

一、 文献信息

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

二、 问题意义

1. 研究问题/背景

在过去的几年里，数百头死去的鲸鱼被冲上了北美海岸。在许多情况下，这些物种的主要死亡原因与人类活动有直接联系，包括船只碰撞和渔具缠绕。加拿大政府已开始在保护区中存在濒危鲸鱼时，实施速度限制和临时捕鱼禁令。但是，如何确定是否存在要保护的鲸鱼还是一个问题。所以，研究和开发能够在声音记录中检测濒危鲸鱼发声的自动化系统是一项重要的任务。

2. 研究意义

在这篇文章中，作者介绍了一个端到端检测系统。该检测系统使用基于区域的卷积神经网络 (R-CNN) 进行训练，用于检测三种濒危须鲸（蓝鲸，长须鲸和塞鲸）的发声记录和标记的发声边界框。

这样，R-CNN 可以在噪声和其他非生物来源的环境下，根据时间和频率检测发声。通过连续不断地实时监测濒危物种的存在，决策者可以更有效地实施对船只和捕鱼的限制。

三、 思路方法

1. 研究思路

多年来，研究和开发用于检测声音记录中的海洋哺乳动物的自动化系统一直是一个令人感兴趣的话题。最近，几位研究人员使用 CNNs 开发了具有普遍性的系统，能够确定在完整记录的样本中是否存在特定的物种。但是，它们只能在时间上确定存在或不存在。当同一样本中存在来自不同物种的多种发声，它们也仅限于检测每个样本中的一个物种。

针对海洋哺乳动物发声检测系统中的多物种检测这一问题，作者想到，可以利用深度学

习加以解决。

2. 研究方法

作者收集发声记录和声谱图数据，并投入模型中进行训练。作者采用了 Mask R-CNN 架构，并在 Python 中实现。具体来说，作者使用 ResNet-50（残差神经网络）进行特征提取，并将特征金字塔网络（FPN）作为主干。然后，FPN 的 256 个输出特性被传递给区域建议网络（RPN），每个训练实例产生 1000 个感兴趣区域（RoI）建议。之后，1000 个 RoIs 通过 RoiAlign 过程和由全连接层组成的网络进行分类和边界框回归。

1. 发声记录。

作者于 2015 年夏季和 2016 年秋季，在加拿大大西洋海岸捕获了大量声音录音。录音是以 8 kHz 和 250 kHz 采样的。然而，由于大多数濒危须鲸的发声都远远低于 1000 Hz，作者将训练数据限制在较低的采样率。

对收集到的发声记录，一些专家利用已有技术对其进行分析，在我们关注的三种濒危须鲸的发声周围标记边界框。如图 1 中的第一行。

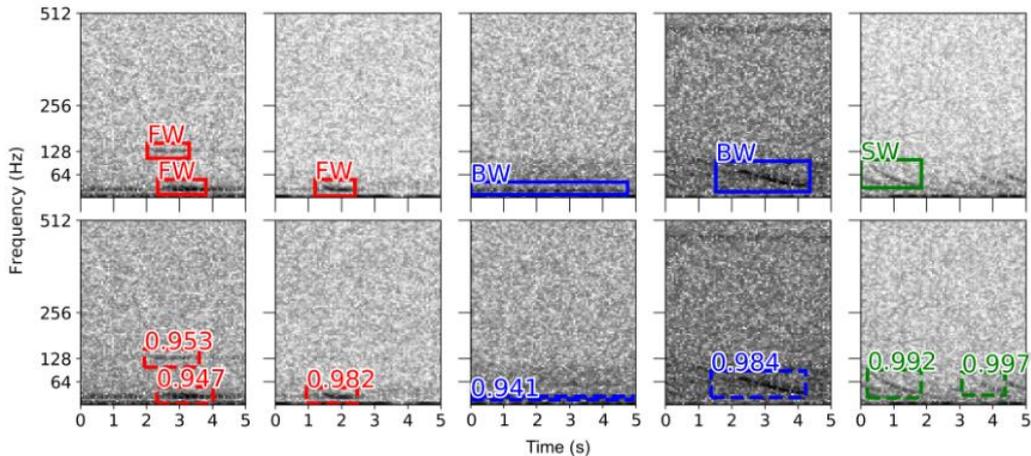


图 1 几种声谱图

作为比较，地位行是利用 R-CNN 检测得到的。可以看出，主要区别在于第五列。此时，似乎有几个 SW 发声连续出现，然而现有技术只标记了一个发声，而 R-CNN 检测到了两种发声。

2. 声学信号的频谱图。

声谱图可以由如下的短时傅里叶变换的绝对值的平方得到。

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

其中，时间 n 是离散的，频率 ω 是连续的， w 是窗口函数。

在许多相关文献中，不论是传统上使用的卷积模板和人工工程特征分类，还是最近使用的深度学习，用于检测海洋哺乳动物发声的算法都是在频域中工作的，即针对频谱图进行工作。原因之一是：使用频谱图可以对人类听觉范围内的信号进行快速分析。

3.训练数据。

作者使用分层抽样程序和 70/15/15 的随机分割率分别生成用于训练、验证和测试 R-CNN 的不同数据集。在实际应用中，谱图并不直接使用上述方程生成，而是使用一种快速傅里叶变换（FFT）算法。设置窗长为 2048 个样本、窗重叠为 512 个样本的 Hann 窗函数，作者利用 FFT 产生了 5 秒长的信号谱图，并截取低频部分作为训练谱图。

四、 实验结论

作者使用不同的随机数生成器种子值来产生不同的训练、验证和测试数据集，得到了十次训练运行的中值。

当考虑到 IoU 较低时，R-CNN 表现良好。值得注意的是，R-CNN 的性能优于 JASCO Applied Sciences 生产的长须鲸和塞鲸的当前检测算法。这些算法是通过使用更传统的方法来实现的，方法是：首先将候选检测结果提取为频谱图中的轮廓，然后把从轮廓中提取的特定特征与地面真实模板进行比较。

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

表 1 AP 和 AR 指标的中位数

自 1980 年代后期禁止商业捕鲸以来，许多鲸类的种群一直在增加，其中包括一些蓝鲸和长须鲸。但是，海上人类活动的增加对这些物种的生计构成了重大威胁。为了减少这种威胁，政府决策者必须继续施加速度限制，并暂时中止易受影响地区的捕捞活动。作者相信，通过 PAM 和本文概述的 R-CNN 的实施，持续监控鲸类动物和有效的政策决策是可行的。

五、 启发思考

近几年，神经网络与深度学习特别火，但是我没想到还可以用到保护濒危动物上去。这篇文章读完，我感觉我们现在的技术真的很强大，感觉现实中很多之前难以建模的东西都可以用深度学习来完成。

另外，我感觉技术的发展不是一蹴而就的，都是不断经历修修补补才趋于成熟的。

最初，大多数研究集中在使用模板来设计特定的检测算法，这些系统无法推广到算法设计期间未考虑到的情况中去。这种系统就比较笨重和僵硬。之后，人们使用 CNN 来开发更通用的系统，该系统能够确定特定物种是否存在，但在同一样本中存在来自不同物种的多次发声时，它们也仅限于每个样本检测一种物种。再到作者将深度学习融合进去，解决之前的问题。

我很明显地感受到技术是一代代发展的，现在的高精尖技术也是一代代人的奉献的总和！