

# 信息网络专题研究课文献阅读报告（应用层）

## 一.文献信息

1. 作者: M. Thomas , B. Martin , K. Kowarski , B. Gaudet , and S. Matwin
2. 题目: Detecting Endangered Baleen Whales within Acoustic Recordings using R-CNNs
3. 发表途径: AI for Social Good workshop at NeurIPS2019
4. 发表时间: 2019 年

## 二.问题意义

### 1. 研究背景

研究表明,人类活动对 40%以上的海洋生态系统是有害的。各种海洋生物的种群减少与人类活动直接相关,其中,21 世纪须鲸数量下降的主要原因就是船只碰撞和渔具缠结。因为人们难以确认某一片海域内是否存在濒危鲸鱼,所以通常在事故发生后才会对濒危鲸鱼出现的保护区内采取限速和禁止捕鱼的措施,但是这并不是一项安全措施,我们只有及时地发现濒危鲸鱼才能更好地保护它们。

目前,用于检测濒危鲸鱼是否存在的一种较为可靠的方法是被动声学监测(PAM)。PAM 通过对记录的声波进行分析,PAM 是非侵入性的,不易受恶劣天气条件的影响,因此比 GPS 标记和视觉测量更受欢迎。通常,PAM 通过使用装有水听器的专用硬件和几兆字节的存储器来收集和记录所需要的声音。

### 2. 研究问题及意义

本篇论文采用了一种基于区域的卷积神经网络(R-CNN)来检测濒危鲸鱼的声音。R-CNN 是一种利用深度学习开发的海洋哺乳动物发声检测系统,能够在时间和频率上处理多物种检测。通过使用 R-CNN 对三种濒临灭绝的须鲸(蓝鲸、长须鲸和西须鲸)的发声进行了声学记录和标记训练,发现 R-CNN 在频域中使用频谱图来检测三种濒危须鲸的效果很好。

R-CNN 可以实时地用于船只、海洋滑翔机和具有电信能力的系泊设备,以确定是否存在濒危鲸鱼物种。通过使用 R-CNN 实时地持续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,进而降低碰撞和纠缠的风险。

## 三.思路方法

### 1. 研究思路

#### (1) 数据收集和处理

①录音:2015 年和 2016 年的夏秋两季,在大西洋沿岸一个叫苏格兰大陆架的区域捕获了大

量濒危鲸鱼的声音。这些录音的采样率分别为 8kHz 和 250kHz，因为大多数濒危的须鲸的发声频率远低于 1000Hz，所以将训练数据限制在较低的采样率。与人工分析相比，R-CNN 标注了连续出现的声音，提高了测试实例的报告指标。

② 声信号的频谱：大多数情况下，检测海洋哺乳动物发声的算法都在频域中实现，主要原因是人们可以通过频谱图对听觉范围以外的信号进行快速分析。信号  $x$  的频谱图可以用短时傅里叶变换绝对值的平方来获得，表示如下：

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

其中时间  $n$  是离散的，频率  $\omega$  是连续的， $\omega$  是一个窗口函数。

③ 训练数据和实验装置：

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。单个训练实例可以理解为一个张量，每个通道对应着一个频谱。在实际应用中，使用快速傅立叶变换（FFT）计算离散信号的频谱。FFT 利用窗长为 2048 个样本、窗重叠为 512 个样本的 Hann 窗函数来产生 5 秒长的频谱图。因为主要研究发声的低频特性，所以训练期间将频谱图截断，最大频率为 512Hz。

实验采用 Python 中的开源的深度学习库 PyTorch 来实现 Mask R-CNN 架构。具体来说，使用 ResNet-50 进行特征提取，并以特征金字塔网络（FPN）为骨干。然后，把 FPN 的 256 个输出特征交给标准区域候选网络（RPN），每个训练实例产生 1000 个 RoI 提案。然后，1000 个 RoI 通过 RoIAlign 过程和由全连接层组成的头网络进行分类和边框回归。

训练数据使用 4 个 NVIDIA P100 Pascal GPU，每个 GPU 有 16GB 的内存，批量大小为 4。随机梯度下降算法的初始学习率为 0.003，当验证集的损失停止下降时，初始学习率衰减了 10 倍。

## （2）初步结果

分析实验数据可知：**R-CNN** 在考虑标注数据的边界框和预测之间的低交并比（IoU）时表现良好。当 IoU 的值大于 0.7 时，**R-CNN** 的性能下降。结果表明：**R-CNN** 比目前 JASCO Applied Sciences 生产中使用的用于检测长须鲸和西须鲸的算法都要好。对蓝鲸发声的检测算法不可能全面的比较，因为 JASCO 探测器需要 30 秒的样本来检测蓝鲸，而 **R-CNN** 仅需要 5 秒的样本来进行训练。

2. 研究方法：对比分析，理论结合实际，科学验证。

## 四.实验结论

保护濒危鲸鱼对于维持海洋生态系统稳定具有至关重要的作用，人类已经逐步意识到这一点并采取了許多有效的措施。PAM 和 R-CNN 的实施为濒危鲸鱼的长期持续监测提供了可能，人们可以及时地发现濒危鲸鱼，进而采取更加有效的措施来保护它们，保护生态系统。R-CNN 检测动物发声的模型在进一步完善中，未来 R-CNN 算法一定还会应用到其他濒危物种的监测中去。

## 五.启发思考

1、本文介绍了用 R-CNN 算法实现发声检测的过程，与传统的机器学习算法相比较，主要区别是利用了深度学习的卷积神经网络结构。在大量训练数据的支持下，深度学习可以展现出很好的效果，可以让机器得到更好的学习，极大地减少了人工参与的过程，可以长期持续的做同一件事，并且可以做的很好。这就是人工智能的优势和意义，希望未来通过我们的努力人工智能可以更好的造福人类和社会。

2、大作业利用 python 的开源数据库 PyTorch 实现了 CNN 卷积神经网络结构的搭建、训练和测试，来进一步研究 CNN 的运行过程和特点。