通过无监督分类提升鸟鸣录音数据集的质量

Félix Michaud [1,*], Jérôme Sueur [1], Maxime Le Cesne [1], Sylvain Haupert [1]

[1] 系统学、进化与生物多样性研究所 (ISYEB)，法国国家自然历史博物馆，法国国家科学研究中心、索邦大学、高等研究实践学院、安的列斯大学，居维叶街57号，75005 巴黎，法国

* 通讯作者：系统学、进化与生物多样性研究所 (ISYEB)，法国国家自然历史博物馆，法国国家科学研究中心、索邦大学、高等研究实践学院、安的列斯大学，居维叶街57号，75005 巴黎，法国，+33 1 40 79 33 98，felix.michaud@mnhn.fr

亮点

- 开放音频数据库为生态学提供独特信息
- 然而，其中大多数都是弱标注，存在信号混淆的风险
- 提出了一种基于分割、特征化和聚类的标注函数
- 在 Xeno-Canto 数据库上进行的测试显示，标签错误显著减少
- 无监督标注有助于促进音频档案的利用

摘要

Xeno-Canto 等开放音频数据库被广泛用于构建数据集，以探索鸟鸣曲目，或利用深度学习算法训练鸟类声音自动分类模型。然而，这类数据库存在一个问题：鸟类声音仅为弱标注，即每段音频录音都标注了物种名称，却没有提供时间戳来定位感兴趣鸟鸣的发生时间。人工标注可以解决这一问题，但耗时、依赖专家，且无法应用于大型数据集。另一种解决方案是使用标注函数，在为每个分割后的音频样本分配标签之前，先自动分割音频录音。尽管引入标注函数是为了加快强标签分配，其分类性能却仍鲜为人知。为解决这一问题并减少大型鸟鸣数据集中的标签噪声（错误的标签分配），我们提出了一种以数据为中心的新型标注函数，由以下三个连续步骤组成：1）时频声音单元分割；2）计算每个声音单元的特征；3）使用无监督 DBSCAN 算法或有监督 BirdNET 神经网络，将每个声音单元分类为鸟鸣或噪声。我们使用 44 种西古北界常见鸟类物种的歌声，对该标注函数进行了优化、验证和测试。首先，我们发现，仅对鸟鸣进行分割，就会因物种不同而产生 10% 至 83% 的标签噪声。我们还证明，该标注函数能够显著降低数据集中的初始标签噪声，最多可将其减少至原来的三分之一。最后，我们探讨了设计合适标注函数的不同方法，以最少的专家标注工作构建高质量的动物发声数据集。

关键词

生物声学、生态声学、以数据为中心、鸟鸣、聚类、深度学习、标注函数、标签噪声

1. 引言

约 70 年来，鸟鸣一直是大多数生物声学研究的核心，目前也是陆地环境中的声景生态学和生态声学研究的关键要素（Xie et al., 2020）。鸟类发出的声音，尤其是它们的领地鸣唱，通常使用抛物面麦克风或定向麦克风录制，以便让代表单一物种的某个个体成为录音中强度最高的声音。这类录音称为焦点录音，会被存入并分享于公共和私人声音库中，而这些声音库的规模还在持续扩大（Toledo et al., 2015）。

Xeno-Canto 是使用最广泛、内容最丰富的鸟鸣数据库（Xeno-canto，2021年）。这个社区项目旨在收集并整理由专家和非专家贡献者提供的鸟类声音，目前已收录来自全球各主要地区的 10,300 个物种的约 678,000 条音频录音。Xeno-Canto 可用于分类学、行为、生物地理学和生态学研究（Planqué 和 Vellinga，2008）。近年来，Xeno-Canto 已成为机器学习和深度学习挑战的主要数据来源，这些挑战旨在通过声学实现鸟类物种的自动识别（Joly 等，2015）。用于监测鸟类物种（Denton 等，2022年；Dufourq 等，2022年；Nolasco 等，2022年；Stowell，2022年）、研究方言（Cohen 等，2022年）或识别个体（Martin 等，2022年）的算法也在迅速发展。其中，卷积神经网络 BirdNET 目前似乎是最有前景的开源、易用的鸟类自动识别解决方案（Kahl 等，2021）。

Xeno-Canto 非常有用，但其标注较弱：每段录音都标注了一个物种名称，却没有时间戳来定位鸟鸣序列的时间位置。人们主要采用三种方法，通过在时间上定位感兴趣的声音来减少这种不准确性。第一种方法是请鸟类专家对录音进行人工标注。专家通过聆听音频并查看其声谱图表示，选出每个感兴趣声音的起点和终点，有时还会直接在声谱图上手动绘制边界框。这一过程能够确保数据质量较高，但需要稀缺的专业知识，耗时较长，而且不同专家添加的标签可能不一致，使流程变得更加复杂（Donmez and Carbonell, 2008；Gil-González et al., 2021；Sheng et al., 2008）。第二种方法是半自动标注，将自动聚类与人工标注相结合。这种方法已应用于多种动物发声，包括声景分析（Arthur et al., 2021；Fukuzawa et al., 2020；Phillips et al., 2018；Ruff et al., 2021；Steinfath et al., 2021；Zhong et al., 2021）。本研究关注第三种方法：该方法无需人类专家直接参与，而是开发一个能够自动为每个声音单元分配标签的程序。这类程序称为标注函数（Ratner et al., 2016；Zhang et al., 2022）。

在利用开放数据库构建定制鸟鸣数据集时，大多数标注函数都通过应用分割方法，自动定位显著声音；方法从简单的能量阈值（Fodor, 2013；Lasseck, 2013；Potamitis, 2014）到更复杂的信号与图像处理（Sprengel et al., 2016）不等。随后，系统会自动将已识别鸟类的弱标签分配给所有分割出的声音。然而，分割出的声音可能并非目标声音。这种错误的标签分配称为标签噪声，被指出是开发稳健鸟鸣检测算法的主要局限之一（Kahl 等，2021）。事实上，数据集中的标签噪声会损害分类任务（Frenay and Verleysen, 2014），降低算法的学习性能，甚至阻碍模型收敛（Arpit et al., 2017；Chen et al., 2019；Karimi et al., 2020；Sheng et al., 2008），还可能导致类别不平衡问题。

为了减少标签噪声，人们提出了三种解决方案（Frenay 和 Verleysen，2014年）。第一种方案依赖于某些算法在训练过程中对噪声具有鲁棒性的隐含能力。例如，研究表明，规模极大的人工神经网络能够拟合标签噪声高达 90% 的数据集（Rolnick 等，2017年）。然而，这种以模型为中心的方案需要大量数据和计算能力，难以轻松部署，还会增加计算时间和能源消耗，并加剧 CO_2 排放（Schwartz 等，2020年；Strubell 等，2019年）。鸟鸣检测与分类领域广泛采用了这一方案（Joly 等，2021年；Stowell，2022年）。第二种方案是通过算法显式处理标签噪声（Algan 和 Ulusoy，2021年；Frenay 和 Verleysen，2014年）。这类方案称为噪声容忍方法或基于噪声模型的方法，需要充分了解标签噪声，以便在模型中纳入对噪声的表征。本文关注第三种方案，旨在从源头处理标签噪声，使用上游过滤器检测并从数据集中移除错误标注数据（Brodley 等，1999年）。通过直接从源头移除标签噪声，可以简化学习过程（Sambasivan 等，2021年）；这也符合以数据为中心的方法的主张：采用更干净、更小的数据集，为更轻量的训练过程创造良好条件（Motamedi 等，2021年；Ng 等，2021年）。

为了解决这一问题并减少大型鸟鸣数据集中的标签噪声，我们提出了一种以数据为中心的新型标注函数，由以下三个连续步骤组成：1）时频声音单元分割；2）计算每个声音单元的特征；3）使用无监督 DBSCAN 算法或有监督 BirdNET 神经网络，将每个声音单元分类为鸟鸣或噪声。我们使用西古北界 44 种常见鸟类物种的歌声，对该标注函数进行了优化、验证和测试。

2. 材料与方法

图 1 展示了用于设计该标注函数的完整工作流程概览。

2.1. 音频数据库

该音频数据库包含西古北界雀形目、鸻形目、鸮形目、鹃形目和鸮形目的常见鸟类物种。泛欧洲普通鸟类监测计划（PECBMS，2021年）的实践表明，对这些物种进行监测对生物多样性评估至关重要。音频数据库分为三个数据集：训练集、验证集和测试集。训练集和验证集涉及相同的 22 种鸟类物种（表A.1.）。物种的选择依据一项先前研究，该研究在舍夫勒斯谷地区自然公园（24,300 公顷）开展；该保护区位于法国巴黎西南方向 40 千米处（Depraetere 等人，2012）。第三个数据集包含另外 22 种常见物种，这些物种从全球生物多样性信息机构（GBIF，2021年）数据库中选出（表A.2.）。

所有录音均从 Xeno-Canto 自动获取。这个国际在线档案库汇集了目标鸟类的鸣唱录音。为在合理的标注时间内获得具有代表性的鸟类声音数据集，首先根据以下元数据条件，为每个物种最多获取 100 个文件：(1) 鸣唱类型；(2) 质量等级较高（A 或 B）；(3) 时长在 20 秒至 180 秒之间。随后为每个物种随机选取 20 个文件。最后，从每个文件中首次出现鸣唱的起始处选取 20 秒并进行标注。

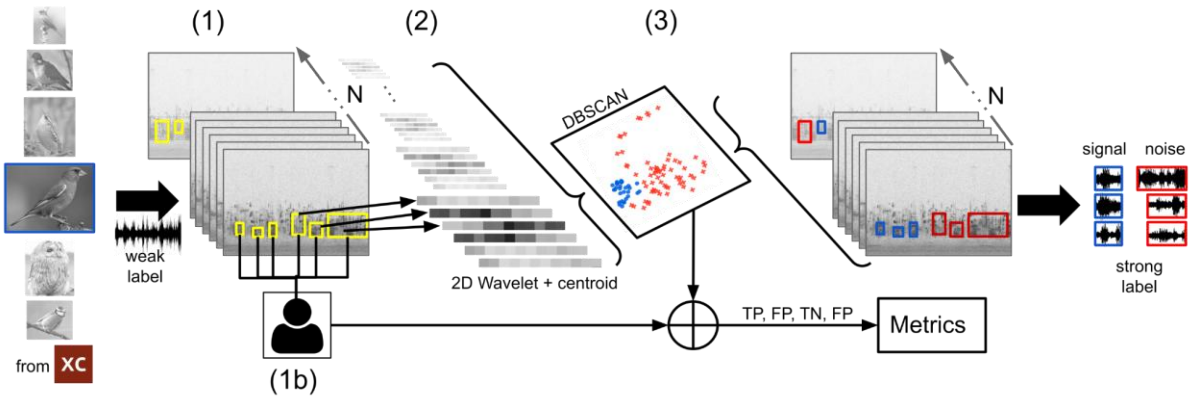

图 1. 从弱标签到强标签：标注函数的工作流程。从 Xeno-Canto 数据库收集目标物种的 N 条弱标注音频录音后，标注函数的工作流程包括：(1) 将每段音频录音分割为时频声音单元；(2) 计算每个声音单元的 49 个特征（二维小波系数（n = 48）+ 频谱质心（n = 1））；(3) 将所有声音单元聚类为信号

使用 DBSCAN 将声音分为目标物种的鸟鸣或噪声（其他所有声音）。最后，为每个声音单元赋予一个强标签。同时，专家会对所有声音单元进行标注（1b），以计算用于评估标注函数性能的指标。TP 代表真正例，FP 代表假正例，TN 代表真负例，FP 代表假正例

2.2. 感兴趣区域（ROI）的自动分割

2.2.2. 预处理

以 44.1 kHz 的采样率载入 MP3 音频文件。应用 100 Hz 至 18,000 Hz 的带通滤波器，以去除低频背景噪声以及由 MP3 压缩导致的高频伪影。使用包含 2,048 个采样点的傅里叶窗、1,024 个采样点的重叠部分和汉宁窗计算短时傅里叶变换（STFT）（图 2a）。随后对生成的声谱图在时间和频率维度上进行缩放，使紧邻的像素合并在一起（图 2b）。通过直接在声谱图上估计平稳噪声，部分去除背景噪声。针对每个频带，估计时间轴上的平均振幅，以获得沿频率轴的原始噪声轮廓。然后，使用跨越 25 个频带的滑动平均对噪声轮廓进行平滑。最后，从缩放后的整体声谱图中减去噪声轮廓（Towsey，2013年），并将生成的声谱图的振幅值转换为分贝（图 2c）。

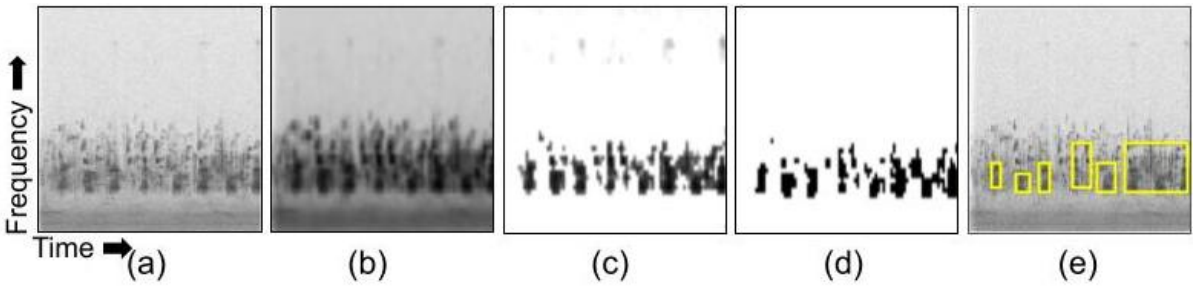

图 2. 声谱图预处理与分割。首先将音频文件转换为声谱图 (a)，然后调整声谱图的时间和频率尺寸 (b)。之后，从声谱图中减去噪声轮廓，并将数值转换为分贝 (c)。此外，使用滞后阈值二值化方法对声谱图进行二值化 (d)。最后一步是划定边界框 (e)。

2.2.3. 分割

在声谱图上进行分割时，围绕显著声音划定时频矩形边界框。这些感兴趣区域（感兴趣区域）的划定采用了一种称为滞后阈值二值化的区域生长分割方法（Ulloa 等，2021）。该方法基于两个分贝阈值：高阈值用于选取高强度像素，低阈值则用于合并相邻像素（图 2d）。在时间上间隔小于 0.24 s 且间隔

频率上相差小于 170 Hz 的边界框会被合并。持续时间少于 0.36 s 的边界框不予考虑，因为它们太短，不足以构成一个鸟鸣单元（图 2e）。

2.2.4. 优化与指标

预处理和分割过程取决于四个参数：声谱图的时间和频率缩放因子，以及滞后阈值二值化的两个阈值。在部分文件上估计这四个参数的最佳组合。使用 Audacity 时频标注工具，在声谱图视图中对这些文件手动划定边界框（Audacity v3.1.3，2021年）。随后，针对每个物种，在特定标注的 20 秒音频片段上，针对四个参数的多种组合应用自动提取流程。时间缩放因子范围为 6 至 18，频率缩放因子范围为 4 至 30，上阈值范围为 10 至 40 分贝，下阈值范围为 7 至 37 分贝。

使用交并比（IOU）测量手动划定与自动划定之间的距离；交并比是一种衡量两个边界框之间面积重叠的指标（Zou 等，2019年）。针对每种参数组合和每个文件分别计算 IOU 的中位数，然后再针对某一物种的所有文件计算中位数。随后选取所有物种中最佳的 IOU 得分，以获得最佳参数组合。

确定四个参数的最佳组合后，根据感兴趣区域的边界，自动提取所有物种20个文件中的20秒音频片段里存在的全部感兴趣区域。感兴趣区域在时间上由垂直边界划定，在频率上由水平边界划定。随后，首先围绕感兴趣区域的起止时间裁剪音频录音，然后应用一个十五阶带通巴特沃斯滤波器来提取声音，其较低和较高截止频率分别对应感兴趣区域的较低和较高频谱边界。

2.3. 感兴趣区域的自动分类

2.3.1. 特征

无监督分类是在声谱图上进行的。不过，声谱图的分辨率和带通滤波器与用于感兴趣区域分割的设置不同。采样率设为 24 kHz，傅里叶窗取 512 个采样点，带通滤波器的范围为 250 Hz 至 11 kHz。无监督分类基于两个信号描述符。第一个描述符是由二维小波分析得到的数值向量（Ulloa 等人，2018年）。对感兴趣区域应用一系列不同尺度的 Gabor 滤波器进行卷积，以提取符合所选分辨率的时频

特征。该滤波器组由 48 个 Gabor 滤波器组成，包含两个垂直滤波器、两个水平滤波器，以及四个对角滤波器；对角滤波器分别采用 45° 和 135° 两种角度，并设有六种不同尺度。第二个描述符基于傅里叶分析，通过测量每个感兴趣区域的频谱质心得到。因此，用于描述一个感兴趣区域的特征向量维度为 49 × 1。每个感兴趣区域都对应一个向量。

2.3.2. 聚类

鸟鸣和噪声的分类采用名为基于密度的带噪声空间应用聚类（DBSCAN）的无监督方法完成（Ester 等人，1996）。我们选择无监督方法，是为了让该工作流程能够应用于任何拥有足够音频录音的鸟类物种，而无需经过学习阶段。基于基于密度的簇定义，DBSCAN 无需预先设定簇的数量，并且适用于非球形簇。此外，DBSCAN 能够识别可视为噪声的离群点，而这是标注函数的先决条件。DBSCAN 从任意点 g 开始搜索核心样本，并检索所有距离小于欧几里得距离（Eps）且其邻域内至少有 MinPts 个点的点。若样本不属于任何簇，则将其定义为噪声。Eps 和 MinPts 并无理想取值，因此设定并不容易（Schubert 等人，2017）。我们使用 K-距离图为每个物种单独估算 Eps；该图按升序绘出每个点与其 k 个最近邻之间的平均距离。拐点（或“膝点”）表示簇中核心点（信号）的最大距离与簇中的离群点（噪声）之间的界限。我们使用名为 Kneed 的膝点定位算法自动检测该点（Satopaa 等，2011）。

通过在训练集上对每个物种运行聚类算法，并将 MinPts 设置为感兴趣区域数量的 1% 至 50 % 之间的不同数值，确定了 MinPts。针对每个百分比，计算了所有物种的宏平均精确率。最终，所有数据集的 MinPts 均确定为 10%。

为评估分类结果，我们将属于该簇的点标注为正样本，将离群点标注为负样本。当存在多个簇时，我们仅保留最大的簇，因为我们推测其他较小的簇可能是其他物种的鸟鸣、汽车交通噪声等人为噪声，或属于目标物种的异常鸟鸣。在已知假正例、真正例、假负例和真负例的情况下，我们采用机器学习中常用的三个性能指标来评估该自动化流程：(1) 精确率的宏平均（真正例数量除以真正例与真负例数量之和），(2) 宏平均

召回率（真正例数量除以真正例与假负例数量之和），以及 (3) F1 分数的宏平均（精确率和召回率的组合）。最后，我们引入了另外两个指标。初始噪声百分比是指被标记为噪声的感兴趣区域数量占感兴趣区域总数的比例。最终噪声百分比是指假正例数量占真正例与假正例数量之和的比例，相当于 1-精确率。

2.3.3. 人工标注

由自动分割得到的 5,315 个感兴趣区域由我们中的一人（MLC）使用 Audacity 的声谱图视图进行人工标注，分为“信号”（部分或完整的鸟鸣）或“噪声”（不是鸟鸣、鸟鸣与其他声音重叠，或鸟鸣片段过短而无法由人类专家识别）。

2.4. 使用 BirdNET 进行监督分类

为了将 DBSCAN 的输出与有监督方法进行比较，我们使用卷积神经网络（CNN）BirdNET-Lite（Kahl 等，2021）对鸟鸣进行分类。BirdNET 专为对全球 984 种鸟类进行鸟类声音分类而设计，被认为是最先进的鸟类音频识别工具。BirdNET 的地理位置设置为法国，季节设置为春季（儒略历第十五周）。我们保留了默认参数：灵敏度为 1，最低置信度为 0.1。不同长度的感兴趣区域通过补零处理，使其长度一致。随后将这些感兴趣区域输入卷积神经网络进行分类。

2.5. 实现

两种方法均可在个人计算机或虚拟机上运行。对于后者，在配备单个 2.2 GHz 时钟频率核心和 13 GB RAM 的情况下，处理验证数据集（包含 22 个物种的 20 个音频文件）所需的计算时间为：从 Xeno-Canto 下载文件用时 7 分钟，提取感兴趣区域用时 6 分钟，使用 DBSCAN 对感兴趣区域进行分类（特征提取和聚类）用时 13 分钟，使用 BirdNET 对感兴趣区域进行分类用时 25 分钟。

整个流程使用 Python 语言编写，采用了 scikit-maad v1.3.0（Ulloa 等，2021）、scikit-learn v0.23.2（Pedregosa 等，2011）、librosa v0.8.0（McFee 等，2015）和 kneed v0.8.1（Satopaa 等，2011）这些软件包。完整代码以 Python 软件包的形式发布在 GitHub（https://github.com/ear-team/bam）。

3. 结果

3.1. 音频文件

从 Xeno-Canto 中为每个物种获取了 880 个时长为 20 秒的音频文件（20 个音频文件 * 44 个物种），共计 4 小时 53 分 20 秒的录音。

3.2. 感兴趣区域 (感兴趣区域)

用于感兴趣区域自动分割的四个参数的最佳组合为：声谱图的时间缩小因子和频率缩小因子分别为 10 和 15；滞后阈值二值化的第一和第二阈值分别为 37 分贝和 33 分贝。

训练集共自动提取了 1,855 个边界框。每个物种的感兴趣区域平均数量为 84 个，其中乌鸫 (Strix aluco) 最少，为 41 个；槲鸫 (Turdus viscivorus) 最多，为 134 个。标注函数估算的每个物种的噪声中位数百分比为 36.5%，但不同物种之间差异很大（表 1，表A.1）。斑鸫 (Turdus philomelos) 的初始噪声量最高（78%），乌鸫 (Strix aluco) 最低（15%）。

验证集包含从与训练集相同的 22 个物种中提取的 1,732 个感兴趣区域。欧洲绿啄木鸟（Picus viridis）的感兴趣区域数量最少（40 个），大杜鹃（Cuculus canorus）的数量最多（162 个）。每个物种的噪声百分比中位数为 35.5%（表 1、表A.1.）。斑鸫（Turdus philomelos）的初始噪声占比最高（83%），而欧洲金翅雀（Chloris chloris）的最低（13%）。

测试集由另外 22 个物种组成，包含 1,728 个感兴趣区域。棕扇尾莺（Cettia cetti）的感兴趣区域数量最少（48 个），蓝山雀（Cyanistes caeruleus）的数量最多（123 个）。每个物种集合的噪声百分比中位数为 48%（表 1、表A.2.）。红腹灰雀（Pyrrhula pyrrhula）的初始噪声占比最高（82%），而黑红尾鸲（Phoenicurus ochruros）的最低（10%）。

3.3. 聚类

对于训练集，所有物种的噪声中位数均有所下降（图 3a），从 36.5% 降至 12.5%。验证集也出现了类似的下降，从

35.5% 降至 13.5%（图 3b，表 1）。训练集最终噪声的离散程度高于验证集，这一点可从去噪后更宽的四分位距看出。

		训练	验证	Test
	初始噪声中位数	36.5%	35.5%	48.0%
DBSCAN 聚类	最终噪声中位数	12.5%	13.5%	33.5%
	宏平均精确率	0.81	0.82	0.67
	宏平均召回率	0.65	0.72	0.68
	宏平均 F1 分数	0.67	0.73	0.65
BirdNET	最终噪声中位数	14.0%	14.5%	23.0%
	宏平均精确率	0.81	0.81	0.75
	宏平均召回率	0.82	0.83	0.74
	宏平均 F1 分数	0.81	0.81	0.71

表 1. 三个数据集各标注函数的全局得分。当质量指标的取值范围为 0 到 1 时，噪声水平以百分比表示。

应用 DBSCAN 聚类后，训练集包含两个已完全去噪的物种（大杜鹃 [C. canorus] 和乌鸫 [S. aluco]），验证集包含三个同样已完全去噪的物种（大杜鹃 [C. canorus], 、乌鸫 [S. aluco], 和长尾山雀 [Aegithalos caudatus]）。其中，一个物种斑鸫（T. philomelos）仍然具有较高的最终噪声，训练数据集中的数值为 76%，验证数据集中的数值为 77%。该物种也是初始噪声量最高的物种（训练集为 78%，验证集为 83%）。

对于测试集（由训练集和验证集中未包含的 22 种鸟类物种组成），噪声也有所减少，从 48% 降至 33.5%（图 3c，表 1）。与训练集和验证集相比，噪声减少幅度较小，这与最终噪声分布更分散有关。经过 DBSCAN 聚类后，只有草地鹨（Anthus pratensis）录音中的噪声能够降至零。小鸮（Athene noctua）的噪声从 58% 增至 75%，戴菊（Regulus regulus）的噪声则从 44% 增至 53%。蓝山雀（C. caeruleus）、芦鹀（Emberiza schoeniclus）和红腹灰雀（Pyrrhula pyrrhula）的最终噪声得分较高，分别为 79%、64% 和 70%。

用于估算数据集纯净程度的宏平均精确率，在训练数据集中高于 0.81，在验证数据集中高于 0.82；而在测试数据集中则降至 0.67（表 1）。三个数据集的宏平均召回率均低于 0.8，介于训练集的 0.65 和验证集的 0.72 之间（表 1）。宏平均 F1 分数介于测试集的 0.65 和验证集的 0.73 之间。

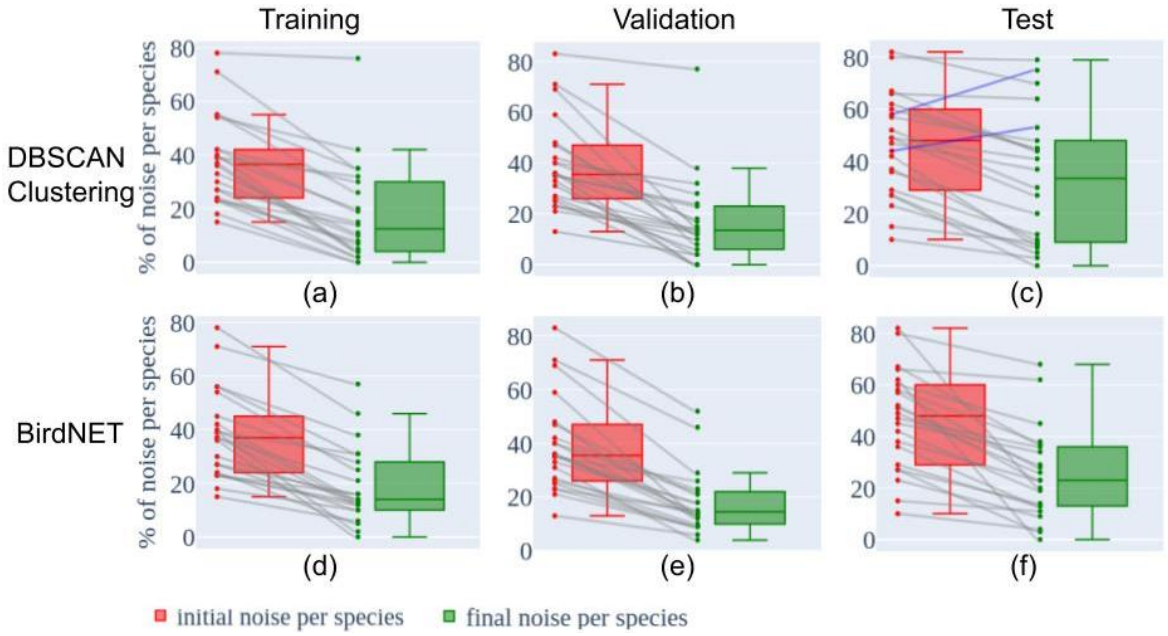

图3. 对三个数据集应用 DBSCAN 无监督聚类（a、b、c）或 BirdNET 监督分类（d、e、f）时的初始和最终噪声量。初始噪声百分比对应于被标记为噪声的感兴趣区域数量占感兴趣区域总数的比例。最终噪声百分比对应于假正例数量（即被自动判定为信号、但实际上是噪声的感兴趣区域）占真正例与假正例数量之和的比例。它也可以表示为 100% – 精确率（即真正例数量占真正例与假正例数量之和的比例）。每个箱体表示中位数（中心线）以及 25%（Q1）和 75% 四分位数（Q3）（箱体边界）；而对于下须线和上须线，须线分别表示低于和高于四分位距（Q3–Q1）1.5 倍范围内的最后一个样本点。为准确呈现分布，每个录音均以点表示，并以连接各处理方法的线相连。

为了解初始噪声量对聚类质量的影响，我们将初始噪声量与最终噪声量进行对比（图4a）。初始噪声超过 40% 的物种，噪声减少幅度较小，尤其是使用聚类标注函数时。对于初始噪声低于 40% 的物种，DBSCAN 聚类比 BirdNET 更有效，因为更多数据点的最终噪声接近 0%。聚类后噪声达到 0% 的所有物种，其初始噪声量均低于 40%。

对聚类结果进行多项式回归（图4a）后发现，对于初始噪声量在10%至35%之间的物种，聚类算法的去噪效果尤为显著，因为这些物种的最终噪声量均为

约5%。超过35%后，去噪效果则不那么理想，因为多项式近似曲线更接近 y = x 直线。

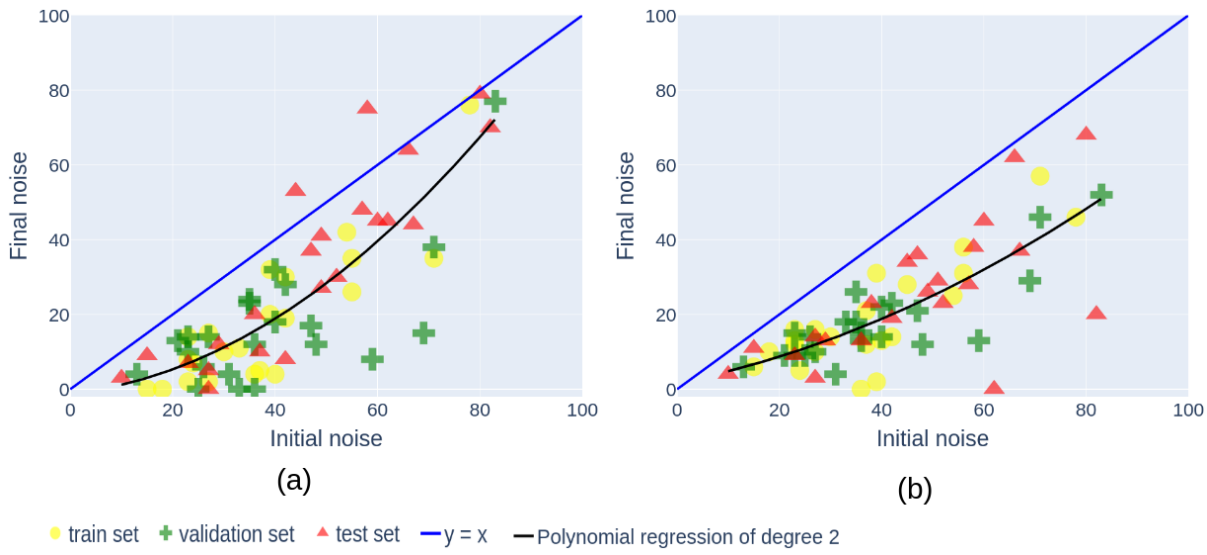

图4. 使用无监督 DBSCAN聚类方法（a）和 BirdNET 算法（b）比较初始噪声与最终噪声。每个点代表一个物种。y = x 蓝线表示最终噪声量等于初始噪声量的情况。黑线是二次多项式回归，表示总体趋势。

3.4. BirdNET

BirdNET 得到的最终噪声中位数高于 DBSCAN 聚类算法的结果：训练集和验证集分别为 14 % 和 14.5%；而测试集的最终噪声中位数较低，为 23%（图 3d–f，表 1）。在三个数据集中，只有两个物种被完全降噪 [欧洲绿啄木鸟（P. viridis] 和紫翅椋鸟 [Sturnus vulgaris]）。与 DBSCAN 聚类不同，预计 BirdNET 分类效率不受每个物种初始噪声量的影响，这一点也得到了初始噪声与最终噪声之间准线性关系的证实（图4b）。

训练集和验证集的宏平均精确率均高于 0.8，低于 DBSCAN 聚类的结果。测试集的宏平均精确率为 0.75，高于 DBSCAN 聚类的结果（表 1）。宏平均召回率介于测试集的 0.74 和验证集的 0.84 之间，高于 DBSCAN 聚类的结果。BirdNET 的召回率高于 DBSCAN 聚类。

3.5. 聚类与 BirdNET

比较了两种方法在所有数据集上得到的最终噪声（图 5）。初始噪声低于 20% 时，DBSCAN 聚类的表现更好。初始噪声介于 20% 到 40% 之间时，两种方法的表现相同。对于初始噪声超过 40% 的物种，BirdNET 标注函数取得了更好的结果。

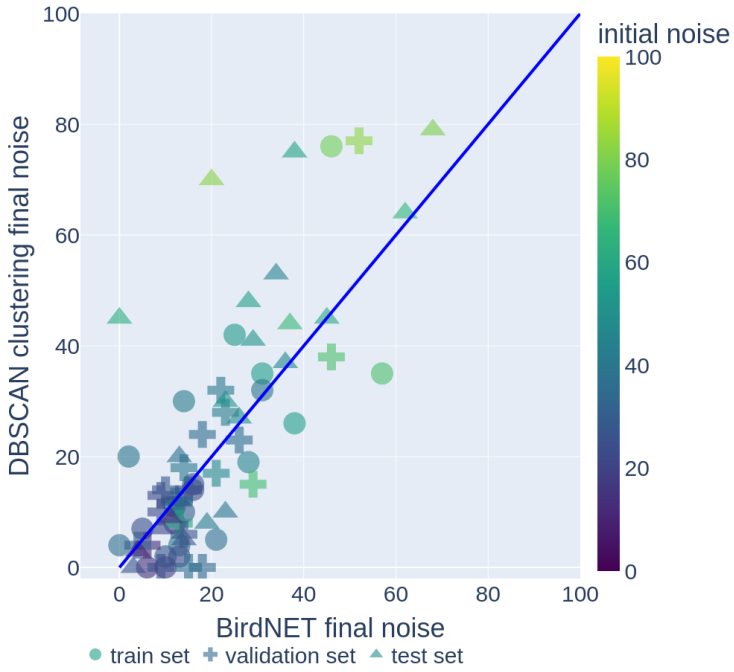

图 5. 与初始噪声量相比，DBSCAN 聚类和 BirdNET 算法减少噪声的效果。对于位于 y = x 蓝线下方的物种，DBSCAN 聚类表现更佳；对于位于 y = x 蓝线上方的物种，BirdNET 表现更佳。

4. 讨论

在这项工作中，我们设计了一种专门的标注函数，通过最大限度地减少标签噪声来提升 Xeno-Canto 鸟鸣数据集的质量。该标注函数分为三个步骤：1）声谱图时频分割；2）特征计算；3）使用无监督（DBSCAN）或有监督（BirdNET）算法进行噪声或信号分类。我们将评述这一新标注函数的性能，讨论工作流程各步骤的优点与不足，并最终提出可能的改进方案。

4.1. 标注函数的性能

通过对目标鸟类录音进行人工标注，我们首次能够评估自动标注产生的标签噪声。分段声音中与目标物种不符的比例介于 10% 和 80% 之间；黑红尾鸲的比例为 10%，而斑鹟

（P. ochruros）、斑鸫（T. philomelos）、红腹灰雀（P. pyrrhula）和蓝山雀（C. caeruleus）的比例则高达 80%。在所有数据集中，44 个物种中有 16 个受到超过 50% 的标签噪声影响。标签噪声水平如此之高，可能有以下原因：1）不同音频设备（从超定向到全向麦克风）造成的固有差异；2）其他声源（包括其他鸟类物种）导致的背景噪声水平不同；或 3）鸟鸣时频结构的多样性（Kershenbaum 等，2016）。

标注函数最多可将标签噪声的中位数降低至原来的三分之一，但不同物种和数据集之间的差异很大。综合考虑所有数据集，使用无监督方法（DBSCAN）时，44 个物种中有 25 个的最终标签噪声低于 25%；使用有监督方法（BirdNET）时，这一比例略有上升，44 个物种中有 31 个达到这一水平。无论采用何种分类方法，标注函数在测试集上的表现都显著低于训练集和验证集。这种性能差异可能是因为分割导致测试集的初始标签噪声水平高于训练集和验证集。由于分割是针对训练集中各物种的鸟鸣进行优化的，调优后得到的超参数可能过于特定，因此导致该流程过拟合。

DBSCAN 聚类的效果因初始标签噪声水平而异。具体而言，初始标签噪声低于 20% 时，DBSCAN 的表现优于 BirdNET；高于 40% 时，则不如 BirdNET。与 BirdNET 不同，DBSCAN 的表现取决于数据集中标签噪声所占的比例。如果初始标签噪声水平超过某个临界值，MinPts 和 Eps 的自适应选择可能会受到一定数量声音的影响而产生偏差，而这些声音并不能代表目标声音（Latifi-Pakdehi 和 Daneshpour，2021）。初始标签噪声水平较高，也可能增加将相似音频样本聚合到同一聚类中的概率，而这些样本并不属于目标物种。

由于声学单元在时间和频率上的多样性很高，鸟鸣的信号分割和分类颇具挑战性（Kershenbaum 等，2016；Priyadarshani 等，2018）。对于蓝山雀（C. caeruleus）或戴菊（戴菊）等鸣声频率较高的鸟类，标签噪声的程度尤其重要。在这种情况下，由于衰减强烈，信噪比预计会较低，因此感兴趣区域不够突出，也更难界定。对于发出短促鸣声的物种（如红腹灰雀（P. pyrrhula）），或像斑鸫（T. philomelos）那样发出长串音符的物种，结果也较差。室外传播造成的声学混响和衰减也会模糊歌声的时间结构，使快速事件尤其难以检测。不同物种之间的相对成功率表明，我们的标注函数中的声学和聚类超参数可能需要针对以下用途进行调整

以适用于其他情境，包括其他鸟类物种，也可能包括哺乳动物、两栖动物或昆虫发出的声音。

4.2. 模块化标注函数

我们的标注函数采用模块化设计。从分割到聚类，各个部分彼此独立，可根据其他需求和情境进行调整。声音分割通常采用时域波形上的单一阈值（Fodor，2013年；Lasseck，2013年；Potamitis，2014年）。我们采用的分割方法是在声音的时频表示上应用双阈值过程。双阈值过程可以选取不一定最显著的声音，从而仍能检测到中距离的声音。与 Sprengel 等人（2016年）设计的检测器相比——该检测器被广泛用于构建训练深度学习模型（如 BirdNET）的鸟鸣数据集（Kahl 等，2021年）——我们的方法不仅在时间维度上分割感兴趣的信号，也在频率维度上进行分割，以便将音色纳入考量。这一策略使我们能够识别两个在时间上重叠、但频率上不重叠的感兴趣声音。我们的分割基于传统图像处理方法，但近年来，图像深度学习解决方案得到了广泛发展。这些颇具吸引力的方案包括时间卷积网络（TCNs）（Bai 等，2018年；Steinfath 等，2021年）、图像目标检测器架构（如 Faster R-CNN 或 YOLO）（Coffey 等，2019年；Kong 等，2017年；Narasimhan 等，2017年；Shrestha 等，2021年；Venkatesh 等，2022年；Wu 等，2022年；Zsebők 等，2019年），以及音源分离方法，例如说话人日志（Denton 等，2022年；Morfi 和 Stowell，2021年）。然而，所有分割方法都需要大量标注数据和高额计算资源；在博物学家专业知识日渐消失、且必须限制能源消耗的当下，这两点构成了主要制约。

为了描述分割后的声音，我们采用了基于 Ulloa 等人（2018年）提出的二维小波分析的手工设计特征。与日益流行的梅尔频率倒谱系数（MFCC）不同，小波变换不以人为中心，因此更适合描述动物发声（Zottesso 等人，2018年）。此外，未经过学习的二维小波核与卷积神经网络（CNN）前几层学习到的卷积核有着惊人的相似性（Meng 等人，2019年；Yosinski 等人，2014年）。

与其使用手工设计特征，另一种方法是使用学习特征，即来自已在动物发声上训练的深度学习模型的所谓深度音频嵌入（Goffinet 等，2021年；Stoumpou 等，2022年）。沿着这一思路，近期一项研究成功测试了多种学习音频潜在表示的方法

，以将横斑林鸮（Strix varia）、灰鹤（Grus grus）和普通潜鸟（Gavia immer）的鸣声与噪声分离开来（Tolkova 等，2021年）。这些方法包括一个卷积自编码器和两个预训练网络：VGGish 模型（Hershey 等，2017年）以及 Wavegram-Logmel-CNN16 架构（Kong 等，2020年）。这两个卷积神经网络均在 Google AudioSet 数据集上进行了预训练；该数据集是从 YouTube 视频中收集的大型声音标注集合（Gemmeke 等，2017年），包含 600 多个声音类别，但只有一个类别是“鸟类”。另一种方案是使用 BirdNET 提供的 512维嵌入。但无论采用哪种方法，都必须特别注意根据每个物种的音频样本数量调整嵌入向量的维度，以避免遭遇维数灾难（Assent，2012年）。深度音频嵌入前景可期，但其缺乏可解释性，且仍需通过使用动物信号和声景丰富训练数据集来加以改进。

最后，近期研究提出使用统一流形逼近与投影（UMAP）等非线性无监督降维方法（McInnes et al., 2020），直接从声谱图生成潜在空间表示。随后直接对 UMAP 生成的嵌入进行聚类（Sainburg et al., 2020; Thomas et al., 2021）。尽管这种方法看起来非常直接，但仍需要在不同鸣叫/歌声之间进行非常精确的起始点检测，以避免类内差异（Thomas et al., 2021）。

标注函数的最后一个模块是使用 DBSCAN 聚类将目标声音与噪声分离。DBSCAN 及其他基于密度的非参数聚类算法此前已成功应用于海洋哺乳动物的其他任务，例如将蓝鲸（Balaenoptera musculus）的鸣叫与噪声分离（Cuevas et al., 2017）、无监督分类虎鲸（Orcinus orca）的发声（Poupard et al., 2019），以及识别抹香鲸的尾声音型库（Amorim et al., 2020; Gero et al., 2016）。最近，DBSCAN 的一种层次变体 HDBSCAN（McInnes et al., 2017）已用于构建狐獴（Suricata suricatta）（Thomas et al., 2021）、角雕（Harpia harpyja）和巨獭（Pteronura brasiliensis）（Schneider et al., 2022）的发声 repertoire。迄今为止，仅有三种基于密度的非参数聚类算法（DBSCAN、HDBSCAN 和 OPTICS [Ankerst et al.,1999]）在动物发声方面经过测试，但近期还设计了许多其他变体（Latifi- Pakdehi and Daneshpour, 2021），或可加以改进，用于设计更高效的标注函数。

5. 展望与结论

生物声学和生态声学在理解动物行为与生态学方面日益取得成功，这主要得益于大规模录音档案的收集与自动分析。如今，借助个人录音设备和自主录音单元进行野外录音已成为可能，但要自动检索出感兴趣的声音以供进一步

描述仍然具有挑战性。在此，我们利用规模最大的鸟鸣音频数据集，展示了基于时频分析和聚类方法的标注函数能够帮助筛选合适的信号，从而提高所收集数据的质量。仍需进一步开发，以便将此类工具应用于声景录音，因为其中混合了多个物种的感兴趣声音。

参考文献

Algan, G., Ulusoy, I., 2021年。存在噪声标签时使用深度学习进行图像分类：综述。Knowl.-Based Syst. 215, 106771。https://doi.org/10.1016/j.knosys.2021.106771 Amorim, T.O.S., Rendell, L., Di Tullio, J., Secchi, E.R., Castro, F.R., Andriolo, A., 2020。西大西洋抹香鲸的尾声音型库与发声族群。Deep Sea Res. Part Oceanogr. Res. Pap. 160, 103254。https://doi.org/10.1016/j.dsr.2020.103254 Ankerst, M., Breunig, M.M., Kriegel, H.-P., Sander, J., 1999。OPTICS：通过对点排序识别聚类结构。美国计算机协会 SIGMOD Rec. 28, 49–60。https://doi.org/10.1145/304181.304187 Arpit, D., Jastrzębski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y., 2017。在：国际机器学习会议中，深入探究深度网络中的记忆现象。PMLR，第233–242页。Arthur, B.J., Ding, Y., Sosale, M., Khalif, F., Kim, E., Waddell, P., Turaga, S.C., Stern, D.L., 2021。Songexplorer：用于发现和分割动物声学通信信号的深度学习工作流程。bioRxiv。Assent, I., 2012。高维数据聚类。WIREs Data Min. Knowl. Discov. 2, 340–350。https://doi.org/10.1002/widm.1062 Audacity, T., 2021。Audacity (R)：免费音频编辑器和录音机。Bai, S., Kolter, J.Z., Koltun, V., 2018。用于序列建模的通用卷积网络和循环网络的实证评估。https://doi.org/10.48550/arXiv.1803.01271 Brodley, C. E., & Friedl, M. A., 1999。识别错误标记的训练数据。《人工智能研究杂志》11, 131-167。Chen, P., Liao, B.B., Chen, G., Zhang, S., 2019。在：国际机器学习会议中，理解并利用通过噪声标签训练的深度神经网络。PMLR，第1062–1070页。Coffey, K.R., Marx, R.G., Neumaier, J.F., 2019。DeepSqueak：一种用于检测和分析超声发声的深度学习系统。Neuropsychopharmacology 44, 859–868。https://doi.org/10.1038/s41386-018-0303-6 Cohen, Y., Nicholson, D.A., Sanchioni, A., Mallaber, E.K., Skidanova, V., Gardner, T.J., 2022。使用分割声谱图的神经网络自动标注鸟鸣。eLife 11, e63853。https://doi.org/10.7554/eLife.63853 Cuevas, A., Veragua, A., Español-Jiménez, S., Chiang, G., Tobar, F., 2017。在：2017年智利电气、电子工程、信息与通信技术会议（CHILECON）中，利用多个时频特征进行无监督蓝鲸叫声检测。发表于2017年智利电气、电子工程、信息与通信技术会议（CHILECON），第1–6页。https://doi.org/10.1109/CHILECON.2017.8229663 Denton, T., Wisdom, S., Hershey, J.R., 2022。在：ICASSP 2022-2022 IEEE声学、语音与信号处理国际会议中，利用无监督声音分离改进鸟类分类。IEEE，第636–640页。Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., Sueur, J., 2012。利用声学指数监测动物多样性：在温带林地中的应用。《生态指标》13, 46–54。https://doi.org/10.1016/j.ecolind.2011.05.006 Donmez, P., Carbonell, J.G., 2008。主动式学习：使用多个不完美预言机进行成本敏感主动学习。在：第17届美国计算机协会信息与知识管理会议论文集，第619–628页。Dufourq, E., Batist, C., Foquet, R., Durbach, I., 2022。被动声学监测

利用迁移学习监测动物种群。《生态信息学》70, 101688。https://doi.org/10.1016/j.ecoinf.2022.101688 Ester, M., Kriegel, H.-P., Xu, X., 1996。用于在大型空间数据库中发现含噪声簇的基于密度的算法 6。Fodor, G., 2013。第九届年度机器学习与信号处理竞赛冠军作品。2。Frenay, B., Verleysen, M., 2014。存在标签噪声时的分类：综述。IEEE Trans. Neural Netw. Learn. Syst. 25, 845–869。https://doi.org/10.1109/TNNLS.2013.2292894 Fukuzawa, Y., Webb, W.H., Pawley, M.D.M., Roper, M.M., Marsland, S., Brunton, D.H., Gilman, A., 2020。Koe：用于对动物发声中的声学单元进行分类并分析序列结构的网页软件。Methods Ecol. Evol. 11, 431–441。https://doi.org/10.1111/2041-210X.13336 GBIF, 2021。全球生物多样性信息机构。https://www.gbif.org/（访问日期：2021-10-06）。Gemmeke, J.F., Ellis, D.P.W., Freedman, D., Jansen, A., Lawrence, W., Moore, R.C., Plakal, M., Ritter, M., 2017。Audio Set：音频事件本体与人工标注数据集。在：2017年IEEE声学、语音与信号处理国际会议（ICASSP）中。发表于2017年IEEE声学、语音与信号处理国际会议（ICASSP），第776–780页。https://doi.org/10.1109/ICASSP.2017.7952261 Gero, S., Whitehead, H., Rendell, L., 2016。抹香鲸尾音序列中的个体、单元和发声族群层级身份线索。R. Soc. Open Sci. 3, 150372。https://doi.org/10.1098/rsos.150372 Gil-González, J., Valencia-Duque, A., Álvarez-Meza, A., Orozco-Gutiérrez, Á., García-Moreno, A., 2021。面向多个标注者的正则化链式深度神经网络分类器。Appl. Sci. 11, 5409。Goffinet, J., Brudner, S., Mooney, R., Pearson, J., 2021。学习到的低维特征空间量化发声 repertoire 中的个体与群体差异。eLife 10, e67855。https://doi.org/10.7554/eLife.67855 Hershey, S., Chaudhuri, S., Ellis, D.P.W., Gemmeke, J.F., Jansen, A., Moore, R.C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., Slaney, M., Weiss, R.J., Wilson, K., 2017。面向大规模音频分类的卷积神经网络架构。在：2017年IEEE声学、语音与信号处理国际会议（ICASSP）中。发表于2017年IEEE声学、语音与信号处理国际会议（ICASSP），第131–135页。https://doi.org/10.1109/ICASSP.2017.7952132 Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.-P., Planqué, R., Rauber, A., Palazzo, S., Fisher, B., 2015。LifeCLEF 2015：多媒体生命物种识别挑战。在：欧洲语言跨语言评测论坛国际会议中。Springer，第462–483页。Joly, A., Goëau, H., Kahl, S., Picek, L., Lorieul, T., Cole, E., Deneu, B., Servajean, M., Durso, A., Bolon, I., Glotin, H., Planqué, R., de Castañeda, R.R., Vellinga, W.-P., Klinck, H., Denton, T., Eggel, I., Bonnet, P., Müller, H., 2021。LifeCLEF 2021概述：机器学习物种识别与物种分布预测评测。在：Candan, K.S., Ionescu, B., Goeuriot, L., Larsen, B., Müller, H., Joly, A., Maistro, M., Piroi, F., Faggioli, G., Ferro, N.（编），《实验性信息检索与多语言、多模态和交互》，《计算机科学讲义》。施普林格国际出版，查姆，第371–393页。https://doi.org/10.1007/978-3-030-85251-1_24 Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021。BirdNET：用于鸟类多样性监测的深度学习解决方案。《生态信息学》61, 101236。https://doi.org/10.1016/j.ecoinf.2021.101236 Karimi, D., Dou, H., Warfield, S.K., Gholipour, A., 2020。含噪声标签的深度学习：探索医学图像分析中的技术与补救方法。Med. Image Anal. 65, 101759。Kershenbaum, A., Blumstein, D.T., Roch, M.A., Akçay, Ç., Backus, G., Bee, M.A., Bohn, K., Cao, Y., Carter, G., Cäsar, C., Coen, M., DeRuiter, S.L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T.M., Garland, E.C., Gustison, M.,

Harlow, H.J., Tucker, R., Thomas, J.A., 2022。声学生态学中的时间尺度。《生态学趋势与进化》37，1102–1113。https://doi.org/10.1016/j.tree.2022.08.003 Hershey, S., Chaudhuri, S., Ellis, D.P.W., Gemmeke, J.F., Jansen, A., Moore, R.C., Plakal, M., Platt, D., Saurous, R.A., Seybold, B., Slaney, M., Weiss, R.J., Wilson, K., 2017。用于大规模音频分类的CNN架构。发表于2017年IEEE国际声学、语音与信号处理会议（ICASSP），第131–135页。https://doi.org/10.1109/ICASSP.2017.7952132 Huet, C., Gasc, A., Joly, A., Servajean, M., Bonnet, P., Glotin, H., Vellinga, W.-P., Planqué, R., Stöter, F.-R., Müller, H., 2017。LifeCLEF 2017鸟类检测挑战赛概述。发表于CLEF（工作笔记），第1–8页。Ioffe, S., Szegedy, C., 2015。批归一化：通过减少内部协变量偏移来加速深度网络训练。发表于国际机器学习会议，第448–456页。PMLR。Jiang, L., Cai, R., Wang, D., Bond, S., Chen, I., Chen, C., 2014。用于高维聚类的新型基于密度的算法。发表于2014年国际联合人工智能会议（IJCAI），第1–7页。IEEE。Joly, A., Goëau, H., Kahl, S., Deneu, B., Servajean, M., Cole, E., Picek, L., de Ruiz, R.A., Bolon, I., Durso, A., Kim, J., Steinmetz, R., Spytsina, A., Bonnet, P., Stöter, F.-R., Sawtell, N., Hogeweg, L., Klinck, H., Palmer, J., Joly, C., Mercado III, E., Narins, P.M., Piel, A., Rice, M., Salmi, R., Sasahara, K., Sayigh, L., Shiu, Y., Taylor, C., Vallejo, E.E., Waller, S., Zamora-Gutierrez, V., 2016年。非人类动物的声学序列：教程式综述与展望。《生物学评论》91，13–52。https://doi.org/10.1111/brv.12160 Kong, Q., Cao, Y., Iqbal, T., Wang, Y., Wang, W., Plumbley, M.D., 2020。PANNs：用于音频模式识别的大规模预训练音频神经网络。《IEEE/美国电气与电子工程师协会音频、语音与语言处理汇刊》28，2880–2894。https://doi.org/10.1109/TASLP.2020.3030497 Kong, Q., Xu, Y., Plumbley, M.D., 2017。用于弱标注鸟类音频检测的联合检测与分类卷积神经网络，载于：2017年第25届欧洲信号处理会议（EUSIPCO）。发表于2017年第25届欧洲信号处理会议（EUSIPCO），第1749–1753页。https://doi.org/10.23919/EUSIPCO.2017.8081509 Lasseck, M., 2013。野外录音中的鸟鸣分类：NIPS4B 2013竞赛获奖方案，载于：生物声学神经信息缩放国际研讨会论文集，Sabiod. Org/Nips4b，与NIPS联合举办，内华达州。第176–181页。Latifi-Pakdehi, A., Daneshpour, N., 2021。DBHC：基于DBSCAN的层次聚类算法。《数据与知识工程》135，101922。https://doi.org/10.1016/j.datak.2021.101922 Martin, K., Adam, O., Obin, N., Dufour, V., 2022。Rookognise：使用多任务神经网络在野外录音中检测并识别单只秃鼻乌鸦 41。McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., Nieto, O., 2015。librosa：Python中的音频与音乐信号分析。发表于Python科学会议，奥斯汀，得克萨斯州，第18–24页。https://doi.org/10.25080/Majora-7b98e3ed-003 McInnes, L., Healy, J., Astels, S., 2017。hdbscan：基于层次密度的聚类。《开放源代码软件杂志》2，205。https://doi.org/10.21105/joss.00205 McInnes, L., Healy, J., Melville, J., 2020。UMAP：用于降维的统一流形逼近与投影。https://doi.org/10.48550/arXiv.1802.03426 Meng, F., Wang, X., Shao, F., Wang, D., Hua, X., 2019。通过遗传算法训练方法优化神经网络中的节能Gabor核。《电子学》8，105 10 3390 8010105。https://doi.org/ . /electronics Morfi, V., Stowell, D., 2021。少样本生物声学事件检测：DCASE 2021挑战赛中的一项新任务。《IEEE多媒体汇刊》17，1733–1746。https://doi.org/10.1109/TMM.2015.2428998 Motamedi, M., Sakharnykh, N., Kaldewey, T., 2021。一种以数据为中心的方法：用更少数据训练深度神经网络。https://doi.org/10.48550/arXiv.2110.03613 Narasimhan, R., Fern, X.Z., Raich, R., 2017。使用卷积神经网络对鸟鸣进行同步分割与分类，载于：2017年IEEE声学、语音与信号处理国际会议（ICASSP）。发表于2017年IEEE声学、语音与信号处理国际会议（ICASSP），第146–150页。https://doi.org/10.1109/ICASSP.2017.7952135 Ng, A., Laird, D., He, L., 2021。以数据为中心的人工智能竞赛。DeepLearning AI。在线链接：https://https-deeplearning-ai. github.io/data-centric-comp/(。Nolasco, I., Singh, S., Vidana-Villa, E., Grout, E., Morford, J., Emmerson, M., Jensens, F., Whitehead, H., Kiskin, I., Strandburg-Peshkin, A., 2022。DCASE 2022挑战赛中的少样本生物声学事件检测。arXiv预印本arXiv220707911。PECBMS, 2021。泛欧洲普通鸟类监测计划。https://pecbms.info（访问日期：2021-10-01）。Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011。Scikit-learn：机器

学习在Python中的应用。《机器学习研究期刊》12，2825–2830。Phillips, Y.F., Towsey, M., Roe, P., 2018。通过聚类与可视化揭示长时环境音频录音中的生态内容。《PLOS ONE》13，e0193345。https://doi.org/10.1371/journal.pone.0193345 Planqué, B., Vellinga, W.-P., 2008。Xeno-canto：欣赏新热带区鸟鸣的21世纪方式。《新热带鸟类》3，17–23。Potamitis, I., 2014。对野外录制的丰富分类群群落进行自动分类。《PLOS ONE》9，e96936。https://doi.org/10.1371/journal.pone.0096936 Poupard, M., Best, P., Schlüter, J., Symonds, H., Spong, P., Lengagne, T., Soriano, T., Glotin, H., 2019。虎鲸发声的大规模无监督聚类：描述虎鲸通信系统的模型，载于：第二届人类、动物与机器人之间语音互动国际研讨会。伦敦，英国。https://doi.org/10.7287/peerj.preprints.27979v1 Priyadarshani, N., Marsland, S., Castro, I., 2018。复杂声学环境中的自动鸟鸣识别：综述。《鸟类生物学杂志》49，jav-01447。https://doi.org/10.1111/jav.01447 Ratner, A.J., De Sa, C.M., Wu, S., Selsam, D., Ré, C., 2016。数据编程：快速创建大型训练集。《神经信息处理系统进展》29。Rolnick, D., Veit, A., Belongie, S., Shavit, N., 2017。深度学习能够抵御大规模标签噪声。arXiv预印本arXiv170510694。Ruff, Z.J., Lesmeister, D.B., Appel, C.L., Sullivan, C.M., 2021。用于自动识别动物声音的工作流程与卷积神经网络。《生态指标》124，107419。https://doi.org/10.1016/j.ecolind.2021.107419 Sainburg, T., Thielk, M., Gentner, T.Q., 2020。发现、可视化并量化不同动物发声 repertoire 中的潜在结构。《PLOS计算生物学》16，e1008228。https://doi.org/10.1371/journal.pcbi.1008228 Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., Aroyo, L.M., 2021。“人人都想做模型工作，而不是数据工作”：高风险人工智能中的数据级联，载于：2021年CHI人因计算系统会议论文集。发表于CHI '21：人因计算系统会议，美国计算机协会，日本横滨，第1–15页。https://doi.org/10.1145/3411764.3445518 Satopaa, V., Albrecht, J., Irwin, D., Raghavan, B., 2011。大海捞“针”：检测系统行为中的拐点，载于：2011年第31届分布式计算系统国际会议研讨会。发表于2011年第31届分布式计算系统国际会议研讨会，第166–171页。https://doi.org/10.1109/ICDCSW.2011.20 Schneider, S., Hammerschmidt, K., Dierkes, P.W., 2022。介绍CASE软件（声音事件聚类与分析）：通过动物发声比较不同聚类方法与音频转换技术 26。Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X., 2017。重新审视DBSCAN：为何以及如何（仍然）应该使用DBSCAN。《美国计算机协会数据库系统汇刊》42，19:1-19:21。https://doi.org/10.1145/3068335 Schwartz, R., Dodge, J., Smith, N.A., Etzioni, O., 2020 63，54 63。绿色人工智能。《美国计算机协会通讯》–。Sheng, V.S., Provost, F., Ipeirotis, P.G., 2008。再获取一个标签？利用多个带噪标注者提升数据质量与数据挖掘，载于：第14届美国计算机协会知识发现与数据挖掘特别兴趣小组国际会议论文集。第614–622页。Shrestha, R., Glackin, C., Wall, J., Cannings, N., 2021。使用 Faster R-CNN进行鸟类音频说话人分离，载于：Farkaš, I., Masulli, P., Otte, S., Wermter, S.（编），人工神经网络与机器学习– ICANN 2021，《计算机科学讲义》。施普林格国际出版，查姆，第415–426页。https://doi.org/10.1007/978-3-030-86362-3 34¯ Sprengel, E., Jaggi, M., Kilcher, Y., Hofmann, T., 2016年。使用深度学习技术进行基于音频的鸟类物种识别。Steinfath, E., Palacios-Muñoz, A., Rottschäfer, J.R., Yuezak, D., Clemens, J., 2021年。使用深度神经网络快速、准确地标注声学信号。《eLife》10，

e68837. Stoumpou, V., Vargas, C.D.M., Schade, P.F., Boyd, J.L., Giannakopoulos, T., Jarvis, E.D., 2022. 小鼠发声交流分析 (AMVOC)：一种深度无监督方法，用于快速检测、分析和分类超声发声。《生物声学》1–31. https://doi.org/10.1080/09524622.2022.2099973 Stowell, D., 2022. 深度学习计算生物声学：综述与路线图。PeerJ 10, e13152. https://doi.org/10.7717/peerj.13152 Strubell, E., Ganesh, A., McCallum, A., 2019. 自然语言处理深度学习的能源与政策考量。arXiv 预印本, ArXiv190602243. Thomas, M., Jensen, F.H., Averly, B., Demartsev, V., Manser, M.B., Sainburg, T., Roch, M.A., Strandburg-Peshkin, A., 2021. 生成基于声谱图的动物发声潜在空间表示的实用指南：无监督方法。https://doi.org/10.1101/2021.12.16.472881 Toledo, L.F., Tipp, C., Márquez, R., 2015. 视听档案的价值。《科学》347, 484–484. https://doi.org/10.1126/science.347.6221.484-b Tolkova, I., Chu, B., Hedman, M., Kahl, S., Klinck, H., 2021. 使用深度音频嵌入解析鸟鸣。https://doi.org/10.48550/arXiv.2108.09203 Towsey, M., 2013. 去除源自环境自然录音的波形和声谱图中的噪声 | 昆士兰科技大学电子文库 [网页文档]。网址 https://eprints.qut.edu.au/61399/（访问于 2021年3月30日）。Ulloa, J.S., Aubin, T., Llusia, D., Bouveyron, C., Sueur, J., 2018. 使用无监督多分辨率分析估算热带环境中的动物声学多样性。《生态指标》90, 346–355. https://doi.org/10.1016/j.ecolind.2018.03.026 Ulloa, J.S., Haupert, S., Latorre, J.F., Aubin, T., Sueur, J., 2021. scikit-maad：用于 Python 定量声景分析的开源-模块化工具箱。Methods Ecol. Evol. 2041–210X.13711. https://doi.org/10.1111/2041-210X.13711 Venkatesh, S., Moffat, D., Miranda, E.R., 2022. 你只需听一次：一种用于音频分割和声音事件检测的类 YOLO 算法。《应用科学》12, 3293. https://doi.org/10.3390/app12073293 Wu, S.-H., Chang, H.-W., Lin, R.-S., Tuanmu, M.-N., 2022. SILIC：一种跨数据库框架，基于目标检测和微型训练数据集，自动从声景录音中提取可靠的生物多样性信息。《生态信息学》68, 101534. https://doi.org/10.1016/j.ecoinf.2021.101534 Xeno-canto, 2021. 分享来自世界各地的鸟类声音。https://xeno- canto.org/（访问日期：2021-06-05）。Xie, J., Hu, K., Zhu, M., Guo, Y., 2020. 1991年至2018年生物声学和生态声学全球研究趋势的数据驱动分析。《生态信息学》57, 101068. https://doi.org/10.1016/j.ecoinf.2020.101068 Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. 深度神经网络中的特征可迁移性如何？，载于：神经信息处理系统进展。Curran Associates, Inc. Zhang, J., Hsieh, C.-Y., Yu, Y., Zhang, C., Ratner, A., 2022. 程序化弱监督综述。arXiv 预印本，ArXiv220205433. Zhong, M., Taylor, R., Bates, N., Christey, D., Basnet, H., Flippin, J., Palkovitz, S., Dodhia, R., Ferres, J.L., 2021. 通过深度卷积神经网络检测区域性稀有鸟类。《生态信息学》64, 101333. Zottesso, R.H.D., Costa, Y.M.G., Bertolini, D., Oliveira, L.E.S., 2018. 使用声谱图和差异度方法识别鸟类物种。《生态信息学》48, 187–197. https://doi.org/10.1016/j.ecoinf.2018.08.007 Zou, Z., Shi, Z., Guo, Y., Ye, J., 2019. 20年来的目标检测：综述。https://doi.org/10.48550/arXiv.1905.05055 Zsebők, S., Nagy-Egri, M.F., Barnaföldi, G.G., Laczi, M., Nagy, G., Vaskuti, É., Garamszegi, L.Z., 2019. 使用开源深度学习目标检测方法自动分割鸟鸣和音节———白领姬鹟的案例研究*(Ficedula albicollis)*。《Ornis Hung.》27, 59–66. https://doi.org/10.2478/orhu-2019-0015

经费

本研究是在索邦大学索邦人工智能中心（SCAI）的支持下完成的（IDEX SUPER 11-IDEX-0004）。

	训练数据集					验证数据集				
		DBSCAN		BirdNET			DBSCAN		BirdNET	
	IN	FN	R	FN	R	IN	FN	R	FN	R
长尾山雀 (银喉长尾山雀)	55	26	52	38	91	33	0	32	18	96
欧亚云雀 (欧亚云雀)	40	4	71	13	87	48	12	92	12	88
树鹨 (树鹨)	24	7	90	5	85	23	14	69	15	94
短趾旋木雀 (短趾旋木雀)	23	2	98	13	100	35	23	92	26	98
欧洲金翅雀 (Chloris chloris)	55	35	34	31	77	13	4	31	6	75
欧鸽 (Columba palumbus)	23	8	28	12	90	69	15	100	29	100
大杜鹃 (Cuculus canorus)	18	0	12	10	99	25	0	36	9	83
黄鹀 (Emberiza citrinella)	37	5	47	21	88	47	17	69	21	90
欧亚鸲 (Erithacus rubecula)	42	19	67	28	60	36	12	87	13	64
苍头燕雀 (Fringilla coelebs)	27	2	94	10	88	21	13	80	9	98
夜莺 (Luscinia megarhynchos)	39	20	90	2	69	40	32	89	22	57
大山雀 (Parus major)	54	42	68	25	71	71	38	50	46	100
普通红尾鸲 （普通红尾鸲）	71	35	56	57	96	40	18	85	14	90
欧洲绿啄木鸟 （Picus viridis）	36	4	90	0	76	59	8	71	13	76
篱雀 （Prunella modularis）	23	14	83	16	96	26	6	56	14	92
普通鸸 （普通鸸）	39	32	31	31	70	35	24	43	18	82
乌鸫 （Strix aluco）	15	0	74	6	97	36	0	56	15	91
黑顶林莺 （黑顶林莺）	42	30	100	14	69	42	28	100	23	79
欧亚鹪鹩 (鹪鹩)	33	11	77	12	85	27	14	86	10	75
乌鸫 (乌鸫)	30	10	57	14	82	31	4	97	4	75
斑鸫 (Turdus philomelos)	78	76	88	46	54	83	77	95	52	48
槲鸫 (Turdus viscivorus)	27	15	17	16	89	23	10	61	9	80

表A.1. DBSCAN 聚类和 BirdNET 方法在训练集和验证集上的得分。IN：初始噪声，FN：最终噪声，R：召回率。

	测试数据集				
		DBSCAN		BirdNET	
	IN	FN	R	FN	R
芦苇莺 （芦苇莺）	49	27	41	26	52
东方大苇莺 (大苇莺)	67	44	59	37	71
草地鹨 (草地鹨)	27	0	37	3	70
小鸮 (小鸮)	58	75	14	38	55
欧洲金翅雀 (欧洲金翅雀)	42	8	70	19	94
棕扇尾莺 (Cettia cetti)	60	45	84	45	84
蓝山雀 (Cyanistes caeruleus)	80	79	42	68	88
芦鸦 (芦鸦)	66	64	67	62	83
家燕 (家燕)	36	20	98	13	94
北石鹎 (普通石鹎)	23	7	90	9	59
煤山雀 （煤山雀）	37	10	75	23	70
黑红尾鸲 （Phoenicurus ochruros）	10	3	92	4	85
叽喳柳莺 （叽喳柳莺）	47	37	71	36	81
柳莺 (柳莺)	29	12	80	13	94
红腹灰雀 (Pyrrhula pyrrhula)	82	70	50	20	44
戴菊 (戴菊)	44	53	25	34	97
欧洲金翅雀 (欧洲金翅雀)	49	41	81	29	84
灰斑鸠 (灰斑鸠)	27	5	90	14	95
紫翅椋鸟 (Sturnus vulgaris)	62	45	80	0	14
园林莺 (园林莺)	52	30	95	23	92
白喉莺 (白喉莺)	15	9	84	11	71

田鹀 (田鹀)	57	48	79	28	50

表A.2 DBSCAN 聚类和 BirdNET 方法在测试集上的得分。IN：初始噪声，FN：最终噪声，R：召回率