读书笔记

一、文献信息

论文题目：Detecting Endangered Baleen Whales within Acoustic Recordings usingR-CNNs

作者：Mark Thomas[1,2,*] , Bruce Martin[2] , Katie Kowarski[2] , Briand Gaudet[2] , and Stan Matwin[1,3]

1 Faculty of Computer Science, Dalhousie University, Halifax, Canada

2 JASCO Applied Sciences, Dartmouth, Canada

3 Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

发表时间：2019 年

二、问题意义

相关保护研究表明，人类活动对海洋生态系统的有害占比超过 40%，一些情况下，人类活动与海洋生物数量减少直接相关。例如，在过去的几年里，数以百计的鲸鱼尸体被冲上了北美海岸，这些物种的死亡的很大因素是包括船只碰撞和渔具缠结。通常情况下，由于不能及时检测到濒危物种的存在，当不幸的伤亡事故已经发生时，相关措施（限速和暂时性禁渔指令）才作为补救措施被实施，大大损害濒危物种的生命安全。

目前一个比较可靠的检测方法是通过对声音记录的分析来确定濒危鲸鱼的存在与否（动声学检测 Passive Acoustic Monitoring,PAM）, PAM 具备非入侵性（相比于 GPS），对恶劣天气条件的敏感性也较低（相比于目测）。通常，PAM 的声学记录来源为安装了水听器和几个 tb 存储空间的专门的硬件，可产生大量数据并保持静态几个月，或者，PAM 可以在船尾安装水听器进行实时监测。同时，基于声音记录的自动化检测系统的研究与发展也是很多年来的热点话题。很多研究都集中在设计专门的检测算法上（针对特定物种声音特性设计特定算法），这样对于新的噪声源或算法中没有考虑的物种将不能适用。最近，更通用化的系统 CNNs 被利用，它能够在完整记录的样本种判定特定物种的存在。**问题在于，上述系统都需要时间来确定物种是否存在，并且，即使同一样本种存在来自不同物种的多种声音，他们也只能检测到一个物种。**

基于上述问题，本文介绍了使用基于区域的卷积神经网络(Region-based Convolutional Neural Network,R-CNN)开发端到端检测系统的初步工作，这是一个利用深度学习开发相关系统的研究，基于上述提到的前一类声音记录（系泊设备收集的大量录音），针对三种濒危长须鲸（BW、FW、SW），对该网络发声的声谱图频域表示和边界框标定进行训练（深度学习）。**R-CNN 可以在环境噪声和其他非生物声源的背景下，根据时间和频率测定发声，并处理多物种检测**。此外，R-CNN 可以在船舶、海洋滑翔机或具有远程通信能力上实时使用，以确定濒危物种的存在与否，进而通过持续监测，决策者可以更有效地对船只和渔业进行限制，从而避免悲剧的发生，达到真正的保护意义。

三、思路方法

本文的目的就是利用现有注释的数据，提出 R-CNN 深度学习架构，对 R-CNN 进行训练，使其深度学习，并与现有传统的基于设计模型的声音检测算法进行比较。

1. 数据收集与处理

首先，作者采用 2015 年和 2016 年的夏天和秋天，加拿大大西洋海岸外（Scotion Shelf）捕获的大量声音记录（第一种静态收集记录）。录音采样在 8kHz 和 250kHz，限制训练数据采样率较低（濒危长须鲸的发声频率低于 1000Hz）。这些样本由海洋生物学家针对特定物种做了部分的边界框标注。如图 3.1.1。正是由于频率是生物学家注释过程主要使用的资源，各相关研究，无论是传统卷积模板式还是深度学习，都以声谱图为主要工具。

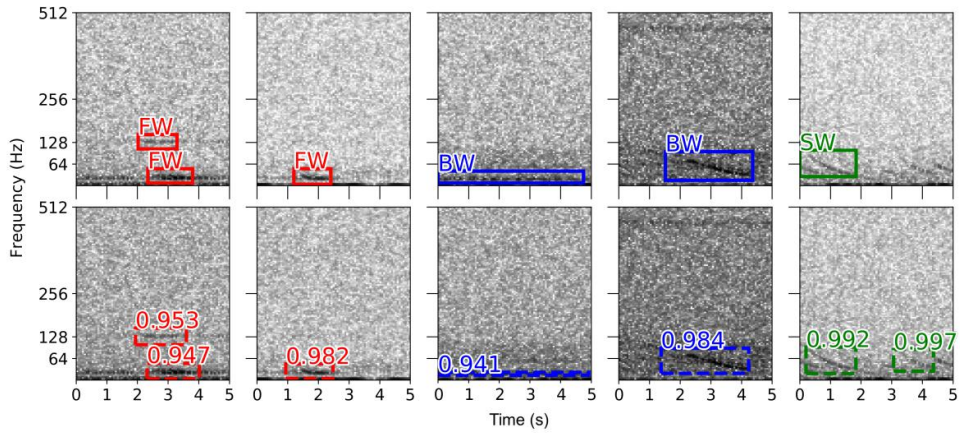

图 3.1.1 三个感兴趣的物种所产生的几个发声例子的示例注释(顶部一行)和由 R-CNN 做出的相应的预测（底部一行)。

如图 3.1.1 中的每个示例描述了称为声谱图的声学信号的可视化表示。粗略来讲，信号 x 的谱图表示可以用短时傅里叶变换绝对值的平方表示：

$$X(n,\omega)=\sum_{m=-\infty}^{\infty}x[m]w[m-n]e^{-j\omega m} \quad (1)$$

时间 n 离散，频率 ω 连续，w 是窗口函数。

2. 训练数据和实验设定

作者对数据和实验进行设定，用于训练、验证和测试 R-CNN 不同数据集分别使用抽样程序并以 70/15/15 的随机分割比生成。每一个训练实例可以被解释为一个带有一个与光谱图对应的频道的张量。实践中，谱图由快速傅里叶变换（FFT）生成。

FFT 用来产生 5 秒长的信号的声谱图，需对信号进行加窗截取，得到信号随时间变化的频谱图，本文使用 Hann 窗口（窗口长度为 2048，窗口重叠为 512）。

作者采用 Mask R-CNN 架构，在 Python 中使用开源的深度学习框架 PyTorch 实现：

使用 ResNet-50 与特征金字塔网络(Feature PyTorch Network,FPN)[6]为骨架耦合进行特征提取。然后将 FPN 的 256 个输出特性交给标准区域建议网络(Region Proposal Network,RPN)，每个训练实例生成 1000 个感兴趣区域(Region of Interest,RoI)建议。然后，

1000 个 roi 通过 RoIAlign 过程和由完全连接的层组成的 head 网络进行分类和边界盒回归。R-CNN 接受了 100 个 epoch 的训练，在丢失验证集时早停止评估。4 个 NVIDIA P100 Pascal gpu，每个 gpu 有 16GB 内存，批处理大小为 4 个用于训练。将随机梯度下降算法的初始学习率设置为 0.003，当验证集的损失停止下降时，学习率衰减 10 倍。

四、实验结论

接着，作者对结果进行分析，该结果描述了使用不同随机数生成器种子值生成不同训练、验证和测试数据集的 10 次训练运行的中位数。

		AP@.5		mAP@[.5:.95]		AR@.5		mAR@[.5:.95]	
Species	**Label**	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 3.3.1 依据各种 loU 阈值评估的平均经度（Average Precision，AP）和平均召回率（Average Recall，AR）的中位数

如上表，R-CNN 在考虑地面真实边界框和预测之间的低相交联合（low Intersect over Union，loU）时表现良好（例如 AP/AR@.5），通过 mAP/mAR@[.5]反映出，loU 值大于 0.7 时性能下降。R-CNN 在检测 FW 和 SW 的表现都优于 JASCO 的算法（一些传统算法）。

作者指出通过 PAM 和 R-CNN 的实施，可以很好的对鲸类动物持续监测并有利于有效的政策决策的实施去保护物种生存。此外，该论文只是他们在进行的更大规模的依据声学记录检测物种的项目的一部分概述。他们还将考虑用转移学习来对项目描述的模型进行微调，来检测其余物种的声音。限制于部分注释的数据，作者还要研究学习能力更强的方法，弱监督或半监督或主动学习等。此外还要对模型进行压缩以使网络能够实时使用。

五、启发思考

作者将机器学习-深度学习应用于环境保护领域，设计学习算法，解决现有算法刚性的服务于检测特定物种的问题。并提出进一步工作规划。这篇文章只是一篇概述，并没有详细讲解 C-RNN 的具体算法，我将在网上对 C-RNN 算法进行具体的了解和学习。从此文章中，我了解了深度学习的一些性能表现，智能化的学习算法能省掉更多的人力，开发更多的资源，应用于更多的场景。实时监测也离不开通信系统的支撑，更快的通信速率，以及海上通信信号的传输，也是待解决的问题。