

AI4MARS:火星地形感知自动驾驶数据集

R. Michael Swan、Deegan Atha、Henry A. Leopold、Matthew Gildner、Stephanie Oij
Cindy Chiu 和 Masahiro Ono

加州理工学院喷气推进实验室

美国加利福尼亚州帕萨迪纳 91109

robert.m.swan@jpl.nasa.gov; deegan.j.atha@jpl.nasa.gov; henry.leopold@jpl.nasa.gov;

stephanie.l.oij@jpl.nasa.gov; matthew.gildner@jpl.nasa.gov; cindy.h.chiu@jpl.nasa.gov;

masahiro.ono@jpl.nasa.gov

抽象的

深度学习已迅速成为地球上自动驾驶汽车的必需品。相比之下,火星上的自动驾驶车辆,包括计划于 2021 年 2 月登陆火星的美国宇航局最新的火星车 Perseverance,

仍然由经典的机器视觉系统驱动。深的学习能力,例如语义分割和物体识别,将大大有利于安全以及正在进行和未来的红色任务的生产力行星。为此,我们创建了第一个大规模数据集, AI4Mars,用于训练和验证地形分类 Mars 的模型,由来自 Curiosity、Opportunity 和 Spirit 漫游者的 35K 图像上的约 326K 语义分割完整图像标签组成,通过众包收集。

每张图片都由 10 个人标记,以确保更多众包标签的质量和协议。它也是包括约 1.5K 由流动站注释的验证标签来自 NASA 的 MSL (火星科学)的规划者和科学家实验室)任务,操作好奇号火星车,和 MER (火星探索漫游者)任务,该任务运行精神和机遇漫游者。我们训练了一个 AI4Mars 训练数据集上的 DeepLabv3 模型和整体分类准确率达到 96% 以上测试集。数据集已公开。^{1 2}

一、简介

美国宇航局的火星探测器拥有自主十多年的驾驶能力,称为 AutoNav [3]于 2004 年开始探索红色世界的“勇气号”和“机遇号”漫游车,以及后来的“好奇号”漫游者,于 2012 年登陆。

¹©2021。加州理工学院。政府赞助承认。

²<https://data.nasa.gov/d/cyxx-2qix>

AutoNav 是为最新的火星车 Perseverance 实施的,在撰写本文时它正在前往火星的途中

[18, 26]。尽管如此,它的感知系统仍然纯粹基于经典机器视觉算法,包括立体用于 3D 重建和障碍物检测的匹配,以及用于状态估计的视觉里程计[26]。这意味着 AutoNav 仅评估地形的可穿越性基于几何信息。然而,像车辆地球,我们凭经验知道地形类型对可穿越性有重大影响。例如,圣灵

流动站被沙坑固定;好奇号火星车,同样,在沙质表面上行驶时几乎嵌入在隐谷;好奇号的铝轮是当它在坚硬表面上的尖石上行驶时被刺穿。

事实上,地球上的人类漫游车司机严重依赖来自流动站图像的语义信息以规划路径。人类漫游车司机手动计划的驾驶是能够使用有关地形类型的语义信息在更长的距离和具有挑战性的地形上行驶否则会在使用 AutoNav 时导致故障并危及流动站的硬件。出于这个原因,绝大多数火星探测器在现有任务中的驾驶都是执行的

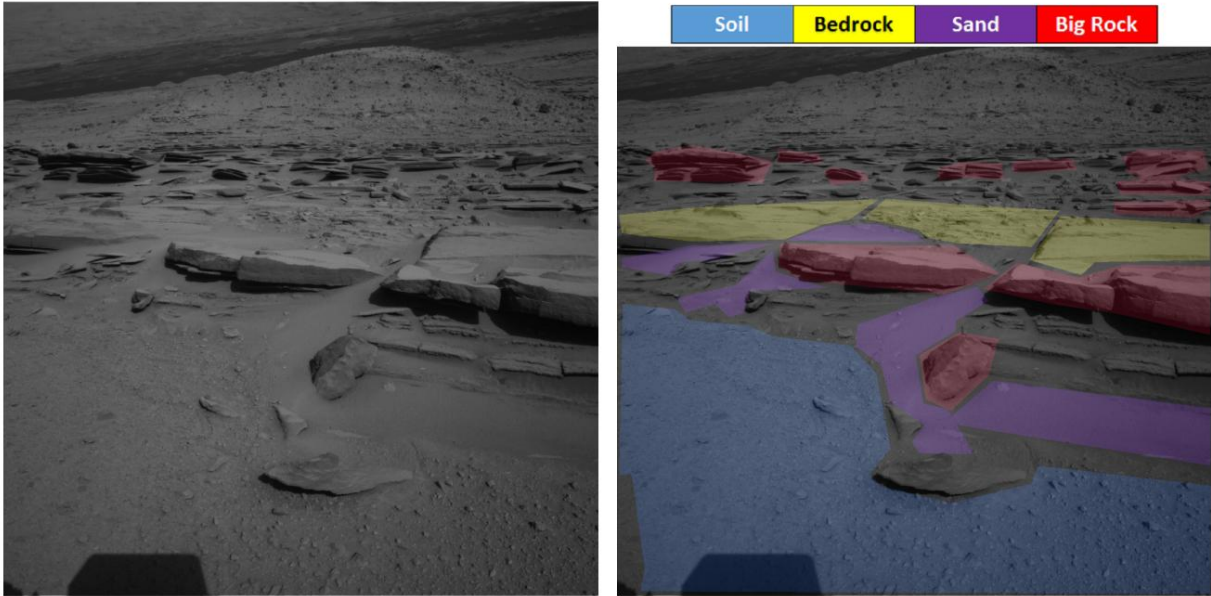
通过人工进行地形评估和路径规划,而 AutoNav 的使用仅限于无法手动控制行驶路径的情况

[19]。

如果有能力识别船上的地形类型,漫游者将能够预测滑移[8],计划

最小化驱动能量[12, 17]或定位误差[13]的路径(两者都取决于地形类型),以及自主识别科学观测的有希望的目标[17]。这种地形感知的基础

自主性是从机上对地形进行分类的能力图片。出于这个原因,罗斯洛克等人。开发了一个基于机器学习的火星地形分类器命名 SPOC (土壤性质和对象分类)[20],其中已部署在Cu-的地面操作系统上



(a) 原始 MSL NAVCAM 图像。 (b) 高等级 AI4MARS 标签。

图 1:金伯利地区,由好奇号的 NAVCAM 在 Sol 574 上拍摄 (2014 年 3 月 18 日)。图片来源:NASA/JPL

iosity 并在地球上用测试漫游车进行了现场测试。尽管这些地面部署的结果是高度有希望,进一步提升 SPOC 的可靠性水平并满足对板载算法的高要求标准,需要大规模、高质量的标记数据集。

毫不奇怪,地球是唯一可供人类使用的用于深度学习的各种公共大规模数据集的星球。此前,我们要求内部专家

创建火星训练数据集,但标签的数量我们能收集到的东西非常有限 (最多几千个) 由于专家的可用性。最近,欧空局的 LabelMars 项目成功进行了众包工作 并从 Spirit,Opportunity 和 Curiosity [23] 收集了 5,000 张图像的语义分割标签。他们的标签涉及约 20 种地质/地貌地形类别,例如“结核/结核”、“深色岩浆露头”和“浅色沉积浮岩”。

这种分类只能由训练有素的专家解释,因此对将数据集扩展到

高度准确的深度学习模型所需的体积,其准确性似乎与数据集成对数大小[24]。这与地面数据集形成鲜明对比 自动驾驶,例如 KITTI [10]、Cityscapes [7]、nuScenes [5]、Berkeley DeepDrive [27] 和 Oxford Robot Car [14],它们主要由常见的物体组成 可由非专家解释,例如人、动物、车辆和路标。

我们的主要观察结果是,虽然 LabelMars 中的地貌地形分类对于

行星科学,一个更简单和直观的分类,如“沙子”、“岩石”和“土壤”,足以满足 火星探测器的可穿越性评估,包括 预测打滑、驱动能量和车轮磨损。简化的地形分类允许公众

以最少的培训参与众包, 可以通过一个简短的基于网络的教程来提供。同时 时间,高质量的数据集对于系统验证是必要的,以满足太空任务的严格要求。 出于这个原因,我们还收集了一个小得多的 (~1.5K 标签)用于模型测试的保留数据集,已标记 由经验丰富的 Mars 组成的大约十位领域专家 MSL 和 MER 任务的漫游车司机和项目科学家。总而言之,AI4Mars 的 策展方式 数据集是通过众包收集一个大规模的、有针对性的数据集进行训练, 结合高质量、 用于测试的专家标记数据集。

2. 相关作品

我们不是第一个利用众包来收集行星图像上的标签的人。有很多令人兴奋的 为在轨道图像上寻找感兴趣的科学特征而创建的太空公民科学项目。 例如, 月球动物园[4]动员公民科学家识别和 在获得的高分辨率月球图像上描述陨石坑的撞击位置和其他地质特征

来自 NASA 的月球勘测轨道器 (LRO) 航天器。该项目成功使用了一个名为 Zooniverse 的在线公民科学平台,该平台允许基于网络的

公众对图像的注释。至于火星，
第四行星[2]项目成功地使用公民科学来识别一些感兴趣的科学特征

由火星勘测轨道器拍摄的火星南极地区的 221 张高分辨率图像

（维修保养）。第四行星专注于直观可识别的特征,例如扇形和斑点,这些特征表明火星上的海季变化。COSMIC 项目[9]开发

一种检测火星表面变化的算法，
例如新的撞击坑和雪崩。它还利用了
Zooniverse 从公民科学家那里收集训练数据集。这些项目取得了巨大成功，
部分原因是在十个展品的轨道图像上的宏观科学特征展示了可以为非专家直观地解释的独特模式,例如“瑞士奶酪地形”[25]

和“蜘蛛通道”[2]。

相比之下,行星表面的公民科学项目
尽管有大量在月球和火星表面拍摄的图像,但图像相对罕见。

这可能是因为原位地质学通常需要具有专业知识的分析,例如解释

地层学和岩石类型分类。尽善尽美
据我们所知,LabelMars [23]是公民科学家对行星原位图像的唯一大规模标签收集工作。使用他们自己的网站（www.labelmars.net），

它在从火星探测器（例如好奇号）获得的五千张图像上积累了地形分类标签

和 Spirit,计划使用图像扩展数据集
来自机遇号和毅力号漫游车。他们的标签涉及约 20 种地质/地貌地形类别,例如“结核/结核”、“深色岩浆露头”和“浅色沉积浮岩”。

由于这种分类需要专业知识,因此标签
从合格的专家那里收集,主要包括
地质学专业的本科生。

虽然 AI4Mars 使用与 LabelMars 相同的数据源，
我们的数据集在两个方面与 LabelMars 不同:i)
标签数量大于两个数量级
LabelMars（我们目前已经收集到 326K 个标签），
ii) 我们采用了一个更简单的、四向分类的地形:沙子、土壤、基岩和大岩石，

非专家也能直观理解
为评估火星的可穿越性提供了丰富的信息
流浪者。数据集的目的也不同于
上面提到的数据集,因为 AI4Mars 不是用于平面科学、地质学或地貌学研究，
而是用于
训练深度学习模型以实现安全的自动驾驶
火星。据我们所知,这项工作第一个
成功收集超过 10 万条语义分割
在任何天体表面拍摄的图像的标签
除了地球。

数据集	图像训练测试
MSL NAVCAM	17K 160K 943
MSL 桅杆摄像机	9K 82K 待定
机会 NAVCAM 6K 54K 精神 NAVCAM 3K 30K	573
总计（合并）	35K 326K 1.5K

表 1:数据集摘要。MER（即精神和机遇）标签使用联合测试集,因为流动站是相同的。

3.3节中描述的合并标签计数等于
图像的数量。Mastcam 测试集的生产是
未来的工作。

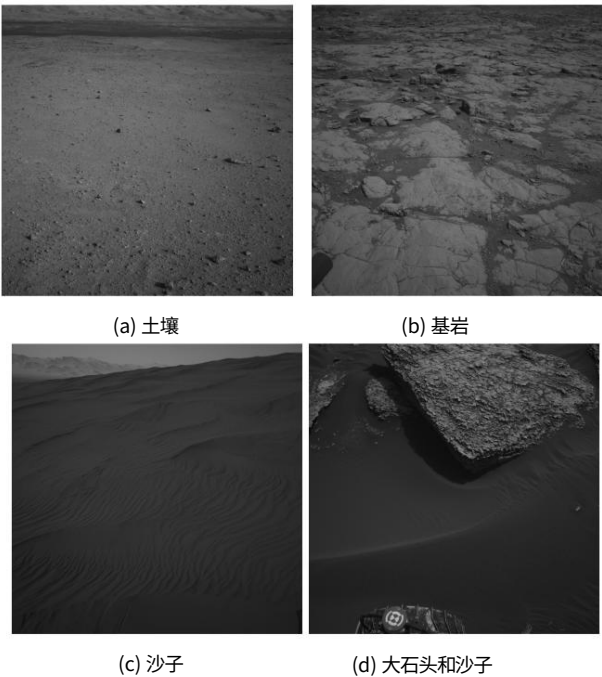


图 2:每个类的代表性示例。注意
几乎没有大岩石本身的图像。

3. AI4Mars 数据集

AI4Mars 数据集包括在火星表面拍摄的大部分现有高分辨率图像。

它来自行星数据系统3 (PDS) 的约 35K 图像组成,涵盖灰度导航相机 (NAVCAM) 和彩色桅杆相机 (Mastcam)

好奇号 (MSL) 的图像,以及灰度
来自 Opportunity (MER) 和 Spirit 的 NAVCAM 图像 (MER) 火星探测器,如表1所示（注意数据集
计数与 2020 年 11 月的初始版本有关）。数据集中值得注意的排除是来自火星的显微图像
手持镜头成像仪 (MAHLI) 和望远镜图像来自

3 <https://pds-imaging.jpl.nasa.gov/index.html>

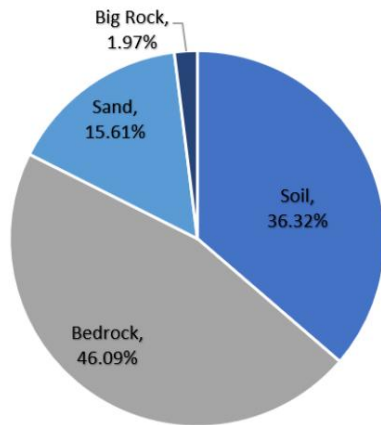


图 3:按类别划分的 MSL NAVCAM 标签组成。

化学和相机 (ChemCam) 因为这些图像通常对可遍历性评估没有帮助。这来自安装的鱼眼避障相机的图像因为 NAVCAM 也不包括在流动站的身体上和 MASTCAM 图像更适合地形分类,因为它具有更高的安装点和更好的角度解析度。全景相机 (Pancam) 的图像精神和机遇计划被纳入未来。由于每张图片都收集了多个标签,标签的数量大大超过了实际的图片数量;这将在第3.1节中进一步讨论。这

数据集分为四种不同的标签类型,它们是:土壤、基岩、沙子和大岩石。与每个相关的结果这些类在第4节中讨论。数据集还提供流动站时间和位置数据、相机参数和可用的深度数据。

深度全部提供 MSL NAVCAM 图像和许多 MER NAVCAM 图像具有相关的立体范围数据。图像的立体测距误差线性增加为距离的函数。对于 1-30m 的距离,MSL 如[16]中所述,NAVCAM 的误差为 0.0005-0.4m。对于 1-30m 的相同范围,MER NAVCAM 图像如[15]中所述,误差为 0.001-0.5m。

3.1.收藏

训练集我们涉及的标签收集方法从不同的标注器中为同一张图像收集 3 个或更多的完整图像标签。一个“标签”将包含所有在单个图像中识别的区域和类别。93% 图像有 9 个或更多相关标签。这张单图多标签方法确保所有图像都可能包含至少一个质量标签,并且可以使用额外的标签对于一些模型训练实验 (更多讨论见

第4.1节)。

标签主要来自志愿者公民 Zooniverse 上的4位科学家获得了简明扼要的完整的基于网络的培训。当贴标机登录时 Zooiverse 项目页面,向他们展示了图像随机顺序并要求提供标签每个图像。这完全是自愿的;没有义务或补偿。收集的标签去了通过算法和人工验收审查 (第3.2),然后是标签合并 (第3.3节)以增强数据集的整体质量。受此训练的模型数据集导致 96% 的整体准确率,评估标签合并后的专家标记测试集,如所述在第4节中详细说明。

测试集一组专家标注人员额外创建了一个小型“黄金标准”测试集,以根据可信参考评估模型性能,如

以及与看不见的最终模型性能进行比较数据。由于高度有限的可用性和领域专家的高劳动力成本,

测试集需要很小,以便可以在合理的时间内对其进行标记。我们针对大约 1% 的训练集图像的测试集。为确保测试

正确设置表示包含的地形的多样性在训练集中,我们从分布在所有主要地形类别上的位置对图像进行采样从轨道图像,如表2所示。与标记中采用的四路局部分类不同,这些或位地形类代表大尺度地质单元,

每个都有独特的成分和外观四个局部地形等级。这个轨道地形测绘由 MSL 项目科学家手动生成[1]重点是与相关的地质地形类别好奇号穿越路径附近的漫游车机动性。轨道地形在任务期间,类别与流动站位置相关。从这些位置,大约相同的数量

从代表每个轨道地形类别的位置中选择了独特的 NAVCAM 图像。测试集然后手动修剪低质量的图像 (通常在黄昏时获得)、高度相似的图像以及那些拥有大量流动站硬件。这导致了一个 Curiosity 的 NAVCAM 图像共有 323 张图像和 205 张 Spirit/Opportunity 的 NAVCAM 图像 (表1中较大的“测试”数字是合并过程之前的原始计数 3.3节中讨论)。

每个轨道地形类别分配了 3 个专家标注员,标注员分配到不超过 2 个类别。给专家贴标签者同样的一般分类作为 Zooniverse 参与者的指导并进行了一些修改。专家贴标机只生成高标签图像中所有地形的置信度和标签覆盖率

轨道地形类	数量 NAVCAM 图像
平坦的地形	38
脊状地形	45
坑坑洼洼的地形	57
高度解剖的地形	57
破碎的地形	41
沙坑和波纹场	46
沙丘	39

表 2:NAVCAM 测试集图像的分布
轨道地形类。每个图像都标记了多个次。

没有被视为优先事项。这意味着一些图像只标记了一小部分地形。标签由专家单独执行,这意味着为特定图像收集的一组标签可能不同于专家对专家。

3.1.1 预处理

在标记图像之前,对图像进行了一些预处理,以使标记者更容易提供

质量标签。图4显示了一个预处理的示例图片。中间的白色梯形标记用作比例尺,其宽度始终为 50 厘米好奇心图像和 20 厘米的精神/机会图像。这些尺寸与 Big Rock 类特别相关,被定义为站立超过30厘米的岩石高,宽度至少 50 厘米。这个类定义之所以引入,是因为高于 30 厘米的岩石被认为有可能让漫游车驶过。教程指出,标注者可能会忽略小于

比例尺的宽度。请注意,图的顶部4变暗;它表示图像的一部分距离摄像机超过 30 m。贴标机被指示忽略超过 30 m 的特征。选择这个距离是为了确保只有更近和更清晰的特征自主所需的将被训练。预处理是使用图像的范围乘积执行的,这些乘积是从图像的立体处理创建的 (该数据集中的所有全分辨率 NAVCAM 和许多

的 Mastcam 图像是立体对)。

3.1.2 贴标机培训

AI4Mars 数据集的分类类别为旨在让那些年仅四年级的学生可以使用基于 Web 的快速教程正确标记。这首次打开 la beling 网站时会自动显示的教程,提供 la beling 的基本指南

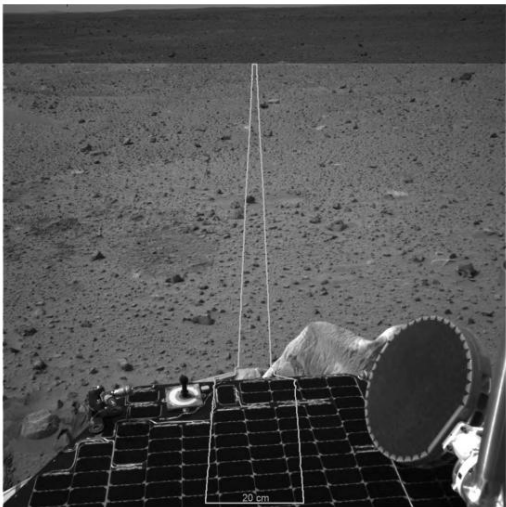


图 4:呈现的预处理图像示例
给 Zooniverse 上的贴标机;图像是由圣灵拍摄的 Sol 9 (火星着陆后的第 9 天)上的漫游车。

beling (例如,标签之间没有重叠)和直观用例子解释了四个地形类别。前启动公民科学项目,我们进行了 beta 测试,138 名志愿者参与,以获取反馈教程和用户界面。对于专业贴标者,在除了基于网络的教程外,我们还举行了一系列会议,以就我们的标签政策达成共识并确保测试集的一致性。

3.1.3 后处理

标记图像后,应用蒙版去除与流动站本身重叠的任何注释 (如果它在图像中)和覆盖区域的注释超过 30 m。

3.2.清洁和标签验收

为了提高训练集的整体质量,我们对公民科学家提交的标签进行了算法和人工审查。包含的标签

后处理后没有有效的注释被抛出。将每个标签与同一图像的其他标签进行比较,以提供基于

众所周知的 Jaccard 指数/mIoU 在公式1 中注明。

J(A, B) =

$|A \cap B|$

$|A \cup B|$

(1)

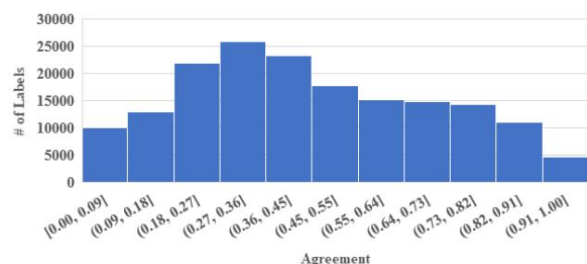
我们将图像的所有标签集定义为 L 并将这些标签的所有成对组合 (无需替换) 定义为 C 根据公式2。

C =

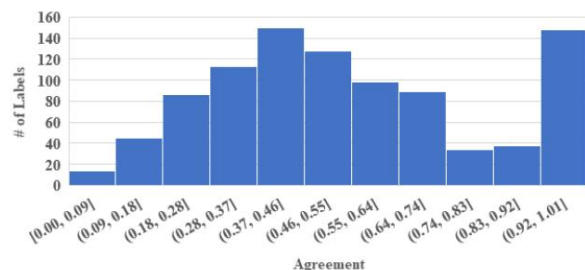
大号

2

(2)



(a) MSL NAVCAM 训练集, 由公民科学家标记



(b) MSL NAVCAM 测试集, 由领域专家标记

图 5: 比较粗训练集和细粒度测试集的一致性分数的直方图。

然后将某个标签 $i \in L$ 的一致性分数定义为

C 中所有对的 Jaccard 指数的平均值, 包含

根据公式 3 和 4 标签 i 和任何其他标签。例如, 给定 3 个标签, 我们希望在

其中找到协议

标签 1 的得分, 我们找到所有包含的组合对

标签 1: (1,2), (1,3)。然后我们找到对的 Jaccard 指数

(1,2) 和对 (1,3)。这些数字相加并

然后除以对的总数 (2)。

$$C_i = \frac{|\{(x, y) \in C \mid x = i\}|}{|C|} \quad (3)$$

$$A(C_i) = J(C_i) \quad (4)$$

然后, 我们会优先审查具有低一致性分数的标签 (所有标签分布的底部 20%)

团队; 初步调查表明, 低一致性

分数与差评相关。团队审查

通过提供关于质量的 1-5 评级来标记

每个标签的。评级为 1 的标签将被丢弃。剩余评级的可能用法将在第 1 节中讨论

5.

测试集的一致性分数直方图和

图 5 所示的训练集提供了对测试集中标签之间一致性的定量审查。基于

关于粗训练标签和细训练标签之间的显著转变

测试标签, 我们假设更高的一致性分数是

与更高质量的标签密切相关。有可能的

其他项目的独特问题是拥有更多

标签比数据可能能够利用协议作为

质量指标。

3.3. 标签合并

给定图像在数据之后仍然有多个质量标签

清洁和验收, 有一些关于如何

应该使用剩余的标签, 其中大部分在第 4.1 节和第 5 节中讨论。一种方法考虑

在他们同意的地方合并现有标签的想法

其他标签留空 (大多数标签都是稀疏的故意标注)。为此, 一个合并的

标签数据集

提供了使用多数规则来确定哪个

要包含的注释或注释的一部分。

为了将单个图像的多个标签合并为单个标签, 使用了两个标准。第一个是为了

要被接受的单个像素, 该像素最常见的标签类别必须由至少三个标记

不同的贴标机。第二个是对于每个像素,

接受的班级必须有超过 65% 的同意

像素的标签总数。请注意, 像素级

这里提到的“协议”是指逐像素标签重叠,

这与中定义的“协议分数”不同

第 3.2 节。此计算忽略未标记的像素。例如, Pixel X 在图像内

由 10 个贴标机贴上标签。细分为: 2 个未标记、6 个土壤、1 个沙子和 1 个基岩标签。合并的

Pixel X 的类将是土壤, 因为它收到了 6 个标签和一个

75% 同意。作为另一个例子, 像素 Y 在

由 9 个标注者注释的图像。细分为: 3

未标记、3 个土壤、2 个沙子和 1 个基岩标签。有

3 个土壤标签, 但只有 50% 的一致性, 所以 Pixel Y

将在合并的注释中未标记。

对于黄金标准测试集, 每个图像都经过注释

由三位不同的专家贴标师。由于标记器数量减少和标签置信度增加, 使用了不同的合并参数。三种不同的金

生成标准测试集。这三个都需要 100%

同意特定像素的标签, 但集合不同

基于所需的最小标记像素数

接受一个标签, 每个像素使用 1、2 和 3 个标签。这

包含合并标签的黄金测试集

每个像素 3 个标签和 100% 的准确度是最有信心的集合, 模型的准确度应该是最好的。

然而, 这也将是最稀疏的测试集, 因此

其他集合有助于更好地理解

在图像的不太置信区域的表现。

专家之间的像素级协议总结

MSL 测试集中的每个图像如图 6 所示。

比如比例较高的 Big Rock 和 Bedrock

仅在两个标签或一个标签之间达成一致 (没有一致) 的像素表明类别的定义或标记它们的方法有一些

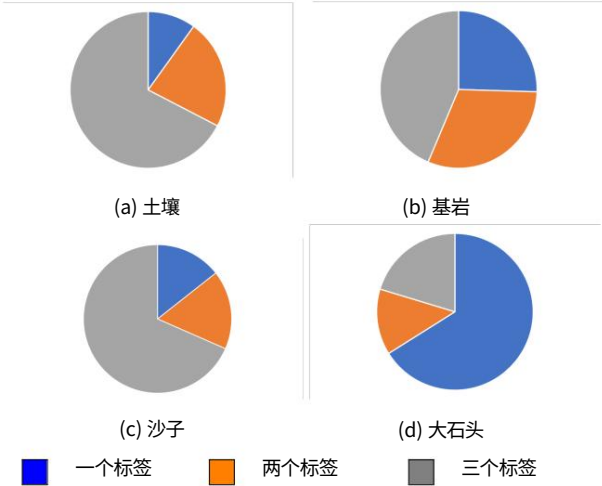


图 6:专家数量占比细分
标签器将像素标记为 MSL 测试集的特定类别。

与其他类相比,固有的模糊性。

4. 实验

对于我们所有的实验,我们都使用了 DeepLabv3+ 在 ImageNet [6, 11, 21]。之所以选择 DeepLab,是因为其代码库的成熟度和它所维护的最先进的语义分割性能。训练是在配备两个 NVIDIA GeForce GTX TITAN X 或两个 NVIDIA 的机器上完成的

特斯拉 P100 GPU。图片从 1024x1024 调整大小 像素到 513x513 以匹配预训练模型的设置。批量大小选择为

可能在遇到 DeepLab 文档建议的 GPU 内存问题之前。

这里显示的所有实验都是使用 MSL 数据完成的。这个数据集最大,最先完成,所以有 更多的时间进行分析。超参数是通过实验确定的;使用相同的超参数

除标签权重外的所有实验。选择标签权重为 1 - 组合,其中

组成是指每个班级所占的百分比 训练集。所有 MSL NAV CAM 标签的组成编号如图 3 所示。在我们的实验中,

发现没有使用标签重量或标签不正确 权重会导致模型收敛速度变慢,但在这些情况下增加步数会导致类似的结果

作为使用 前面提到的加权方法。 鉴于所有指标都可能有更好的表现 进一步测试。这里提到的实验旨在 作为改进的基准。数个可能 这种方法的变化可以提供改进的 性能将在第 5 节中进一步讨论。

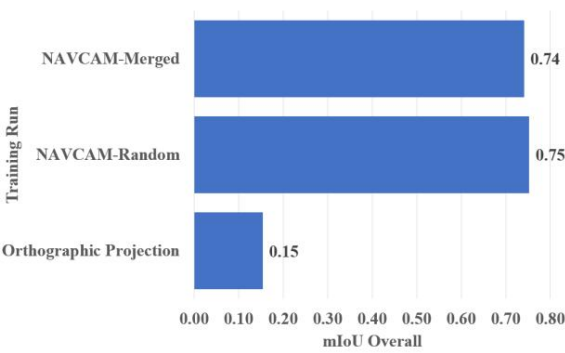


图 7:针对随机计算的总体 mIoU 分数
MSL 数据变体的验证集。 NAVCAM
变体在第 4.1 节中定义

	预料到的			
	土壤	基岩	沙	大岩石
土壤	96.00	0.31	6.15	3.69
基岩	0.25	90.87	2.54	0.44
沙	0.25	3.23	96.51	0.01
大岩石	11.67	0.03	5.48	82.83

表 3:根据 3 标签一致性测试集计算的 MSL NAVCAM-随机混淆矩阵百分比
年龄。
总体准确率为 94.97%。

	预料到的			
	土壤	基岩	沙	大岩石
土壤	99.10	0.32	0.57	0.01
基岩	3.64	94.90	0.88	0.37
沙	0.88	0.37	93.45	0.05
大岩石	6.76	0.05	0.05	93.24

表 4:根据 3 标签一致性测试计算的 MSL NAVCAM-Merged 混淆矩阵百分比
年龄。
整体准确率为 96.67%

4.1. 标签合并与随机选择

检查标签合并与 随机选择可用的标签,一些实验是用合并的数据完成的,如

第 3.3 节和随机选择的数据 (例如给定图像 有 10 个标签,我们随机选择其中一个来使用)。这 我们能够为每种方法实现的最佳结果是 如图 7 和表 4 和 3 所示。初步结果表明 标签合并的明显好处,但修改后的结果 修复实施错误尚无定论。我们提供 他们的完整性并讨论可能的改进 在第 5 节中。

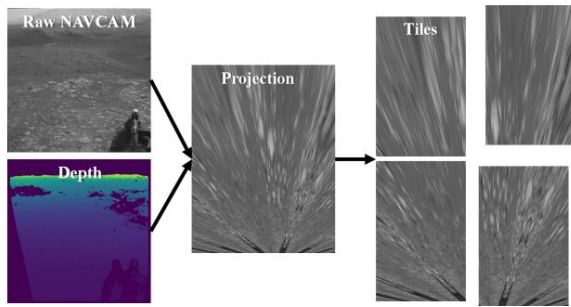


图 8:投影和平铺图像所需的过程。

深度既用于投影,也用于掩盖远处区域。在这种情况下,原始图像的上半部分在投影前被移除。显示的平铺被夸大了;许多图像需要多达 100 个图块。

4.2.平铺和正交投影

将深度数据集集成到训练模型和提高我们在 Big Rock 类上的性能的一个想法是利用正交投影,这是计算机图形学领域中一种成熟的技术[22]。

这种方法在[20]中被使用,并在这项工作中使用几乎相同的投影和平铺实现进行了复制。这个过程总结在图8中。

如图7所示,我们能够达到的 mIoU 分数客观上很差;目前尚不清楚这种糟糕表现的原因是我们的实施问题还是方法本身的问题。我们假设用于平铺的方法平均会导致过多的图像上下文丢失,因此该模型无法始终如一地确定相似类别之间的差异。

5. 讨论和未来工作

未来还有很多工作可以改进,以改进使用该数据集进行的实验以及在其他行星上的自主努力。本文提供了当前数据集的快照。我们计划继续对数据集进行改进和添加,包括来自新火星探测器 Perseverance 的图像。

我们相信,通过使用现有的深度数据(如果可用)或使用单独的岩石实例分类器,可以大大提高数据集中包含的 Big Rock 类的性能。其他工作正在进行中,通过实例检测器识别岩面,然后使用立体数据来估计岩石高度。将来使用这种模型方法,可以训练一种三类语义分割方法,其中岩石检测器填充大岩石类。此外,可以探索全景分割模型来结合这两个网络。

我们在3.3和4.1节中使用的标签合并方法并不是处理额外标签的唯一方法,目前还不清楚标签合并的确切好处是什么。

标签度量的使用可用于创建置信度

神经网络可以集成的标签得分,以提高训练准确性。可能使用的几个指标是第 3.2 节中提到的标签评级和计算的一致性分数。

未来工作的另一个途径是利用由该数据集训练的自动地形分类器。例如,我们凭经验知道滑移和驱动能量与地形类型高度相关[8];漫游者可以通过了解地形类型来选择能量最优路径[17];虽然 AI4Mars 数据集不使用地质地形类别,但它可以帮助漫游者支持科学探索,因为大多数科学观察都集中在基岩中发现的证据,那里的地质环境比沙子、土壤或浮石保存得更好。AI4Mars 支持的高精度火星地形分类将成为这些高级应用的基础。

6. 结论

我们展示了 AI4MARS,这是一个用于火星上地形感知自治的大型数据集。我们对收集和处理数据的过程、其组成的统计数据以及使用标签多于图像的数据的方法进行了广泛的概述。我们的实验为这些数据提供了基线结果,并提供了如何处理这些数据以实现最佳模型性能的想法。我们希望这个数据集将促进未来关于地外自治的工作,并期待阅读利用它的研究。

7. 致谢

我们感谢 Zooniverse 的工作人员,特别是 Cliff Johnson,感谢他们在收集数据集、注释工具和支持方面提供的帮助。我们还要感谢 Brandon Rothrock、Ryan Kennedy 和 Jeremie Pa pon 在复制和扩展 SPOC [20]结果方面的指导,以及帮助标签接受工作的 Hong Xiangyu Hong。我们还要感谢 Keri Bean、Fred Calef、Doug Ellison、Abigail Fraeman、Sharon Laubach、Camden Miller、Tyler del Sesto、Nathaniel Stein、Ashley Stroupe 和 Nathan Williams 提供专家标签。

最重要的是,我们要感谢参与 AI4Mars 工作的数千名公民科学家;如果没有您的帮助,我们将无法取得这些成果。根据与美国国家航空航天局(80NM0018D0004)的合同,这项研究在加州理工学院喷气推进实验室进行。

参考

[1] RE Arvidson,P DeGrosse Jr.,JP Grotzinger,MC Heverly,J Shechet, SJ Moreland,MA Newby,N Stein,AC Steffy,F Zhou 等。将地质单元和移动系统运动学联系起来,导致火星大风口的好奇轮损坏。大地力学杂志, 73:73-93,2017。

[2] K.-Michael Aye,Megan E. Schwamb,Ganna Portyankina,Candice J. Hansen,Adam McMaster,Grant RM Miller,Brian Carstensen, Christopher Snyder,Michael Parrish,Stu art Lynn 等。第四行星:通过绘制南极二氧化碳喷射沉积物来探测火星上的春季风。伊卡洛斯,319: 558-598,2019 年 2 月。

[3] 杰弗里 J. Biesiadecki 和 MarkW.迈蒙。火星探测车表面机动飞行软件: Driving ambition。 2006 年 IEEE 航空航天会议论文集,2006 年。

[4] 罗伯特·布乔拉奇、史蒂文·班福德、保罗·塔尔、尼尔·萨克尔、伊恩·克劳福德、凯瑟琳·H·乔伊、彼得·M·格林德罗德和克里斯·林托特。月球动物园公民科学项目:阿波罗 17 号着陆点的初步结果。伊卡洛斯,271:30-48,2016。

[5] Holger Caesar,Varun Bankiti,Alex H Lang,Sourabh Vora,Venice Erin Liong,Qiang Xu,Anush Krishnan,Yu Pan,Gi ancarlo Baldan 和 Oscar Beijbom。nuscenes:自动驾驶的多模态数据集。在 IEEE/ CVF 计算机视觉和模式识别会议论文集上,第 11621-11631 页,2020 年。

[6] 陈良杰、朱玉坤、乔治·帕潘德里欧、弗洛里安·施罗夫和哈特维格·亚当。用于语义图像分割的具有多孔可分离卷积的编码器-解码器。在 ECCV,2018 年。

[7] Marius Cordts,Mohamed Omran,Sebastian Ramos,Timo Rehfeld, Markus Enzweiler,Rodrigo Benenson,Uwe Franke,Stefan Roth 和 Bernt Schiele。用于语义城市市场景理解的城市景观数据集。在 IEEE 计算机视觉和模式识别会议论文集上,第 3213-3223 页,2016 年。

[8] C. Cunningham,M. Ono,I. Nesnas,J. Yen 和 WL Whitaker。使用高斯过程的行星漫游车的局部自适应滑动预测。 2017 年 IEEE 国际机器人与自动化会议 (ICRA),第 5487-5494 页,2017 年。

[9] G. Doran,S. Lu,M. Liukis,L. Mandrake,U. Rebbapragada,KL Wagstaff,J. Young,E. Langert,A. Braunnegg,P. Hor ton,D. Jeong 和 A.特罗克曼。Cosmic:基于内容的董事会摘要来监控不经常发生的变化。 2020 年 IEEE 航空航天会议,2020 年第 1-12 页。

[10] Andreas Geiger,Philip Lenz,Christoph Stiller 和 Raquel Urtasun。视觉遇到机器人:kitti 数据集。国际机器人研究杂志 (IJRR),2013 年。

[11] 何凯明,张翔宇,任少卿,孙健。用于图像识别的深度残差学习。在 IEEE 计算机视觉和模式识别会议论文集,第 770-778 页,2016 年。

[12] S. Higa,Y. Iwashita,K. Otsu,M. Ono,O. Lamarre,A. Di dier 和 M. Hoffmann。基于视觉的驾驶估计使用深度学习和 terrame chanics 为行星漫游者提供能量。 IEEE 机器人与自动化快报,4(4):3876-3883,2019。

[13] Hiroka Inoue,Masahiro Ono,Sakurako Tamaki 和 Shuichi Adachi。行星探测器的主动定位。 2016 年 IEEE 航空航天会议论文集,2016 年。

[14] 威尔·马德恩、杰弗里·帕斯科、克里斯·莱恩加和保罗·纽曼。 1 年,1000 公里:牛津机器人汽车数据集。国际机器人研究杂志,36 (1) :3-15,2017。

[15] JN Maki,JF Bell,Kenneth E Herkenhoff,SW Squyres,A Kiely,M Klimesh,M Schwochert,T Litwin,R Willson,A Johnson 等。火星探测车工程相机。地球物理研究杂志:行星,108 (E12) ,2003。

[16] J Maki,D Thiessen,A Pourangi,P Kobzeff,T Litwin,L Scherr,S Elliott, A Dingizian 和 M Maimone。火星科学实验室工程相机。空间科学评论, 170 (1-4) :77-93,2012。

[17] Masahiro Ono, Brandon Rothrock, Kyohei Otsu, Shoya Higa, Yumi Iwashita, Annie Didier, Tanvir Islam, Christopher Laporte, Vivian Sun, Kathryn Stack, Jacek Sawoniewicz, Shreyansh Daftry, Virisha Timmaraju, Sami Sahnoun, Chris A. Mattmann, Olivier Lamarre,Sourish Ghosh,Dicong Qiu,Shunichiro Nomura,Hiya Roy,He manth Sarabu,Gabrielle Hedrick,Larkin Folsom,Sean Suehr 和 Hyoshin Park。Maars:用于自动漫游车系统的基于机器学习的分析。 2020 年 IEEE 航空航天会议论文集,2020 年。

[18] Kyohei Otsu,Guillaume Matheron,Sourish Ghosh,Olivier Toupet 和 Masahiro Ono。具有铰接式悬挂系统的漫游车的快速近似间隙评估。现场机器人杂志,37(5):768-785, 2020。

[19] Arturo Rankin,Mark Maimone,Jeffrey Biesiadecki,Nikunj Patel, Dan Levine 和 Olivier Toupet。激发好奇心:火星探测器前七年的移动趋势。 2020 年 IEEE 航空航天会议论文集,2020 年。

[20] Brandon Rothrock,Ryan Kennedy,Chris Cunningham,Jeremie Papon,Matthew Heverly 和 Masahiro Ono。SPOC:用于火星探测器任务的基于深度学习的地形分类。在 AIAA SPACE 2016,第 1-12 页,2016 年 9 月。

[21] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, 等。Imagenet 大规模视觉识别挑战。国际计算机视觉杂志,115(3):211-252, 2015。

[22] 大卫·所罗门。计算机图形中的变换和投影。施普林格科学与商业媒体,2007 年。

[23] SP Schwenzer,M. Woods,S. Karachalios,N. Phan 和 L. Joudrier。Labelmars:通过机器学习创建一个非常大的火星图像数据集。在 2019 年第 50 届月球与行星科学会议上。

[24] Chen Sun,Abhinav Shrivastava,Saurabh Singh 和 Abhinav Gupta。重新审视深度学习时代数据的不合理有效性。在 IEEE 计算机视觉国际会议论文集上,第 843-852 页,2017 年。

[25] RL Tokar,RC Elphic,WC Feldman,HO Funsten,KR Moore,TH Prettyman 和 RC Wiens。火星

南极残余极冠的奥德赛中子传感。地球物理研究快报,30(13),2003。

[26] Olivier Toupet, Tyler Del Sesto, Masahiro Ono, Steven Myint, 约书亚·范德·胡克和迈克尔·麦克亨利。基于 ros 的模拟器,用于测试 mars 2020 漫游者的增强型自主导航。2020 年 IEEE 航空航天

会议论文集,2020 年。

[27] 于费雪,陈浩峰,王欣,冼文琪,莹莹 Chen, Fangchen Liu, Vashisht Madhavan 和 Trevor Darrell。Bdd100k:异构驱动数据集多任务学习。在 IEEE/CVF 计算机视觉和模式识别会议论文集上,页数 2636–2645, 2020。