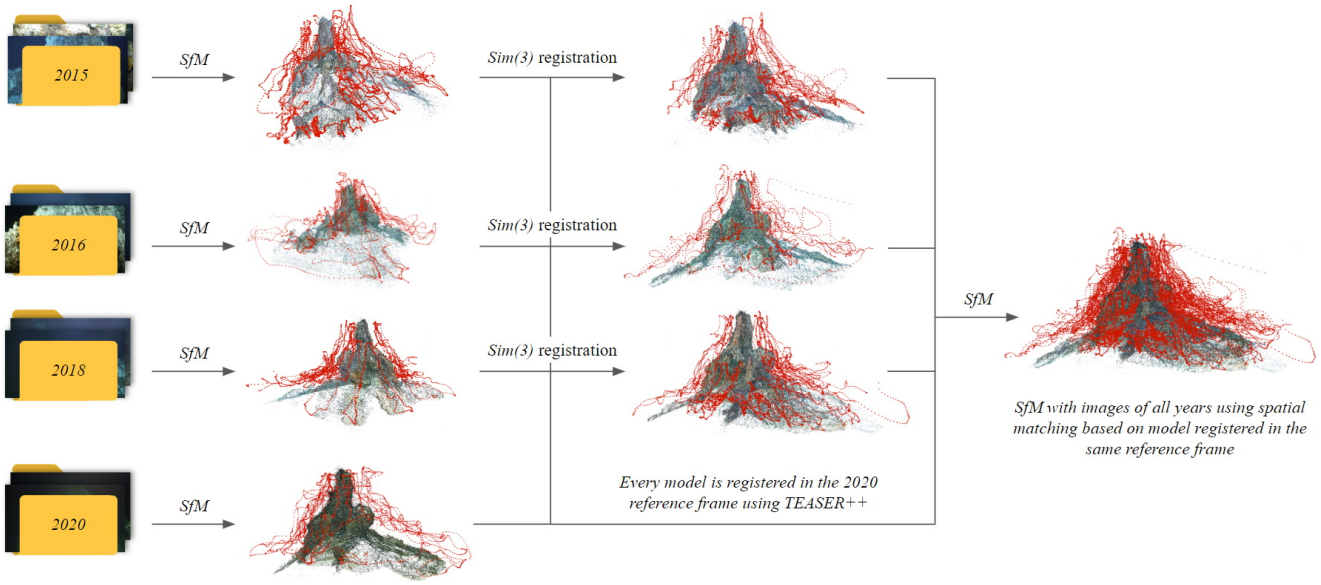


图 4. 用于跨年份匹配图像的运动恢复结构流程。首先针对每一年分别独立构建模型。随后使用 TEASER 将这些模型配准到同一参考坐标系中，即，2020年参考坐标系，并使用迭代最近点++ (Yang 等人，2021) 和迭代最近点 (Zhou 等人，2018)。最后，根据现已处于同一参考坐标系中的各个模型的相机位姿，通过空间匹配计算出一个包含所有年份图像的模型。

重建统计数据。对于每个模型，我们报告配准图像的数量、三角测量三维点的数量、平均轨迹长度（三维点被观测到的图像数量）、每张图像的二维观测平均数量，以及以像素为单位的平均重投影误差。

模型	图像数量	三维点数量	平均轨迹长度	每张图像的平均观测数	平均重投影误差
2015	4,914	525,522	8.48	906.4	1.35 像素
2016	3,699	520,320	5.85	823.5	1.32 px
2018	5,493	618,421	7.09	798.1	1.31 px
2020	3,976	464,331	8.35	975.5	1.33 px
全球	18,082	1,971,726	8.24	898.7	1.39 px

在不同的潜水作业期间，使用 ROV传感器采集了同步视频和导航数据。视频由两台不同的摄像机拍摄，其特性如表1所示。图像通过专门设计的穹顶视窗采集，该视窗配备了校正镜片，以补偿水下折射的影响。由于这些镜片能够在很大程度上补偿空气-玻璃-水介质引起的畸变，因此采用包含 k1 和 k2 畸变系数的二阶径向畸变模型，对水下摄像机进行标定。ROV 配备了一套人工照明系统，由 12 块 LED面板组成，每块面板的光通量均为 20,000 流明。此外，ROV 还集成了一套超短基线声学定位系统（USBL）、一个惯性导航系统（INS）、一个多普勒测速仪（DVL）以及一个深度传感器，并按照Guerrero-Font 等人 (2016年) 的方法对这些数据进行融合，以计算导航数据，从而估算潜水器的位置。该位置由 ROV 的纬度、经度和海拔，以及其偏航角、俯仰角和滚转角组成。然而，由于 USBL 存在不确定性，导航数据仅在每次探访期间保持一致，因为不同探访所对应的帧之间可能出现数米的偏移。

相机位姿与场景几何结构

该视觉定位数据集提供图像及其在统一参考坐标系中表示的六自由度位姿。COLMAP 运动恢复结构（Sch¨onbergerand Frahm 2016）用于获取相机位姿和内参，以及三维场景几何结构。

从每年拍摄的视频中，每隔 3 秒提取一张图像，作为数据集的输入数据。部分图像的边缘带有小型导航叠加层，使用图像修复技术将其去除（Telea 2004）。此外，还人工检查图像并剔除无关图像，例如仅拍摄水柱的图像。

尝试直接对所有年份的图像使用 COLMAP 时，遇到了几个问题。首先，为了创建图像对，COLMAP 依赖词汇树（Sch¨onberger 等，2016年）或 NetVLAD（Arandjelovic 等，2016年）等图像检索方法。虽然这些方法能够成功匹配同一年的图像，但在跨年份图像配对方面表现不佳。其次，为了对图像对进行特征匹配，COLMAP 使用 SIFT描述子。然而，如图3所示，手工设计描述子无法在不同年份的图像之间产生令人满意的匹配结果。

为克服这些问题，我们采用了图 4* 所述的运动恢复结构流程。我们首先为每一年分别构建一个模型

*用于计算所提出的运动恢复结构流程各个步骤的代码可在 github.com/clementinboittiaux/sfm-pipeline 获取。

表 3. 按两张图像拍摄年份统计的配对特征百分比。该表经过归一化处理，每行总和为 100%。表中展示了用于执行运动恢复结构的观测数据跨年份覆盖情况。

观测年份	2015	2016	2018	2020
2015	65.9%	10.7%	15.2%	8.12%
2016	20.3%	50.9%	16.2%	12.6%
2018	16.1%	9.05%	63.1%	11.8%
2020	9.79%	8.00%	13.5%	68.8%

。对于每个独立模型，我们使用导航数据进行图像检索。然而，2015年这次潜水任务中有 3,178 张图像缺少导航数据。在这种情况下，我们使用 NetVLAD（Arandjelovic 等，2016年）进行图像检索。在同一次潜水任务中，NetVLAD 能够有效检索相似图像。我们没有使用手工设计描述子，而是采用 SuperPoint（DeTone等，2018年）和 SuperGlue（Sarlin等，2020年）进行特征匹配。随后，我们使用 TEASER 将各个独立模型生成的三维点云配准到 2020年的点云上++ （Yang 等人，2021），并使用迭代最近点（Zhou 等人，2018）对结果进行优化。通过这种方式，我们在同一参考坐标系中粗略估计了每一年的相机内参和位姿。然后，我们利用这些位姿配对跨年度图像，并使用 SuperPoint 和 SuperGlue 对这些图像进行匹配。最后，我们使用 COLMAP 构建了一个包含所有图像的全球模型。

通过使用 Umeyama 算法（$Sim(3)$），将相机位姿与可用导航数据对齐，获取了每个模型的尺度 Umeyama 1991）。对于包含所有观测数据的全球模型，对齐时仅使用了 2020年 的导航数据。至此，我们获得了每个年份中各视点的 6DOF 位姿，以及处于同一参考坐标系下、具有真实尺度的完整热液喷口 3D 点云。

表2报告了使用 COLMAP 获得的重建统计数据，以说明所提出的真实值的确定性水平。表 3 报告了最终运动恢复结构中各年份之间匹配的三维点百分比。虽然大多数三维点观测都集中在同一年份内，但其中相当一部分在不同年份之间进行了匹配。这确保了不同访问期间场景几何结构和相机位姿的一致性。

数据集格式

该埃菲尔铁塔数据集由潜水作业图像和一个涵盖所有访问记录的全局 COLMAP 模型组成。为了确保可复现性，我们还提供每次访问对应的独立 COLMAP 模型，以及每张图像在可用情况下的插值导航数据。数据集架构详见图 5。

每张图像均以其采集日期命名，格式为 YYYYmmddTHHMMSS.fffZ。其中，YYYY 表示四位数年份，mm 表示月份，dd 表示日期，HH 表示 24 小时制的小时，MM 表示分钟，SS 表示秒，fff 表示毫秒。COLMAP 模型包含场景几何结构、相机内参和位姿。模型由 3 个文件组成：cameras.txt 包含相机内参；images.txt 提供相机位姿，以及

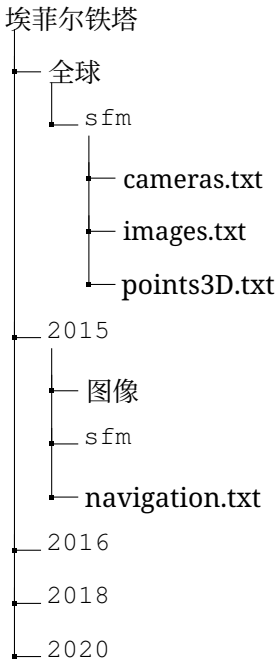


图 5. 数据集文件组织。对于每个年份，均提供图像、导航数据和一个独立的运动恢复结构。另提供一个涵盖所有年份的全球运动恢复结构，用于对视觉定位方法进行基准测试。

每张图像对应的二维关键点列表及其三维观测；points3D.txt 包含三维点的位置和颜色。COLMAP 输出格式的具体说明可在文档[†]中找到。此外，COLMAP 提供了脚本，可轻松读取 C++、Python 和 MATLAB[‡] 中的模型。导航数据采用以下文本格式提供：

image001.png 纬度$_1$ 经度$_1$ 海拔$_1$
image002.png 纬度$_2$ 经度$_2$ 海拔$_2$.....

其中，纬度、经度 和海拔 分别表示载具的纬度、经度和海拔。

跨年变化特征

该数据集记录了多次访问期间出现的诸多外观变化，这些变化给基于视觉的定位任务带来了重大挑战。这些变化性质各异，也就是说，涉及地形、环境或遥控潜水器设备的改动。在这一时期观察并测量到的多项变化，反映了多年来建筑体局部地貌的改变（Van Audenhaege 等人，2023）。烟囱坍塌、露头／巨石脱落或滑塌导致物质流失，而活动喷口区则通过矿物沉积不断扩展，形成新的露头、裙边和尖塔。物质堆积量是流失量的两倍，表明建筑体的体积正随着时间推移而增加。尽管这些变化在局部可能十分剧烈，并影响跨年度三维模型的配准，但它们仅占总表面积的 5%，而且

[†]colmap.github.io/format.html
[‡]github.com/colmap/colmap/blob/dev/scripts

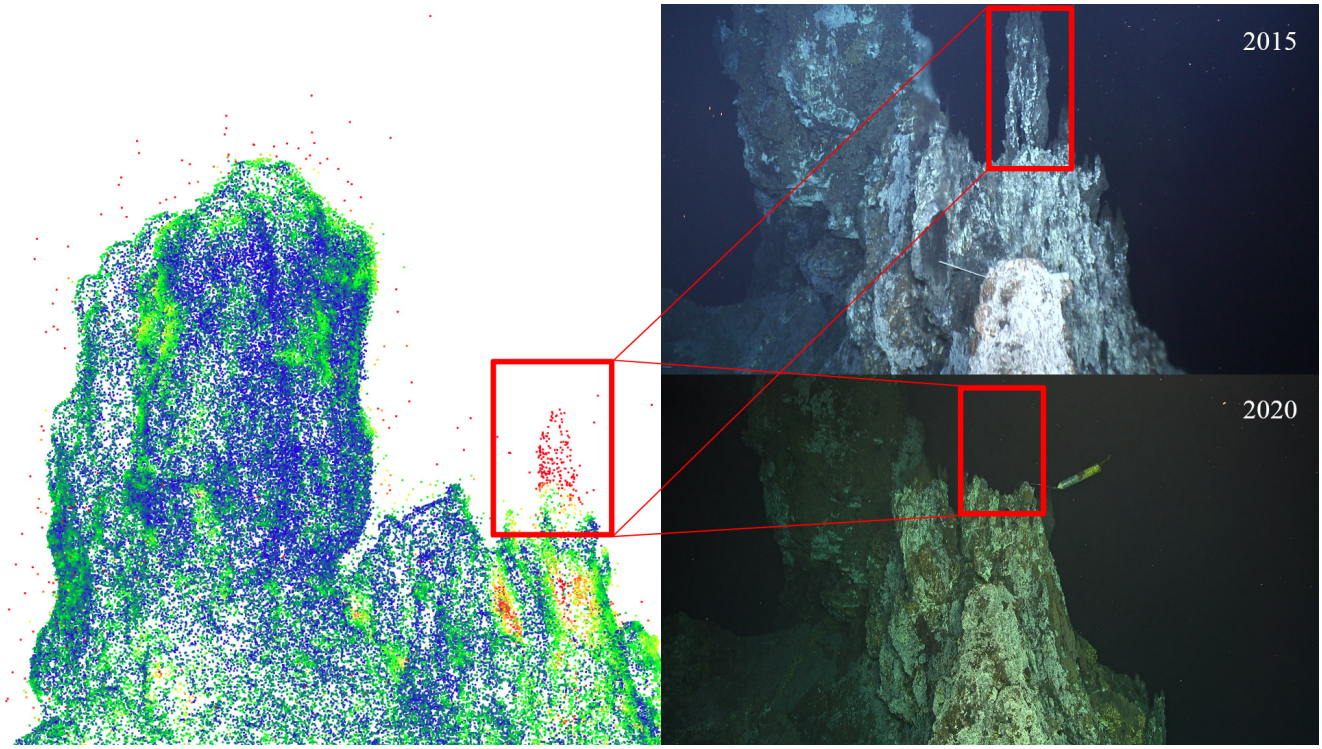


图6。 拓扑变化示意图。左图显示了2015年与2020年模型之间的点云距离。我们注意到，2015年的一部分区域在2020年缺失。这一变化在右侧图像中清晰可见。

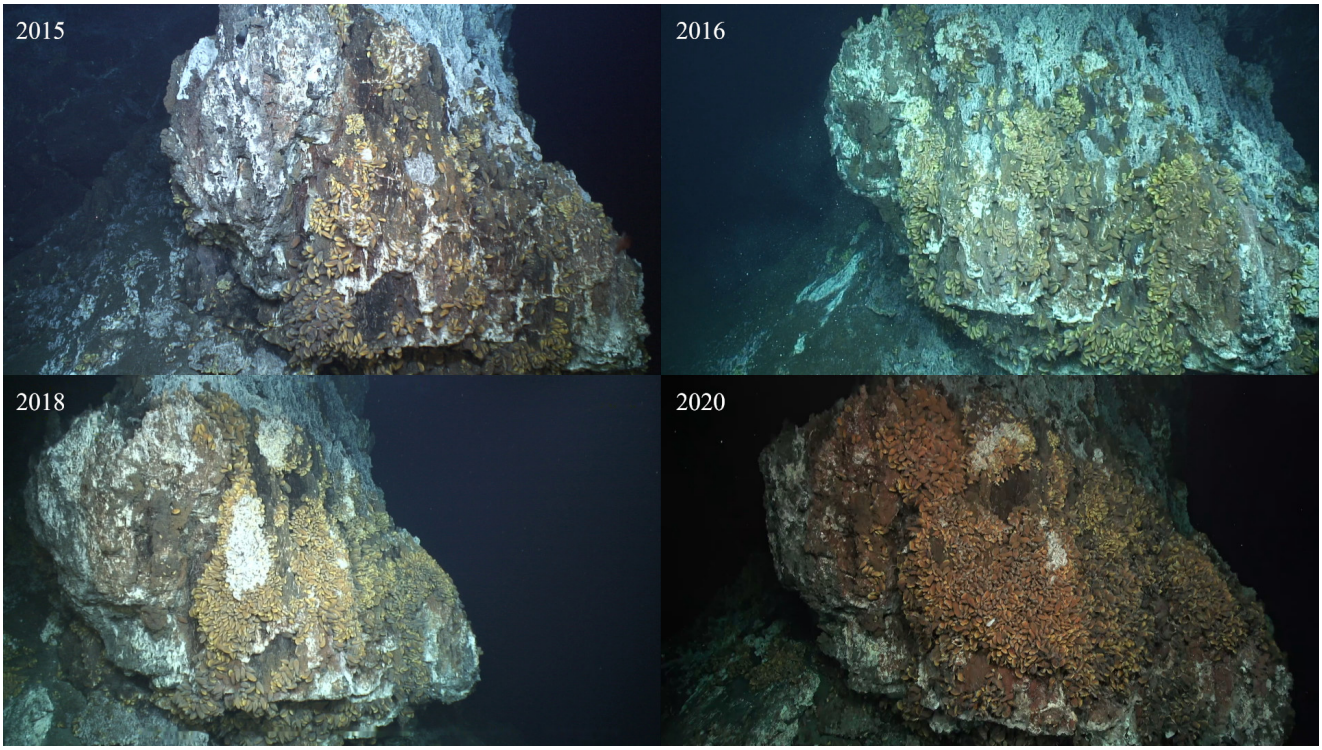
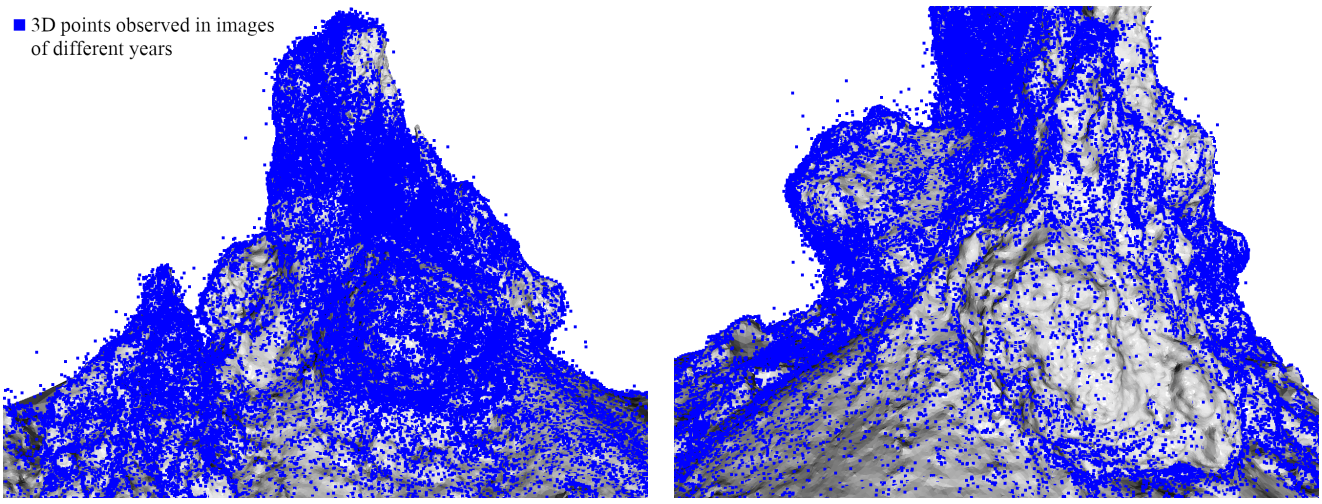


图7。 热液喷口东南侧立面的演变。贻贝种群的增长显著改变了场景的视觉外观，使特定二维点的匹配变得困难。

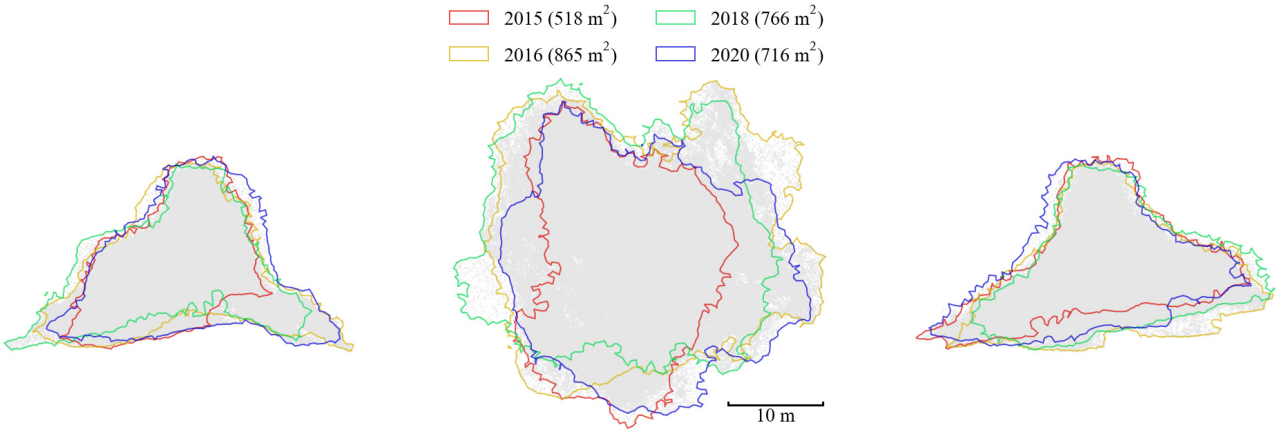
集中在活跃喷口区域。热液活动的局部变化也会导致不同的矿化过程，从而形成沉积物；沉积物的颜色会随流体的温度和化学成分而变化。图6 展示了场景地形的变化。2015年可见的一座烟囱在2020年已不复存在，此外，2015年不存在的温度传感器于2020年部署在附近。

生物变化更为重要，且主要集中在地形变化区域。这些变化源于贻贝种群的生长和迁移，它们逐渐扩散并占据新形成的栖息地（143.97 m^2 从2015年到2020年）（Van Audenhaege 等, 2022）。此外，贻贝还具有动态变化的特征



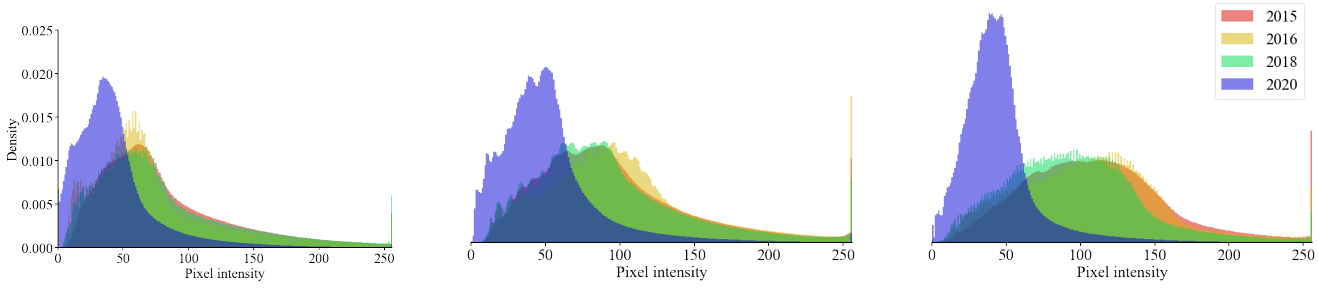
(a)西立面。 **(b)** 东南立面。

图 8. 埃菲尔铁塔建筑结构上，不同年份图像之间经三角测量得到的三维点分布。由于生物变化，东南立面上通过跨年份三角测量得到的三维点较为稀少。



(a)东-上视图。 **(b)**东-北视图。 **(c)**北-上视图。

图9. 遥控潜水器在不同潜水作业期间覆盖的区域。



(a)红色通道。 **(b)**绿色通道。 **(c)**蓝色通道。

图10. 各年份各颜色通道的像素强度直方图对比。

每日重新定向。此外，2015年至2020年期间，整个构造上的白色微生物席总体消失（-72.85 米^2）。尽管这些变化不会影响结构的一般地形，但会显著改变模型的颜色和纹理。图 7 展示了贻贝种群多年来的演变如何改变场景的地形以及热液喷口的颜色。此外，由于这些有机变化，不存在真正匹配的三维点

烟囱不同年份之间的大部分区域都会发生变化。这使得特定二维点的匹配变得困难。图 8 展示了这些生物变化如何影响该建筑体的全球三维重建。虽然热液喷口在不同年份之间总体上匹配良好，但东南立面的某些具体区域会受到这种变异来源的影响，模型主要依赖同一年份图像之间的匹配。

图9 展示了遥控潜水器每年覆盖的区域。我们注意到，该潜水器在不同区域的覆盖情况并不均匀

表4。 M定位误差中位数，以及在给定米和度阈值内完成定位的位姿百分比 .

方法	误差中位数	1 cm, 1°	2 cm, 2°	3 cm, 3°	5 cm, 5°	25 厘米, 2°	50 厘米, 5°	500 厘米, 10°
PoseNet	1.98 米, 10.73°	0.00%	0.00%	0.00%	0.02%	0.24%	3.58%	45.75%
同方差	1.32 米, 6.27°	0.00%	0.00%	0.00%	0.00%	0.79%	10.20%	**62.03%**
单应性	1.23 米, 8.30°	0.00%	0.00%	0.00%	0.02%	0.53%	8.47%	57.83%
hLoc	0.09 米, 1.11°	**15.04%**	**28.37%**	**36.04%**	43.49%	53.87%	**57.94%**	60.07%
PixLoc	6.55 米, 41.00°	0.41%	1.75%	3.50%	6.49%	13.76%	15.16%	18.46%
hLoc+PixLoc	**0.08 米**, **1.10°**	13.41%	28.18%	35.80%	**44.08%**	**53.95%**	**57.94%**	60.07%

。2015年潜水覆盖的范围是所有年份中最小的。

图 10 比较了各年份所有图像中像素强度的直方图。首先，我们观察到，与绿色和蓝色通道相比，红色通道的像素强度整体较低。这很容易用水下环境中波长衰减差异来解释（Akkaynak 和 Treibitz 2019；Ber man 等人 2021）。我们还注意到，2020 年的潜水活动中像素强度发生了变化，这可能是由于相机更换所致。

视觉定位基准

我们根据遥控潜水器每年覆盖的区域划分训练集和测试集。如图 9所示，2016年、2018年和2020年覆盖的总区域几乎包含了2015年覆盖的全部区域。因此，我们选择2016年、2018年和2020年作为训练集，2015年作为测试集。

采用上述训练集/测试集划分，我们使用知名的视觉定位方法对埃菲尔铁塔数据集进行了基准测试：采用不同损失函数的 PoseNet（Kendall 等人，2015年；Kendall 和 Cipolla，2017年；Boittiaux 等人，2022年）、hLoc（Sarlin 等人，2019年）和 PixLoc（Sarlin 等人，2021年）。PoseNet 会针对每个场景训练一个不同的神经网络，而 hLoc 和 PixLoc 则依赖于在陆地数据集上训练的基于深度学习的特征。下面将详细介绍各方法所用的参数。

PoseNet [§]：该网络按照（Kendall 等人 2015年）中的描述重新实现，但将 GoogLeNet 骨干网络替换为更新的 MobileNetV2（Sandler 等人 2018年）。对于 PoseNet 损失，我们采用了 $\beta = 500$，如（Kendall等人2015年）针对室外剑桥数据集所建议的那样。我们按照（Kendall 和 Cipolla 2017年）中的建议初始化同方差损失，即，$\hat{s}_x = 0.0$ 和$\hat{s}_q = -3.0$。对于单应性损失，我们选择局部 x_{min} 和 x_{max}作为 $2.5th$ 和 $97.5th$ 百分位数，如（Boittiaux 等人 2022年）所述。

hLoc：我们使用（Sarlin 等人，2019）中提出的流程，即，使用 NetVLAD 进行图像检索，并使用在室外场景上预训练的 SuperPoint 和 SuperGlue 进行局部匹配。

PixLoc：我们使用了在 MegaDepth 数据集（Li 和 Snavely 2018年）上预训练的网络权重。

hLoc+PixLoc：首先使用 hLoc 检索相机位姿，然后使用 PixLoc 进行细化。该流程由 Sarlin 等人（2021 年）提出。

上述所有方法在该数据集上的结果报告于表4 中，可作为该与其他长期视觉定位方法的比较。与 hLoc 和 Pixloc 不同，基于 PoseNet 的方法是端到端网络。与 Sattler 等人提出的结果一致Sattler 等人 (2019)，端到端网络获得的位姿估计精度最低。此外，由于真实值是通过运动恢复结构（SfM）构建的，因此采用相同工作模式的方法（例如 hLoc）具有优势，因为它们优化的是相同的度量指标（Brachmann 等人 2021）。

hLoc 和 PixLoc 基于使用陆地数据训练的网络，我们可以预期，使用水下数据训练这些网络能取得更好的结果。然而，这仍然是一项挑战，因为所需的数据量超过了水下环境中现有的可用数据量。例如，NetVLAD 使用 Google 街景时间机器的数据进行训练。另一种方法是尽量减少水下环境中的物理现象造成的变化，使图像更接近陆地图像，例如使用 *Sea-thru*（Akkaynak 和 Treibitz 2019年）或 SUCRe（Boittiaux 等人 2023年）。

结论

本文提出了一个新颖的数据集，用于评估深海环境中的视觉定位方法。与现有数据集不同，埃菲尔铁塔 呈现了水下场景的长期变化，例如地形、种群和物种分布、后向散射以及颜色衰减。我们分析了这些变化，并在所提出的数据集上评估了多个定位流程。所得结果可作为未来水下视觉定位系统研究的基准。除用于视觉定位外，该数据集还可用于检测深海环境中场景几何结构的变化。更广泛地说，它也有助于研究水对各种计算机视觉算法的影响。

致谢

所有数据采集工作均由 Ifremer 完成。作者谨向科研船 *Pourquoi Pas?* 和*L'Atalante*的船员、Victor6000 遥控潜水器的驾驶员，以及所有协助采集这些数据的工作人员表示感谢。MM 和 LVA 得到了欧盟地平线2020研究与创新项目 iAtlantic 的支持，该项目资助协议编号为 818123。

[§]提供了包含全部三种位姿回归损失的 PoseNet 实现，网址为 github.com/clementinboittiaux/homography-loss-function。

参考文献

Akkaynak D 和 Treibitz T（2019）Sea-Thru：一种从水下图像中去除水体影响的方法。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第 1682–1691 页。Arandjelovic R、Gronat P、Torii A、Pajdla T 和 Sivic J（2016）NetVLAD：用于弱监督地点识别的 CNN 架构。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。美国拉斯维加斯，第 5297–5307 页。Berman D、Levy D、Avidan S 和 Treibitz T（2021）利用雾霾线和新的定量数据集进行水下单幅图像颜色恢复。*IEEE*《模式分析与机器智能汇刊》论文集43(8)：2822–2837。Boittiaux C、Marxer R、Dune C、Arnaubec A、Ferrera M 和 Hugel V（2023）SUCRe：利用场景结构进行水下颜色恢复。审稿中。Boittiaux C、Marxer R、Dune C、Arnaubec A 和 Hugel V（2022）基于单应性的相机位姿回归损失函数。*IEEE*机器人与自动化快报7(3)：6242–6249。Brachmann E、Humenberger M、Rother C 和 Sattler T（2021）视觉相机重定位中伪真实值的局限性。载于：*IEEE*国际计算机视觉会议（*ICCV*）论文集。第 6198–6208 页。Burnett K、Yoon DJ、Wu Y、Li AZ、Zhang H、Lu S、Qian J、Tseng WK、Lambert A、Leung KY、Schoellig AP 和 Barfoot TD（2023）Boreas：多季节自动驾驶数据集。《国际机器人研究杂志》42(1-2)：33–42。Campos R、Garcia R、Alliez P 和 Yvinec M（2015）一种用于精细水下三维光学测绘的表面重建方法。《国际机器人研究杂志》34(1)：64–89。Cannat M 和 Sarradin PM（2010）MOMARSAT：大西洋中脊监测。DOI：10.18142/130。Cernea D（2020）OpenMVS：多视图立体重建库。DeTone D、Malisiewicz T 和 Rabinovich A（2018）SuperPoint：自监督兴趣点检测与描述。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）研讨会论文集。第 337–33712 页。Ferrera M、Creuze V、Moras J 和 Trouv´e-Peloux P（2019）AQUALOC：用于视觉—惯性—压力定位的水下数据集。《国际机器人研究杂志》38(14)：1549–1559。Girard F、Sarrazin J、Arnaubec A、Cannat M、Sarradin PM、Wheeler B 和 Matabos M（2020）洋流和地形决定活跃热液丘上的群落分布。《海洋学进展》187：102397。Griffith S、Chahine G 和 Pradalier C（2017）Symphony 湖泊数据集。《国际机器人研究杂志》36(11)：1151–1158。Guerrero-Font E、Massot-Campos M、Negre PL、Bonin-Font F 和 Codina GO（2016）一种由 USBL 辅助的野外自主水下航行器多传感器导航系统。载于：*IEEE*智能系统多传感器融合与集成国际会议（*MFI*）。第 430–435 页。Kendall A 和 Cipolla R（2017）用于深度学习相机位姿回归的几何损失函数。载于：论文集

*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第5974–5983页。Kendall A、Grimes M 和 Cipolla R（2015）PoseNet：一种用于实时六自由度相机重新定位的卷积网络。载于：*IEEE*国际计算机视觉会议（*ICCV*）论文集。第2938–2946页。Langmuir C、Charlou JL、Colodner D、Costa I、Desbruyeres D、Desonie D、Emerson T、Fornari D、Fouquet Y、Humphris S、Fiala-Medioni A、Saldanha L、Sours-Page R、Thatcher M、Tivey M、Dover C、Damm K、Wiese K 和 Wilson C（1993）Lucky Strike——亚速尔高原新发现的热液区。海岭事件4(2)：3–5。Li C、Guo C、Ren W、Cong R、Hou J、Kwong S 和 Tao D（2019）水下图像增强基准数据集及其拓展。*IEEE*图像处理汇刊 29：4376–4389。Li Z 和 Snavely N（2018）MegaDepth：从互联网照片中学习单视图深度预测。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第2041–2050页。Mallios A、Vidal E、Campos R 和 Carreras M（2017）水下洞穴声呐数据集。《国际机器人研究杂志》36(12)：1247–1251。Matabos M、Barreyre T、Juniper SK、Cannat M、Kelley D、Alfaro-Lucas JM、Chavagnac V、Colac¸o A、Escartin J、Escobar E、Fornari D、Hasenclever J、Huber JA、La¨es-Huon A、Lant´eri N、Levin LA、Mihaly S、Mittelstaedt E、Pradillon F、Sarradin PM、Sarrazin J、Tomasi B、Venkatesan R 和 Vic C（2022）热液喷口环境多学科观测集成（IMOVE）：深海热液喷口观测站的十年进展。《海洋科学前沿》9。Nielsen MC、Leonhardsen MH 和 Schjølberg I（2019）PoseNet在六自由度水下位姿估计中的评估。载于：*OCEANS2019 MTS/IEEE*。第1–6页。Sandler M、Howard A、Zhu M、Zhmoginov A 和 Chen LC（2018）MobileNetV2：倒置残差与线性瓶颈。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第4510–4520页。Sarlin PE、Cadena C、Siegwart R 和 Dymczyk M（2019）由粗到细：大规模场景下的稳健层次定位。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第12716–12725页。Sarlin PE、DeTone D、Malisiewicz T 和 Rabinovich A（2020）SuperGlue：利用图神经网络学习特征匹配。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第4938–4947页。Sarlin PE、Unagar A、Larsson M、Germain H、Toft C、Larsson V、Pollefeys M、Lepetit V、Hammarstrand L、Kahl F 等（2021）回归特征：从像素到位姿学习稳健的相机定位。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第3247–3257页。Sattler T、Maddern W、Toft C、Torii A、Hammarstrand L、Stenborg E、Safari D、Okutomi M、Pollefeys M、Sivic J 等（2018）变化条件下六自由度室外视觉定位基准测试。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第8601–8610页。

Sattler T、Zhou Q、Pollefeys M 和 Leal-Taixe L（2019）理解基于 CNN 的绝对相机位姿回归的局限性。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第 3302–3312 页。Sch¨onberger JL 和 Frahm JM（2016年）重新审视运动恢复结构。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第 4104–4113 页。Sch¨onberger JL、Price T、Sattler T、Frahm JM 和 Pollefeys M（2016年）用于图像检索快速空间验证的投票与验证策略。载于：亚洲计算机视觉会议。Springer，第 321–337 页。Shotton J、Glocker B、Zach C、Izadi S、Criminisi A 和 Fitzgibbon A（2013）用于 RGB-D 图像中相机重新定位的场景坐标回归森林。载于：*IEEE*计算机视觉与模式识别会议（*CVPR*）论文集。第 2930–2937 页。Telea A（2004）一种基于快速行进法的图像修复技术。图形工具期刊9(1)：23–34。Umeyama S（1991）两个点模式间变换参数的最小二乘估计。*IEEE*模式分析与机器智能汇刊论文集13(4)：376–380。Valentin J、Dai A、Niessner M、Kohli P、Torr P、Izadi S 和 Keskin C（2016年）学习在能量景观中导航。载于：*2016*年第四届国际三维视觉会议（*3DV*）。第 323–332 页。Van Audenhaege L、Matabos M、Brind'Amour A、Drugmand J、La¨es-Huon A、Sarradin PM 和 Sarrazin J（2022）长期监测揭示大西洋中脊热液喷口贻贝群落前所未有的稳定性。海洋学进展204：102791。Van Audenhaege L、Sarrazin J、Legendre P、Perrois G、Cannat M 和 Matabos M（2023）复杂热液结构上的生态动态监测：一种新型摄影测量方法揭示了动物群落变异的精细尺度。审稿中。Yang H、Shi J 和 Carlone L（2021）TEASER：快速且可认证的点云配准。*IEEE*机器人学汇刊37(2)：314–333。Zhou QY、Park J 和 Koltun V（2018年）Open3D：用于三维数据处理的现代库。*arXiv:1801.09847* 。