arXiv:2301.02211v1 [cs.CY] 5 Jan 2023

生态学计算机视觉教学

ElijahCole[1] Suzanne Stathatos[1] Bj¨orn L¨utjens[2] Tarun Sharma[1]Justin Kay[1,3] JasonParham[4] BenjaminKellenberger[5] SaraBeery[1,2]1加州理工学院 [2]麻省理工学院 [3]Ai.Fish [4]Wild Me [5]耶鲁大学https://cv4ecology.caltech.edu/

摘要

计算机视觉能够通过自动分析来自相机陷阱、无人机和卫星等传感器的原始图像，加速生态学研究。然而，计算机视觉是一门新兴学科，很少在生态学家中开设相关课程。本文介绍了我们在一场密集、实践性强的暑期研讨会中，指导一群背景多元的生态学家制作原型并评估计算机视觉系统的经验。我们说明了研讨会的结构，讨论常见挑战，并提出最佳实践。本文面向跨学科教授计算机视觉的计算机科学家，同时也可能对正在学习如何使用计算机视觉的生态学家及其他领域专家有所帮助。

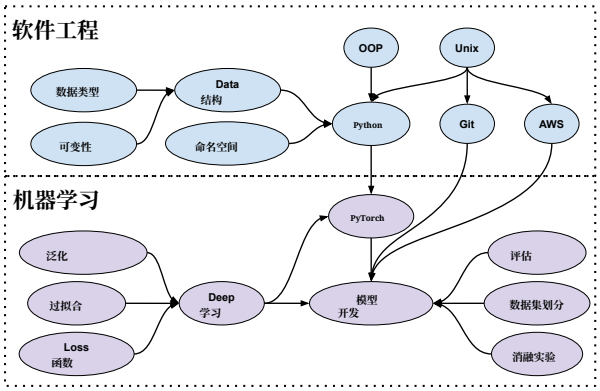


图 1：展示开发计算机视觉系统所需部分技能的简化依赖关系图。这些软件工程和机器学习主题很少被纳入生态学训练中。关键主题及其重要性的目录见附录 B。

1 引言

从图像和视频中提取重要信息通常需要人工标注员付出艰苦的手动劳动。计算机视觉算法可以自动完成这一过程。当数据量过大（例如每天收集的 >100 TB 卫星影像）或标注员数量过少（例如识别图像中的某个物种需要专业知识）而无法进行人工审核时，这一点尤为重要。这两项挑战在生态学中都很常见。

生态学是计算机视觉一个极具吸引力的应用领域。受气候变化影响，我们需要大规模监测动物种群、植被属性以及其他生态系统健康指标 [41]。生态学家正借助相机陷阱、无人机和卫星收集海量原始数据，但专家数量不足以为这些数据进行标注。计算机视觉算法能够高效地将这些原始数据转化为有用的知识，从而加快生态学研究的进程。目前，在动物检测、细粒度物种识别 [18, 25, 29], 个体再识别 [42],物种分布建模 [38], 以及土地覆盖制图 [19,21], 等领域，已经取得了令人鼓舞的进展。这些工作可以放在计算可持续性 [32]这一更广泛的背景下来看，也体现了运用机器学习应对气候变化影响的努力 [23] 。

为了在这一进展的基础上继续前进，我们必须使生态学家掌握理解和应用计算机视觉方法开展研究所需的技能。虽然生态学家通常接受过统计学和编程方面的训练，但他们很少接触计算机视觉所必需的软件工程与机器学习主题之间相互关联的知识体系。我们在图 1 中展示了其中几个主题。

在这项工作中，我们探讨了如何在*Resnick*可持续发展研究所生态学计算机视觉方法暑期研讨会（CV4E研讨会）的背景下向生态学家教授计算机视觉。该研讨会是在加州理工学院举办的为期 3 周的密集研讨会，时间为 2022 [4]。我们将在第 2 节回顾相关工作，然后

在第 3 节介绍该研讨会，在第 4 节讨论主要收获，并在第 5 节概述我们认为有用的教学技巧。

2 相关工作

目前已有越来越多的文献致力于教授机器学习 [20, 34, 37], 深度学习 [30],和计算机视觉 [24, 26, 31, 36]。 [35] 则更狭义地聚焦于机器学习课程项目中的常见错误。然而，这些研究大多关注如何向计算机科学或相关学科的学生传授知识。已有研究探讨了向跨学科受众教授机器学习的具体挑战，例如非计算机科学专业的本科生 [39], 商科学生 [43], 艺术家 [22],材料科学家 [40], 和生物学家 [28]。我们的工作与之互为补充，聚焦于向生态学家（主要是博士生和博士后——见图 2）教授计算机视觉的过程；这些生态学家具备统计学和编程方面的背景知识，但此前几乎没有机器学习经验。此外，我们探讨了一种沉浸式研讨会，研究人员在其中使用自己的研究数据构建原型，而非传统的课堂环境。

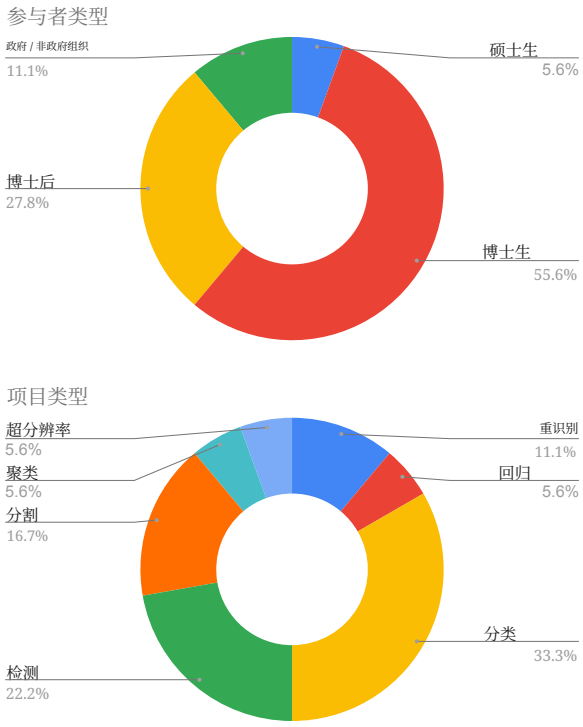


图2：2022 CV4E 研讨会参与者背景（上）和项目类别（下）汇总。完整详情见附录A。

3 CV4E研讨会

首届CV4E研讨会于2022年8月1日至19日在加州理工学院举行。该项目旨在培训生态学家，使其能够在自己的研究中运用计算机视觉。本文概述研讨会的各个阶段。

申请。申请包含五个部分：(i) 一页的项目提案，(ii) 一页的个人陈述，(iii) 编程示例，(iv) 一封推荐信，以及 (v) 一份简历。最重要的部分是项目提案，参与者在其中描述了他们希望用计算机视觉解决的问题、可行解决方案可能产生的影响，以及现有的数据和标签。

遴选流程。CV4E工作人员从机器学习和生态学领域邀请评审人员。每份申请由两位评审人员审阅。最终决定由CV4E工作人员作出。主要评选标准包括：(i) 目标是否明确，(ii) 项目是否可行，(iii) 潜在影响，以及 (iv) 申请者的准备情况。关于2022届学员的详细信息，请参见图2和附录A。为尽可能提高参与便利性，所有参与者在项目期间的交通、住宿和膳食费用均由项目承担。

研讨会前的准备。所有参与者都被添加到 Slack 工作区，该工作区是研讨会的主要沟通渠道。每位参与者都被分配到一个由 CV4E讲师指导的工作小组。在参与者入选与研讨会开始之间的 ~6 个月里，参与者与指导教师会面，以敲定项目计划并解决数据或标签方面的问题。参与者还需要在此期间学习 Python。指导教师会提供 Python 学习资源，并每两周安排一次答疑时间，为参与者提供帮助。

线下研讨会。图3展示了 CV4E研讨会具有代表性的一周日程安排。参与者通过讲座和特邀演讲嘉宾接受课堂教学。每位参与者都加入了一个阅读小组，围绕自己选择的主题（见附录D）开展每周两次的研究论文指导讨论。在工作时间，参与者独立开展项目，CV4E工作人员和工作小组的同伴随时解答问题。每个工作小组都会在小组进展汇报期间讨论进展和遇到的障碍。

成果。在研讨会结束时，18 位参与者都已为各自的项目训练好了模型

。其中一些模型已经表现出色，另一些则还需要进一步研究。此外，参与者和工作人员建立起一个持续至今的社群，并通过 Slack 工作区和持续开展的项目保持联系。

	星期一	星期二	星期三	星期四	星期五
上午9:00	讲座	特邀演讲嘉宾 特邀演讲嘉宾		讲座	特邀演讲嘉宾
上午10:00	工作时间	讲座	讲座	工作时间	讲座
上午11:00	午餐	午餐	午餐	午餐	午餐
中午12:00					
下午1:00	特邀演讲嘉宾 读书小组	特邀演讲嘉宾 读书小组			工作时间
下午2:00	工作时间	工作时间	工作时间	工作时间	
下午3:00					
下午4:00			小组进展汇报		小组进展汇报
下午5:00	休息	休息	休息	休息	休息
下午6:00	晚餐	晚餐	晚餐	晚餐	晚餐
晚上 7:00					

图3：2022 CV4E 研讨会的每周日程大致平均分配为教学时间（讲座、特邀演讲嘉宾、读书小组、小组进展汇报）和由讲师监督的工作时间（工作时间）。实际上，参与者经常会利用部分小组进展汇报、午餐、休息和晚餐 时段作为额外的工作时间。

4 点经验教训

加强 Python 预备学习的结构化安排。 对大多数参与者来说，主要障碍是 Python 准备不足。虽然申请前不要求参与者掌握 Python，但他们需要在抵达前学习 Python。为方便大家学习，工作人员提供了 Python 学习资源，并在 CV4E研讨会 开始前的几个月安排了答疑时间。然而，许多参与者（甚至是能力出众的 R 语言程序员）在整个研讨会期间仍然受到 Python 问题的困扰。事后看来，我们高估了 R 语言经验对快速学习 Python 的帮助。今后，我们将在研讨会开始前对 Python 预备学习进行更有条理的安排。

从简单做起。 在 3 周内从头构建一个可运行的计算机视觉系统颇具挑战。为了尽可能提高成功率，从简单做起至关重要。适当情况下，我们会鼓励参与者采用标准且成熟的流程，例如对经过 ImageNet 预训练的 ResNet-50 进行微调。

分长时段工作。 与较短的工作时段（1–2 小时）相比，参与者在较长的工作时间块（3+小时）内取得的进展要多得多。

将相似的项目归入工作小组。参与者通常都很乐意互相帮助，尤其是在部署相似技术时。组建工作小组时，应尽可能创造这类合作机会。

在工作小组中混合不同经验水平的参与者。有些参与者在机器学习或编程方面经验丰富，因此在指导教师只需提供少量协助的情况下，就能迅速推进项目。经验丰富的参与者经常主动帮助经验较少的参与者，这似乎对双方都有益。今后，我们计划确保每个工作小组都由经验丰富和经验较少的参与者共同组成。

提供明确无歧义的基础设施建议。搭建计算机视觉工作所需的基础设施，有许多合理的方式。举例来说，考虑开发一段需要在虚拟机上运行的代码。一种做法是在本地文本编辑器中编辑代码，再使用 rsync 将代码传到虚拟机，并在本地进行版本控制。另一种做法是使用 VSCode [17] 这类工具，通过 SSH 直接编辑虚拟机上的代码。在这种情况下，版本控制将在虚拟机上进行。第三种做法是在本地编辑代码，将其推送到 GitHub，再拉取到虚拟机上。这种工作流程内置了版本控制。指导教师各有偏好，而且没有哪种工作流程明显更优。让参与者自行选择使用哪种工作流程并没有带来好处。今后，我们将为开发基础设施提供明确统一的建议。

避免使用深度学习库封装。 深度学习库有许多“封装”，旨在让深度学习工具更易于使用。有些是通用型的（例如 PyTorch Lightning [13]），另一些则是特定领域的（例如 Open Soundscape [11]，DeepForest [5]，TorchGeo [16]）。虽然这些封装无疑很有用，但对于我们的参与者来说并不理想，原因有二。首先，它们隐藏了过多复杂性，妨碍了参与者学习训练循环和数据流等内容。其次，即使有讲师协助，这些封装也更难定制和调试。今后，我们将鼓励所有参与者直接使用深度学习库。

避免使用 Jupyter 笔记本。 Jupyter 笔记本提供了许多经验丰富的 R 用户熟悉的功能，例如可以交互式运行代码片段。然而，学习 Python 时依赖 Jupyter 笔记本的学员，在开发计算机视觉系统时，往往难以转向更传统的命令行工作流程。我们现在认为，学习使用 Python

通过命令行进行操作，能为理解机器学习工作流程打下更好的基础。

确保有可用的 GPU。 AWS 和 Azure 等云计算服务通常会为教育和研究提供免费额度。不过，GPU 是否可用可能取决于客户需求。务必向云服务提供商确认是否能提供 GPU。或者，也可以考虑使用本地计算资源或大学集群。

5 教学技巧

本节介绍一些我们发现对 CV4E研讨会很有帮助的教学技巧。

引导式故障排查。 故障排查和调试是机器学习中的重要技能，因此，为参与者提供磨练这些能力的机会十分重要。然而，由于 CV4E研讨会日程紧凑，我们不希望参与者在任何一个问题上耗费太多时间。为兼顾这两个目标，指导教师尝试通过针对参与者遇到的问题提出引导性问题，带领他们完成故障排查流程。对于教育价值有限的特殊问题（例如复杂的配置或安装问题），指导教师会介入并尽快解决。

结对伪代码编写。 在 CV4E研讨会开始时，大多数参与者还不太习惯编写 Python代码，因此我们希望为他们提供频繁动手编程的机会。只要条件允许，指导教师就会避免替参与者编写代码。为防止参与者在代码设计问题上陷入困境，我们采用了结对伪代码编写：

1. 指导教师请参与者说明他们想要实现什么，并通过讨论确保双方都明确目标。

2. 指导教师编写能够解决问题的伪代码，并与参与者一起逐步讲解，帮助他们理解解决方案的逻辑。根据参与者的需求，伪代码可以写得更具体或更概括。

3. 参与者编写 Python代码来解决问题，指导教师则随时准备回答他们遇到的问题。

目标陈述。 在项目初期，一些参与者觉得进展停滞不前，因为代码还没能“运行起来”。为了让他们更清楚地看到自己的进步，一些指导教师要求参与者在每个工作时段开始时写下一份目标陈述，并在每个工作时段结束时检查目标的完成进度。这一策略帮助参与者保持动力，直到取得更切实的成果。

情境化讲座。 由于参与者热情高涨，在 CV4E研讨会期间，保持他们对讲座的兴趣并不是一个突出问题。不过，机器学习主题的讲座很容易变得过于抽象。我们努力确保讲座始终立足于应用和示例。由于每位参与者心中都有自己的应用问题，我们经常暂停讲座，请参与者思考讲座主题如何应用于他们的个人项目。参与者会与全班分享自己的想法，提供具体示例来阐释讲座主题。

6 结论

我们介绍了首届 Resnick可持续发展研究所生态学计算机视觉方法暑期研讨会的经验。我们认为这种形式取得了成功，因为所有参与者都在研讨会结束前为各自的项目训练了模型。不过，我们也讨论了遇到的一些挑战，并指出了改进 CV4E研讨会的机会。我们希望这些观察能为其他跨学科教授计算机视觉的人提供参考。

7 致谢

我们谨向 Resnick可持续发展研究所、Caroline Murphy、Xenia Amashukeli 和 Pietro Perona 致谢，感谢他们促成了 CV4E研讨会的举办。计算资源额度由 Amazon AWS 和微软 Azure 提供。我们还要感谢 CV4E研讨会的首届学员：Ant´on Alvarez、Carly Batist、Peggy Bevan、Catherine Breen、Anna Boser、Tiziana Gelmi Candusso、Melanie Clapham、Rowan Converse、Roni Goldshmid、Natalie Imirzian、Brian Lee、Francesca Ponce、Alixandra Prybyla、Rachel Renne、Felix Rustemeyer、Taiki Sakai、Ethan Shafron 和 Casey Youngflesh。

参考文献

[1] 亚马逊云科技 (AWS)。 https://aws.amazon.com/.[2] Audacity。 https://www.audacityteam.org/。

[3] 计算机视觉标注工具 (CVAT)。 https://github.com/opencv/cvat。

[4] CV4E夏季研讨会。 https://cv4ecology.caltech.edu/。

[5] DeepForest。 https://deepforest.readthedocs.io/。

[6] FFMPEG。https: // ffmpeg.org/doxygen/3.0/index.html。

[7] ImageMagick。 https://imagemagick.org/index.php。

[8] ImgLab。https: // github.com/NaturalIntelligence/imglab。

[9] 微软 Azure。 https://azure.microsoft.com/.[10] OpenCV。 https://opencv.org/.[11] OpenSoundscape。 http://opensoundscape.org/.[12] PyTorch。 https://pytorch.org/.[13] PyTorch Lightning。 https://www.pytorchlightning.ai/.[14] TensorBoard。 https://www.tensorflow.org/tensorboard.[15] TensorFlow。https://www.tensorflow.org/.[16] TorchGeo。https: // torchgeo.readthedocs.io/en/stable/.[17] VSCode。 https://code.visualstudio.com/.[18] Sara Beery、Dan Morris 和 Siyu Yang。高效的相机陷阱图像审查流程。*arXiv* 预印本 *arXiv:1907.06772*，2019年。[19] Elijah Cole、Benjamin Deneu、Titouan Lorieul、Maximilien Servajean、Christophe Botella、Dan Morris、Nebojsa Jojic、Pierre Bonnet 和 Alexis Joly。geolifeclef 2020 数据集。*arXiv* 预印本 *arXiv:2004.04192*，2020年。[20] Adrian A de Freitas 和 Troy B Weingart。我要学什么？！在计算机科学导论课程中向大学一年级新生教授人工智能。载于第 *52* 届 *ACM* 计算机科学教育技术研讨会论文集，第 198–204 页，2021年。[21] Benjamin Deneu、Alexis Joly、Pierre Bonnet、Maximilien Servajean 和 Fran¸cois Munoz。基于遥感影像的超高分辨率物种分布建模：如何捕捉细粒度和大尺度植被生态学与卷积神经网络？《植物科学前沿》，第 13 卷，文章编号 839279，2022年。

[22] Rebecca Fiebrink。面向艺术家、音乐家及其他创意实践者的机器学习教育。*ACM*计算教育汇刊 *(TOCE)*，19(4):1–32，2019年。

[23] Carla Gomes、Thomas Dietterich、Christopher Barrett、Jon Conrad、Bistra Dilkina、Stefano Ermon、Fei Fang、Andrew Farnsworth、Alan Fern、Xiaoli Fern 等。计算可持续性：以计算创造更美好的世界和可持续的未来。《*ACM*通讯》，62(9):56–65，2019年。[24] Tal Hassner 和 Itzik Bayaz。计算机视觉教学：将研究基准引入课堂。*ACM*计算教育汇刊 *(TOCE)*，14(4):1–17，2015年。[25] Benjamin Kellenberger、Diego Marcos 和 Devis Tuia。无人机图像中的哺乳动物检测：使用深度学习处理严重类别不平衡数据集的最佳实践。《环境遥感》，216:139–153，2018年。[26] Sami Khorbotly。面向本科阶段计算机视觉教学的项目式学习方法。载于*2015*年美国工程教育学会年度会议暨展览会，第26–91页，2015年。[27] Zachary C Lipton 和 Jacob Steinhardt。机器学习学术研究中的令人担忧趋势。*arXiv*预印本 *arXiv:1807.03341*，2018年。[28] Chris S Magnano、Fangzhou Mu、Rosemary S Russ、Milica Cvetkovic、Debora Treu 和 Anthony Gitter。面向生物学家的易于入门、灵活且实用的机器学习研讨会。$bioRxiv$，2022年。[29] Jason Parham、Charles Stewart、Jonathan Crall、Daniel Rubenstein、Jason Holmberg 和 Tanya Berger-Wolf。用于身份识别的动物检测流程。载于*2018*年*IEEE*冬季计算机视觉应用会议 *(WACV)*，第1075–1083页，2018年。[30] Simon J.D. Prince。理解深度学习。麻省理工学院出版社，2022年。[31] S.J.D. Prince。《计算机视觉：模型、学习与推理》。剑桥大学出版社，2012年。

[32] Caleb Robinson、Le Hou、Kolya Malkin、Rachel Soobitsky、Jacob Czawlytko、Bistra Dilkina 和 Nebojsa Jojic。利用多分辨率数据进行大规模高分辨率土地覆盖制图。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第12726–12735页，2019年。

[33] David Rolnick、Priya L Donti、Lynn H Kaack、Kelly Kochanski、Alexandre Lacoste、Kris Sankaran、Andrew Slavin Ross、Nikola Milojevic-Dupont、Natasha Jaques、Anna Waldman-Brown 等人。《利用机器学习应对气候变化》。*ACM*计算调查 *(CSUR)*，55(2):1–96，2022年。

[34] Omar Shouman、Simon Fuchs 和 Holger Wittges。《向背景各异的硕士生教授实践机器学习课程的经验》。载于第二届机器学习与人工智能教学研讨会论文集，第62–67页。PMLR，2022年。

[35] James Skripchuk、Yang Shi 和 Thomas Price。《识别开放式机器学习项目中的常见错误》。载于第*53*届*ACM*计算机科学教育技术研讨会论文集，第*1*卷，第216–222页，2022年。

[36] Scott Spurlock 和 Shannon Duvall。《让本科生也能轻松学习计算机视觉》。《高校计算科学杂志》，33(2):215–221，2017年。

[37] Peter Steinbach、Heidi Seibold 和 Oliver Guhr。2020年机器学习教学。载于欧洲机器学习与数据库知识发现原理及实践会议，第 1–6 页。PMLR，2021年。

[38] Charles V. Stewart、Jason R. Parham、Jason Holmberg 和 Tanya Y. Berger-Wolf。动物个体识别问题：持续整理。*arXiv* 预印本 *arXiv:2106.10377*，2021年。

[39] Elisabeth Sulmont、Elizabeth Patitsas 和 Jeremy R Cooperstock。向非专业学生教授机器学习有哪些难点？从指导教师学习目标的分类中获得的见解。*ACM*计算教育汇刊 *(TOCE)*，19(4):1–16，2019年。

[40] Shijing Sun、Keith Brown 和 A Gilad Kusne。向材料科学家教授机器学习：托管教程和竞赛的经验。*Matter*，5(6):1620–1622，2022年。

[41] Devis Tuia、BenjaminKellenberger、Sara Beery、Blair R Costelloe、Silvia Zuffi、Benja min Risse、Alexander Mathis、Mackenzie W Mathis、Frank van Langevelde、Tilo Burghardt 等。野生动物保护领域机器学习的展望。《自然·通讯》，13(1):1–15，2022年。

[42] Grant Van Horn、Oisin Mac Aodha、Yang Song、Yin Cui、Chen Sun、Alex Shepard、Hartwig Adam、Pietro Perona 和 Serge Belongie。iNaturalist物种分类和检测数据集。载于*IEEE*计算机视觉与模式识别会议论文集，第 8769–8778 页，2018年。

[43] Linus Wunderlich、Allen Higgins 和 Yossi Lichtenstein。《面向商科学生的机器学习：一种体验式学习方法》。载于第 *26* 届 *ACM* 计算机科学教育中的创新与技术会议论文集，第 *1* 卷，第 512–518 页，2021年。

2022届学员

A.1 参与者背景

首届 2022 CV4E 研讨会共有 18 名参与者。按当前职业划分，我们的队伍包括:

- 1硕士生;
- 10博士生;
- 5 博士后研究人员；以及
- 2 政府机构或非政府组织的研究人员。

按地理位置划分，我们的学员群体包括:

- 11 名来自美国7个不同州的参与者;
- 5 名来自欧洲国家的参与者；以及
- 2名来自加拿大的参与者。

参与者的学术背景多种多样，包括保护生物学、生物人类学、地理学、机械工程、土木工程、神经科学和生态学。

A.2 参与者项目

项目分为七个主要类别。

1.**个体再识别：** 将从不同相机、地点和时间拍摄的同一动物的图像对应起来。相关的两个项目是: (1) 重新识别熊；(2) 重新识别伊比利亚猞猁。

2. **回归：** 为图像或视频赋予一个连续数值。相关的项目是：(1) 根据树木的俯拍无人机视频分析风速。

3. **分类：**将图像或图像中的部分内容归入一组固定类别。六个相关项目分别是：(1) 判断是否存在狐猴发声，(2) 根据回声定位点击声进行喙鲸物种分类，(3) 根据飞行声音进行熊蜂物种与级型分类，(4) 将蚂蚁划分为不同体型类别，(5) 根据相机陷阱图像识别天气状况，以及 (6) 在城市相机陷阱中识别物种。请注意，项目 (1)、(2) 和 (3) 使用计算机视觉技术，对代表音频信号的图像（频谱图）进行分类。

4. **目标检测：**在图像或视频中定位物体实例。四个相关项目分别检测：(1) 水源周边放牧区，(2) 林地沟谷，(3) 苍蝇，以及 (4) 水禽。项目 (3) 和 (4) 将检测作为计数的中间步骤。

5.**分割：** 根据像素的语义特征对其进行分类。相关的三个项目分别是分割：(1) 北极地区的海象群；(2) 永久冻土；(3) 树木。所有项目均基于遥感影像。

6. **聚类：** 根据某种相似性概念将对象归为一组。相关的项目是：(1) 利用一组相机陷阱影像中的聚类数量，确定某个区域的物种丰富度。

7. **超分辨率：** 提高图像的分辨率。相关的项目是：(1) 使用卫星影像提高地表温度数据的分辨率。

B 关键主题

本节汇总了许多参与者在研讨会期间通过正式教学或自学所学习的主题。我们着重介绍最初对大多数参与者而言较为陌生的工具和概念。对于每个主题，我们都会介绍相关内容，并说明其对参与者的重要性。另请参阅附录 C 中的讲座列表。

B.1 工具

B.1.1 标注工具

内容：使用标注工具为图像 [3,8]或音频 [2]数据添加标签。

动机：标注数据对于训练和评估计算机视觉算法至关重要。由于 CV4E研讨会的参与者使用的是自己的数据，许多人需要学习使用某种标注工具。此外，许多此类工具都能以开源计算机视觉库所需的标准格式导出标签。

B.1.2 Unix 命令

内容：常见的 Unix 命令，如 ls、pwd、rm、mkdir、rmdir、mv、cat、head、tail 等。偶尔也会用到 chmod 或 grep 等不太常见的命令。

学习动机：熟练使用 Unix 命令对于安装软件包、使用虚拟机至关重要，

也有助于使用版本控制。理解 Unix 命令还能帮助建立对绝对路径与相对路径等核心概念的直觉。

B.1.3 基于终端的文本编辑

内容： 使用 nano 等工具，从命令行编辑存储在服务器上的文本。

动机： 配置 SSH 身份验证时，通常需要在虚拟机上编辑文本文件（例如 ~/.ssh/config）。

B.1.4 终端复用

内容： 使用 tmux 或 screen 等工具管理终端会话。

动机： 长时间运行的代码（例如 PyTorch 中的模型训练）应在与 SSH 连接分离的终端会话中执行，以避免笔记本电脑合上或网络连接中断时进程被终止。

B.1.5 SSH

内容： ssh 命令和 SSH 密钥。偶尔也会用到 SSH 隧道。

动机： ssh 命令用于创建连接到虚拟机的终端会话。SSH 密钥等相关主题也很重要，例如，它们可用于验证基于终端的文件传输，以及启用 GitHub 访问。设置 TensorBoard [14] 等工具时，可能需要使用 SSH 隧道。

B.1.6 基于终端的文件传输

内容： 用于传输文件的工具，例如 scp 或 rsync。

动机： 命令行工具是将大量数据从一个位置移动到另一个位置最可靠的方式。这适用于本地传输（例如，从一块硬盘传输到另一块硬盘）和远程传输（例如，从本地硬盘传输到附加到虚拟机的存储卷）。

B.1.7 版本控制

内容： 使用 GitHub 跟踪代码变更。

动机： 计算机视觉项目的代码往往很快变得复杂，而且很容易忘记自上一个可正常运行的版本以来发生了哪些变化。如果某次更改引入了错误，GitHub 等工具可以让人轻松查看代码的早期版本。此外，GitHub 还可用于将代码从本地计算机推送到虚拟机（git push），并允许用户从虚拟机拉取代码（git pull），还可以下载开源计算机视觉代码库（git clone）。

B.1.8 云计算

内容： 通过云计算服务提供商的网页界面进行操作。创建一台配备适当资源（如 GPU、存储空间）的虚拟机。估算并管理成本。

动机： 在计算机视觉工作中，获取 GPU 资源最常见的方式之一是使用云计算服务提供商（如亚马逊云科技（AWS） [1] 或微软 Azure [9]）提供的虚拟机。了解云计算的优点（可扩展性、可靠性）和缺点（成本）非常重要。

B.1.9 虚拟环境

内容： 创建和管理虚拟环境。

动机： 计算机视觉项目通常依赖大型的现有代码库，而这些代码库可能要求安装特定版本的软件包。用户虽然可以更改基础安装，但更好的解决方案是创建一个虚拟环境（例如通过 conda），并在其中安装代码库的依赖项。如果需要“彻底重新安装”，虚拟环境也很有用，因为它们易于创建和删除。

B.1.10 Python

内容： 基本语法、条件语句、循环、字符串解析、文件输入/输出、函数、类和数据结构。

动机： 熟练掌握 Python 对于高效使用基于 Python 的深度学习库至关重要，而计算机视觉领域几乎完全依赖这些库。

B.1.11 Python 库

内容： 常用库，如 numpy、pandas、ipdb、sklearn 和 matplotlib。

动机： Python 拥有许多稳定且高质量的数值计算和数据分析库。ipdb 等库支持内联调试。

B.1.12 深度学习库

内容： 构建深度学习系统时，优先使用 PyTorch [12]，也可以使用 TensorFlow [15]。

动机： 现代深度学习库对于开发和训练计算机视觉系统不可或缺。

B.1.13 图像处理库

内容： 库和命令行工具，例如 OpenCV [10], ImageMagick [7],和 FFmpeg [6]。

动机： 这些工具常用于高效的数据增强和可视化。

B.2 计算机科学概念

贯穿整个课程，有几个计算机科学的核心概念反复出现。

B.2.1 面向对象编程

内容： 类和对象。继承、封装、多态。

动机： 许多重要的库都要求使用者理解面向对象编程的概念。例如，参与者普遍感到困惑的一点，是 PyTorch 数据集类与该类的一个数据集对象之间的区别。理解面向对象编程也有助于理解数据结构。

B.2.2 数据结构

内容： 常见的数据结构（例如列表、元组、字典、NumPy 数组、PyTorch 张量）及其方法、将一种数据类型转换为另一种数据类型，以及检查数据结构。

动机： 例如，Python 列表、NumPy 数组和 PyTorch 张量之间出乎意料的行为差异，如果对数据结构缺乏充分了解，就会令人十分沮丧。

B.2.3 数据类型

内容： 常见数据类型，例如整数、浮点数、双精度浮点数、字符串和布尔值。

动机： 理解数据类型有助于加深对上下文的理解，并能显著影响数据存储大小。

B.2.4 命名空间

内容： 内置、全局和局部命名空间。

动机： 命名空间可以解答许多常见问题，例如，为什么在函数内部定义的变量无法在函数外部访问。

B.2.5 可变性

内容： 可变对象和不可变对象。原地操作。

动机： 可变对象可以原地修改，而可变对象则不能。这是理解在函数内部对对象所做的更改是否会影响函数外部对象的基础。

B.3 机器学习概念

参与者根据各自的项目学习了计算机视觉和机器学习的不同实践与概念方面的内容。不过，所有参与者都必须学习一些核心概念。

B.3.1 泛化

内容： 泛化的概念、不同类型的泛化，以及识别能够反映项目目标的泛化类型。

动机： 在生态学中，泛化有许多不同的概念，因此选择一种能够反映项目目标的概念十分重要。例如，在相机陷阱图像分类中，可能需要泛化到新的地点，或泛化到来自相同地点的未来数据。衡量这些不同的泛化概念，需要采用不同的方法。

B.3.2 数据集划分

内容： 训练、验证和测试数据的作用。设计恰当的数据集划分，以衡量所选的泛化类型。

动机： 训练、验证和测试集的划分应当设计为能够体现适合特定问题的泛化概念。随后必须恰当地处理这些数据集划分（例如，不在测试集上调整超参数），以确保性能测量结果能够反映泛化能力。

B.3.3 过拟合

内容： 定义并识别过拟合。使用正则化技术缓解过拟合。

动机： 所有参与者都在使用深度学习，而过拟合始终是一个值得重视的问题。

B.3.4 评估指标

内容： 不同任务的常用评估指标、各自的优势与局限，以及如何选择能反映高层目标的指标。

动机： 恰当的指标对于确定哪些方法效果最好，以及判断计算机视觉系统是否“足够好”、能否用于实际应用至关重要。

B.3.5 深度学习

内容： 神经网络、损失函数、小批量梯度下降。

动机： 所有现代计算机视觉方法都依赖深度学习。由于参与者正在构建和排查计算机视觉系统的故障，他们也需要了解深度学习基础知识。损失函数是重点内容之一，因为修改损失函数是将现有方法适配到新问题的主要方式之一。

B.3.6 表示

内容： 图像嵌入、嵌入空间中的距离、预训练、迁移学习。

动机： ImageNet 预训练在现代计算机视觉中无处不在，但我们的许多参与者从事的是专业领域的工作，ImageNet 预训练未必适用于这些领域。领域专用预训练需要理解表示学习。图像嵌入的概念也有助于理解许多常见的计算机视觉算法（例如度量学习）和可视化技术（例如 t-SNE）。

B.4 其他技能

B.4.1 批判性阅读机器学习论文

内容： 理解机器学习术语和论文结构，批判性解读论断，评估复杂度与性能之间的权衡。

动机： 探索一个新领域的文献总是令人望而生畏。机器学习领域尤其具有挑战性，因为论文可能写得过于热情，因此读者需要格外谨慎才能清楚地理解一种方法 [27]的缺点和优点。

B.4.2 选择“优秀”的开源库

内容： 识别开源代码中的质量标志。

动机： 开源计算机视觉代码十分丰富，但并非所有代码都可靠或维护良好。参与者必须学会检查代码质量的指标，例如某个库的用户数量，或开发者修复错误的频率。

B.4.3 深入了解库

内容： 阅读文档，找到处理特定任务的代码，了解代码库各组件之间如何交互。

动机： 计算机视觉项目依赖众多复杂但（通常）文档完善的库。理解文档至关重要。有时还需要找到并检查文档所描述的代码片段（例如某个库中的函数），以详细了解其工作原理。

B.4.4 常见故障排查

内容： 错误与警告，以及如何搜索错误消息的更多信息。

动机： 例如，在安装软件包或测试新代码时，错误和警告很常见。在任何编程活动中，最重要的技能之一就是能够使用搜索引擎来理解错误消息。这包括找出错误消息中适合作为搜索词的部分，阅读搜索结果，并选择恰当的后续步骤。

B.4.5 Python代码调试

内容： 错误类型、查找错误来源、使用 print 语句进行调试。

动机： 对于给定的一行代码，可能会出现各种错误。了解不同类型的 Python 错误有助于找出根本原因。在远程机器上运行代码时，打印语句调试对于故障排查也非常有用。

C 讲座列表

1. 简介与课程安排（Sara Beery） 2. 数据集原型设计与可视化（Jason Parham） 3. 在云端工作（Suzanne Stathatos） 4. 数据集划分与避免数据投毒（Sara Beery） 5. 训练模型：确定配置、启动、监控、检查点保存并整理训练任务（Benjamin Kellen-berger） 6. 使用开源计算机视觉代码库：选择基线模型并进行自定义数据加载（Sara Beery） 7. 评估指标（Elijah Cole） 8. 离线评估与分析（Sara Beery） 9. 接下来做什么？改善结果的经验法则（Benjamin Kellenberger） 10. 数据增强（Bj¨orn L¨utjens） 11. 使用模型扩展并改进训练数据集：弱监督、自监督、定向重新标注与异常检测（Tarun Sharma） 12. 公平比较与消融研究：了解哪些因素至关重要（Elijah Cole） 13. 高效模型：速度与准确度（Justin Kay） 14. 模型服务、托管与部署以及质量控制（Jason Parham）

D 读书小组列表

1. 时间序列、谱变换与遥感 2. 数据不平衡与长尾分布 3. 弱监督、无监督学习、微调与迁移学习 4. 偏差、领域偏移与泛化