

信息网络专题（应用层）

一、文献信息

- 1、 论文题目: Detecting Endangered Baleen Whales within Acoustic Recordings using R-CNNs
- 2、 作者: Mark Thomas, Bruce Martin, Katie Kowarski, Briand Gaudet and Stan Matwin
- 3、 发表途径: AI for Social Good workshop at NeurIPS (2019), Vancouver, Canada.
- 4、 发表时间: 2019

二、问题意义

1、研究背景

应用背景: 过去几年中, 在北美洲的沿海地区死了数百只鲸鱼。在大多情况下, 这些鲸鱼的主要死亡原因与人类活动直接相关, 其中就包括船只碰撞和渔具纠缠。为了减少发生碰撞和纠缠的风险, 当保护区中存在濒临灭绝的鲸鱼时, 政府会实施速度限制和临时捕鱼禁令。但是, 确定是否存在鲸鱼可能是一项艰巨的任务, 并且通常是在碰撞等事件发生后才施加上述限制。

学术背景: 在过去几年间, 许多研究都集中在使用模板或手工设计的特征来设计特定的检测算法, 但是这些系统通常未考虑的新噪声源或其他海洋哺乳动物物种; 最近, 一些研究人员已经使用卷积神经网络 (CNN) 来开发更通用的系统, 该系统能够确定完整记录的样本中是否存在特定物种, 但它们仅能够确定时间上的存在/不存在。此外, 即使在同一样品中存在来自不同物种的多次发声, 它们也仅限于每个样品检测一种物种。

2、主要研究问题

文章主要是研究了使用基于区域的卷积神经网络 (R-CNN) 开发端到端检测系统的初步工作。

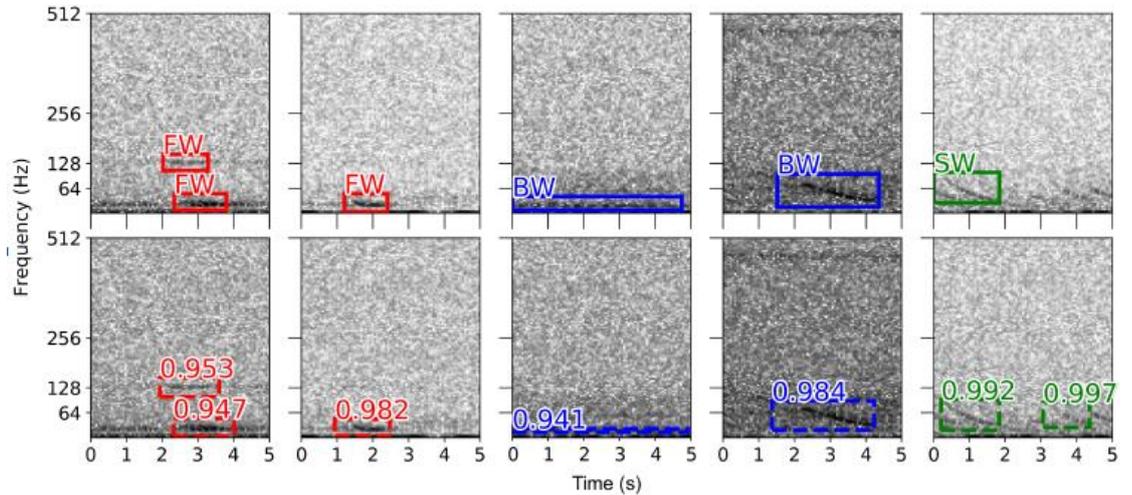
3、研究意义

这个系统使用声学记录的声谱图表示法, 围绕三种濒危须鲸 (蓝鲸、鳍鲸、塞鲸) 的发声标记进行训练。R-CNN可以使用大量的声音记录进行训练, 特别是使用频谱图在R-CNN的频域中进行训练, 来检测三种濒临灭绝的须鲸的声音, 因此人们可以利用R-CNN减轻船只碰撞和渔具纠缠的风险, 从而保护须鲸。

三、思路方法

首先是数据收集: 文章团队在2015年夏季和秋季以及2016年夏季, 在加拿大大西洋沿岸捕获了大量声学记录。这些记录位置位于生物感兴趣的地区, 即斯科蒂架上。由于大多数濒临灭绝的鲸鱼发声都远低于1000Hz, 因此团队将训练数据限制为较低的采样率。团队对一小部分声学记录进行了分析, 随后对其进行注释, 用来产生三种鲸鱼的声音特征范围。

下图中显示的示例每个都描绘了一个声谱图的可视化表示



将信号x的频谱图表示形式表示为短时傅立叶变换的绝对值的平方，如下式：

$$X(n, \omega) = \sum_{m=-\infty}^{\infty} x[m]w[m-n]e^{-j\omega m}$$

其中time (n)是离散的，frequency (ω)是连续的，ω起到一个窗口的功能

文章使用频谱图的原因之一是，它们数据记录注释过程中使用的主要资源，因为它们可以快速分析人类听觉范围内的信号。

之后进行数据训练。文章使用分层采样例程和70/15/15的随机分配比率，生成了用于训练，验证和测试R-CNN的不同数据集。单个训练实例可以解释为具有与频谱图相对应的一个通道的张量。文章的频谱图实际上不是通过上式生成的，而是使用的快速傅里叶算法。由于研究的声音（须鲸的声音）有低频特性，因此在训练过程中使用的频谱图的最大频率为512Hz。文章使用他人提出的Mask R-CNN架构。作者对R-CNN进行了100个时期的训练，并根据验证集的丢失来评估其提前停止。训练使用了四个NVIDIA P100 Pascal GPU，每个GPU具有16GB的内存和4个批次。随机梯度下降算法的初始学习率设置为0.003，并且当验证集的损失停止减少时衰减10倍。

最后将R-CNN与其他检测器进行对比，如下表所示：

表为在各种IoU阈值上评估的平均精度（AP）和平均召回率（AR）指标的中值。

Species	Label	AP@.5		mAP@[.5:.95]		AR@.5		mAR@[.5:.95]	
		R-CNN	JASCO	R-CNN	JASCO	R-CNN	JASCO	R-CNN	JASCO
Overall	-	82.1	-	41.8	-	91.9	-	54.8	-
Blue whale	BW	85.7	-	52.8	-	96.2	-	70.9	-
Fin whale	FW	75.3	65.0	30.8	27.4	89.9	62.6	40.0	35.0
Sei whale	SW	85.4	75.7	41.9	35.0	89.7	34.4	49.4	18.4

可以看出，当考虑到地面真值区域和预测之间的联合时的低交叉点（IoU）时，R-CNN表现良好，对于IoU值大于0.7，表现下降，由表的mAP/mAR@[.5:.95]列反映出来。R-CNN的性能优于JASCO的有鳍鲸和塞鲸的当前检测算法。无法完全比较蓝鲸发声的检测算法，因为JASCO检测器需要30秒的样本才能检测蓝鲸，而R-CNN仅使用5秒的样本进行训练。

四、实验结论

应用结论：通过被动声监测（PAM）和本文概述的R-CNN实施和有效的政策决策，持续监控鲸类动物是可行的

学术结论：R-CNN可以使用大量的声音记录进行训练，特别是使用频谱图在R-CNN的频

域中进行训练，来检测三种濒临灭绝的须鲸的声音，且有相对于现有工作的优势，如上表对比。文章的工作是初步的，也是正在进行的项目的一部分，该项目的重点是利用记录的声音检测濒危和非濒危的各种海洋哺乳动物的声音。

五、启发思考

文章比较短，作者也只是初步的研究了这个系统，后续肯定还会有更好的发展进展。本文是人工智能造福人类社会的一个例子。人工智能也是我们专业的一个方面，我们可以用数据对模型进行训练从而达到预测的目的，不只是这篇文章，是一种趋势。作者写的不是特别的详细，主要是想说一下自己成果有什么作用，而没有在文章中介绍出来如何去应用自己的成果，可能涉及到一些专利或者利益关系。机器学习的过程与人类对历史经验归纳较为相似，人类在成长、生活过程中积累了很多的历史与经验。人类定期地对这些经验进行归纳，获得了生活的规律。当人类遇到未知的问题或者需要对未来进行推测的时候，人类使用这些规律，对未知问题与未来进行推测，从而指导自己的生活和工作。俗话说“历史往往不一样，但历史总是惊人的相似”。通过学习历史，我们从历史中归纳出人生与国家的规律，同样机器学习的思想并不复杂，仅仅是对人类在生活中学习成长的一个模拟，机器学习不是基于编程形成的结果，它的处理过程不是因果的逻辑，而是通过归纳思想得出的相关性结论。文章让我更熟悉了人工智能与机器学习，同时激发了对这方面的兴趣，可能以后向这方面学习时，会因为兴趣而感到轻松。