

OAC-114-015 (研究報告)

基於 Faster Region-Based Convolutional Neural Networks 識別與分類海洋聲
景中的鯨豚聲紋

(期末報告)

海洋委員會補助研究
中華民國 114 年 09 月

「本研究報告僅供海洋委員會施政參考，並不代表該會政策，該會保留採用與否之權利。」

OAC-114-015 (研究報告)

基於 Faster Region-Based Convolutional Neural Networks 識別與分類海洋聲
景中的鯨豚聲紋

(期末報告)

學校：銘傳大學

指導教授：郭庭榕

學生：凌毓淇

研究期程：中華民國 114 年 3 月至 114 年 10 月

研究經費：新臺幣十萬一千元

海洋委員會補助研究

中華民國 114 年 09 月

「本研究報告僅供海洋委員會施政參考，並不代表該會政策，該會保留採用與否之權利。」

「本研究報告絕無侵害他人智慧財產權之情事，如有違背願負民、刑事責任。」

目次

一、前言.....	2
第一節 研究主旨	2
第二節 研究背景及相關研究	3
第三節 預期目標與進度說明	6
第二章 研究方法及過程.....	8
第一節 資料集標記	8
第二節 Faster Region-Based Convolutional Neural Networks (Faster R-CNN)	10
第三節 Focal Loss 損失函數	11
第三章 研究成果.....	13
第一節 資料前處理	13
第二節 未使用 Focal Loss 分類結果.....	13
第三節 使用 Focal Loss 分類結果.....	16
第四章 結論.....	18
致謝.....	18
參考資料.....	18

圖次

圖 1 鯨豚聲音頻譜圖(引用出處[19])	4
圖 2 鯨豚口哨聲的種類(引用出處[19])	6
圖 3 工作甘特圖	7
圖 4 Raven Lite 操作環境	9
圖 5 標記聲紋種類	9
圖 6 標記哨聲聲紋中不同的種類	9
圖 7 鯨豚聲紋分類 (a)上升型 (b)平坦型 (c)非鯨豚	13
圖 8 未使用 Focal Loss 之模型預測結果	14
圖 9 訓練與驗證準確性	15
圖 10 訓練與驗證損失值	15
圖 11 未導入 Focal Loss 之模型混淆矩陣	16
圖 12 導入 Focal Loss 之模型混淆矩陣	17

表次

表 1 特徵分類筆數	8
表 2 Faster R-CNN 與 YOLO 的特性比較	11

摘要

隨著深度學習技術的快速發展，深度學習方法在不同領域的挑戰性問題上的應用引起了廣泛的興趣。卷積神經網路(Convolutional Neural Network, CNN)最近在各種音訊分析任務中取得的成功，本研究使用更快的區域卷積神經網路(Faster Region-based Convolutional Network, Faster R-CNN)在海洋聲景中偵測鯨魚聲音並對其進行分類進行了研究。基於深度神經網路模型，提取鯨魚聲紋的顯著特徵，得到更高層次的抽象表示，從而實現精準分類，明顯優於傳統的基於專家知識的人工特徵分類方法。具體來說，鯨魚的聲音代表不同的行為，因此有效的偵測並識別不同的聲音圖形，將有助於了解鯨魚的生態。

關鍵字：深度學習、更快的區域卷積神經網路、海洋聲景、鯨魚聲紋

一、前言

第一節 研究主旨

海洋聲景(Soundscape)是由生物音(鯨豚、魚類)、環境音(潮汐、風)與船舶這類的人為噪音等建構而成的水下聲音環境[1]，可以得到錄音地點的棲地特性與生態狀況。從生物聲音的強度、頻率與複雜度，可以推估發聲生物的群體數量、行為狀態與物種的多樣性[2]；在多個地點同時進行長時間錄音，也可以用來評估生物多樣性的時空變化。因此借由海洋聲景的分析，可以了解海洋生物的多樣性變化。

歐盟在 2015 年的巴黎氣候協定中，提出 1.5° C 的全球溫升承諾，台灣也在 2022 年發佈台灣 2050 淨零排放路徑及策略總說明，做為國家能源總體發展政策，並且在積極發展再生能源中，設立短期目標以 2025 年太陽光電達到 20GW 及離岸風電 5.6GW 裝置容量為主要發展目標[3]。然而，離岸風電為再生能源中極為重要的一環，在沿岸施工的過程中，所造成的人為噪音是否會影響海洋生物的生存環境，尤以鯨豚此類是對聲音極為敏感之生物，是我們急迫了解的環境永續議題。學者施仲平[4]指出離岸風場的建設過程中，可能因施工噪音和機械振動等大型海上工業活動，對該海域的鯨豚造成不可逆的傷害，包括聽覺系統損傷，或是對於族群間的通訊、覓食能力造成干擾等諸多傷害。此外，離岸風電的海底纜線和基礎設施可能改變海底地貌，對鯨豚的棲息地和覓食區造成破壞。故利用海洋聲景分析來研究海岸工程與鯨豚生存的永續環境之間的關聯，是迫切需要研究之議題。

傳統的海洋聲景分析方式是將水中麥克風(Hydrophone)蒐集到水下聲音信號，藉由傅立葉轉換[5]或梅爾頻譜圖[6]這兩種技術將聲音轉換為聲紋圖後由人工進行識別與分類它們的圖形特徵[7]。然而海洋環境的高度複雜性和變化性的影響，這使得聲音信號含有大量的噪音和干擾[8]。其次，不同種類的生物可能具有相似的聲音頻譜特徵或是它們的聲音被其他環境音所覆蓋。如槍蝦的大螯以每秒 20 公尺的速度瞬間闔上時，其兩片螯指間的水團被以每秒 30 公尺的速度將如子彈般的快速移動，在大螯前方形成了氣穴現象。氣泡崩解時所發出的聲響如喀噠聲，與鯨豚發出的喀噠聲極為類似，故造成判斷上的混淆。這些因素都會大幅度增加了分析的時間、人力與分類的困難度[9]。因此，需要高效率、高準確度的檢測演算法來識別與分類這些水下聲音。

本研究提出使用 Faster R-CNN (Region-Based Convolutional Neural Networks) 模型[10]來分析海洋聲景，並從中檢測並分類鯨豚的高頻聲。Faster R-CNN 在鯨豚聲音檢測中的應用主要基於聲紋圖(Spectrogram)的分析[11]。聲紋圖是一種將

聲音信號轉換為視覺圖像的方法，展示不同時間和頻率的能量分佈來顯示聲音的特徵。Faster R-CNN 將聲紋圖作為輸入圖像進行分析，檢測並標記出鯨豚聲音的特徵區域。Faster R-CNN 通過 Region Proposals Network 來生成候選區域，這些候選區域可能包含鯨豚的聲音特徵。然後，這些候選區域會被送入後續的卷積神經網路進行進一步的分類和位置回歸，以確定它們是否包含目標聲音。因此，Faster R-CNN 不能夠檢測到鯨豚的聲音，還可以識別出具體的物種和聲音類別。

使用 Faster R-CNN 進行聲音檢測相比傳統方法具有多方面的優勢。它能夠同時處理大量的數據，在相對較短的時間內完成大量聲音數據的分析，從而提高了整體監測的效率，這對於大規模的海洋聲學監測是非常重要的。Faster R-CNN 能夠在自動學習和提取聲音特徵，使得模型能夠更快適應不同環境下的聲音檢測任務。因此，本研究提出之 Faster R-CNN 未來可以擴展到其他海洋生物或船舶的聲音檢測中，為海洋生態監測中具有重要的應用價值。

第二節 研究背景及相關研究

鯨豚的聲音主要可以分為三大類，寬頻的喀嗒聲(Click)與脈衝聲(Burst Pulse)以及窄頻的哨叫聲(Whistle)，然而每一類又具有多種型態[12]。目前已知喀嗒聲與脈衝聲如同聲納(Sonar)一般，都具有回聲定位的功能。喀嗒聲主要代表齒鯨正在覓食與探查環境，齒鯨可透過接收到回聲的時間來推斷反射回聲的物體距離，再發出下一個喀嗒聲持續了解偵測物體的位置[13]。當捕捉獵物時，為了精準的追蹤獵物移動的路徑，齒鯨會以很高的頻度發出連續不斷的喀嗒聲來更新獵物位置。此時一連串的喀嗒聲由於其發聲間隔極小而形成了脈衝聲。目前已知所有的齒鯨都是利用類似的機制來搜索、捕捉獵物，因此當脈衝聲的出現往往可以做為齒鯨覓食行為的代表性聲音[14]。故分析一個區域的聲景，若過往在此區域可輕易收集到喀嗒聲，當鄰近有工程施工時鯨豚已無法在此覓食，則可能兩者具有生態關聯性。

鯨豚的哨叫聲是一種窄頻的聲音，此種類型的哨叫聲通常是代表鯨豚彼此之間正在進行溝通。哨叫聲在頻譜圖上具有不同的音頻變化特徵(曲調)，每一種鯨豚的哨叫聲曲譜皆可由數種哨叫聲類別所組成。如同歌曲一樣，鯨豚使用哨叫聲的模式具有相當大程度的變化[15]。雖然哨叫聲被認為是鯨豚個體之間維持聯繫、互相溝通的聲音，但目前對於鯨豚如何使用哨叫聲互相溝通，學術界尚未取得一至的定論，主要是因為哨叫聲的使用模式可能在不同的個體或是群體行為之間改變。例如寬吻海豚(*Tursiops truncatus*)被認為可發出特有的哨聲，如同人類的指紋一樣，以和同類互相保持連繫。當寬吻海豚在進行覓食行為時，母豚可能會暫時離開其子豚，母子會透過獨特的簽名哨聲來確認彼此的位置，以在短暫分離過後重新團聚[16]。然而此種獨特的哨聲，卻又有可能在子豚長大的過程中逐漸改變。中華白海豚(*Sousa chinensis*)群體在遷徙時所發出的哨叫聲曲譜較為單純，但是在

進行社交活動時，除了會利用遷徙所發出的哨叫聲之外，也會使用較為複雜的哨叫聲曲譜組成。根據目前的研究尚無法將特定哨叫聲與特定行為事件互相關聯。

然而不同種類的鯨豚哨叫聲在聲紋頻譜圖上的特徵不盡相同，例如大型的鬚鯨(Mysticeti)僅能發出低頻的聲音，在頻譜圖的特徵圖形較為單純；體型小的海豚聲音頻率則可達到如同超音波的高頻段。因此鯨豚聲音的音頻特徵，大略能做為鯨豚種類辨識的依據之一，但仍有例外。

學者林子皓提出利用鯨豚經常發出聲音的特性，可以透過偵測水下鯨豚聲音來監聽鯨豚是否出現。被動式聲學監測是一種利用水下麥克風來被動傾聽水中聲音的技術，在近年來被廣泛地採用作為監測鯨豚類活動的方式[17]。被動式聲學監測可以輔助目視觀察的不足，在能見度低的時間也能夠收集到鯨豚活動的資料。甚至透過某些特殊的聲音，還能夠辨識出鯨豚當時的活動。以中央氣象局在宜蘭外海所建置的台灣東部海域海纜觀測系統為例，其所配備的水下麥克風可以連續不間斷地收錄當地的水下聲音，並即時傳輸回岸上的資料處理中心。透過哨叫聲偵測器，可以自動化的把各種鯨豚的哨叫聲出現時間偵測出來，同時透過曲調追蹤演算法計算哨叫聲的音頻特徵，並透過電腦自我判斷的方法來分出各類哨叫聲。透過此一自動化的平台，目前發現在監測海域的鯨豚類活動以小型海豚為主，且其主要的活動時間以夜間為活動高峰。此外，在日落之後偵測到許多脈衝聲，代表著這些海豚可能在日落之後較常進行覓食行為[18]。圖 1 繪出鯨豚的聲音頻譜特徵，鯨豚的一段聲音的峰值頻率與峰值時間在圖形上可分為：

- Beginning frequency: 鯨豚聲音開始處的頻率
- Maximum frequency (F_{max}): 鯨豚聲音的最高頻率
- Center time: 鯨豚聲音總長度的一半
- Minimum frequency (F_{min}): 鯨豚聲音的最低頻率
- Ending frequency: 鯨豚聲音結束時的頻率

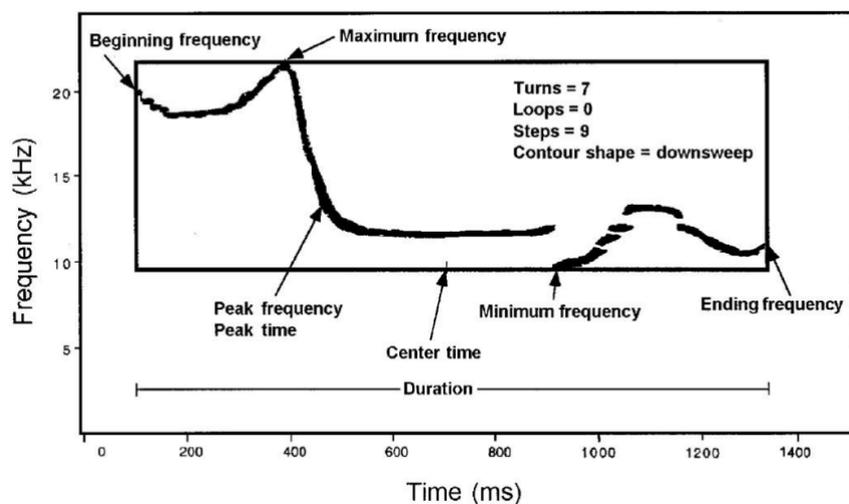


圖 1 鯨豚聲音頻譜圖(引用出處[19])

根據圖 1 所提供的定義，我們引用圖 2 文獻中的鯨豚哨叫聲，敘述如下：

- 平坦/恆定型：頻率變化小於 1kHz，其中頻率變化幅度不到總時長的四分之一。在整個哨叫聲持續時間內頻率變化很少保持恆定。
- 上升型：頻率主要呈現上升趨勢。如果存在任何拐點，則下降頻率部分 \leq 哨叫聲頻率跨距($F_{\max} - F_{\min}$)的一半。
- 下降型：頻率主要呈現下降趨勢。如果存在任何拐點，則上升頻率部分 \leq 哨叫聲頻率跨距($F_{\max} - F_{\min}$)的一半。
- 凹/U 型：至少包含一個拐點，其中頻率首先呈下降趨勢，然後呈上升趨勢。上升和下降部分佔哨叫聲頻率跨距($F_{\max} - F_{\min}$)的一半以上。
- 凸型：至少包含一個拐點，其中頻率首先呈上升趨勢，然後呈下降趨勢。上升和下降部分佔哨叫聲頻率跨距($F_{\max} - F_{\min}$)的一半以上。
- 正弦波型：內含兩個或多個頻率變化，至少包含兩個拐點，其中頻率首先呈上升趨勢，然後呈下降趨勢，依此類推，反之亦然。

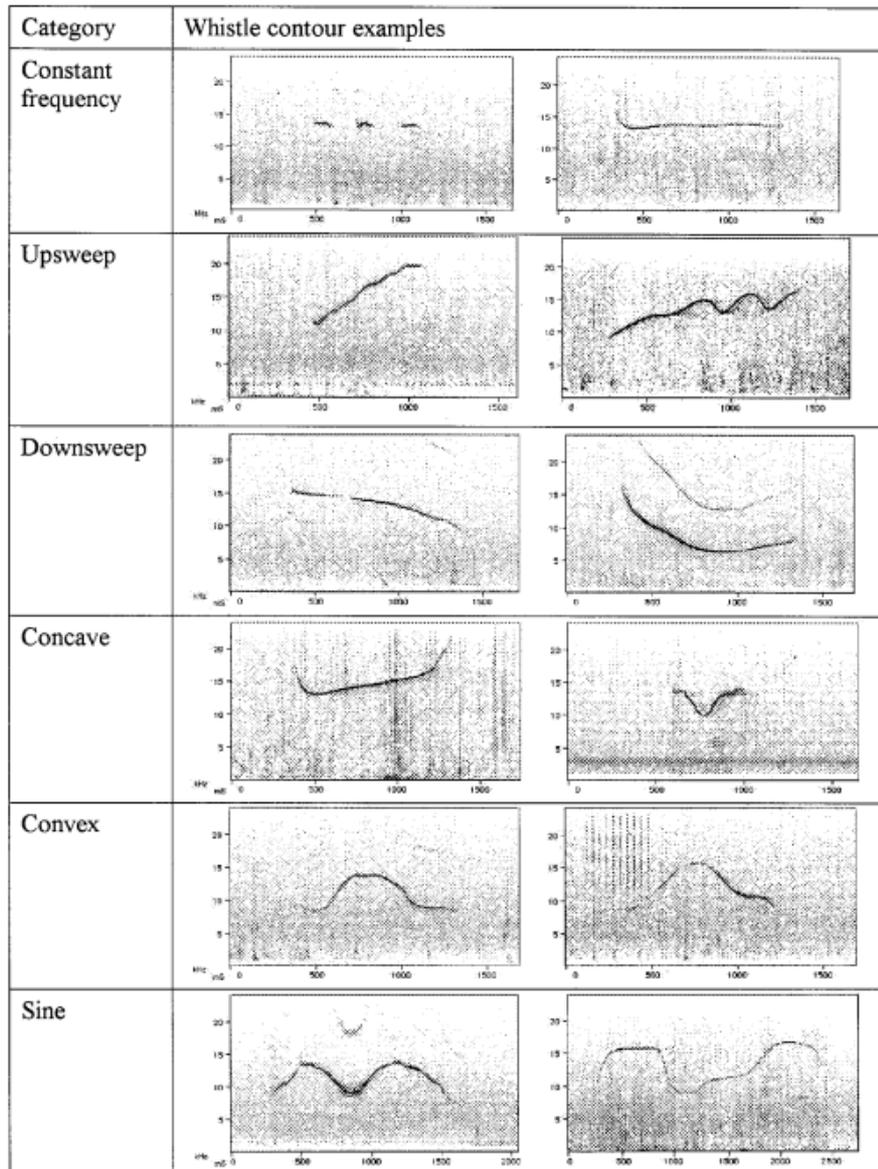


圖 2 鯨豚口哨聲的種類(引用出處[19])

被動式聲學監測聲景可以用來了解當地海域的氣候與海洋生態。然而，不同季節的海域生態以及氣象狀態也可能會改變不同季節的環境噪音特性。除了自然界所產生的聲音，海洋中的背景噪音在近年來隨著長途貨輪運輸的成長也逐年提升。在淺海的人為開發，更伴隨著填海、疏濬、水下打樁等工程產生的高強度噪音，這些噪音除了會增加當地海域環境的吵雜程度，也可能會傷害海洋生物的聽力[20, 21]。未來透過被動式聲學方法，將可協助相關單位掌握台灣鄰近海域的環境狀況，進而提供保育海洋生態、維持人為開發與保育平衡的重要監測平台。

第三節 預期目標與進度說明

本研究規劃之各項工作項目的時程和完成進度如圖 3 所示。在建立鯨魚聲音資料庫的主要來源有兩方面，一是由開放式資料庫 Kaggle 下載由 Google 研究

團隊所建立的資料外，二是由台灣大學工科海洋系陳琪芳特聘教授的研究團隊於雲林外海所搜集到的水下聲景資料庫，再由其中萃取出鯨魚的聲音檔案。接者由台灣大學船舶研究中心方銀營助理研究員與洋聲公司蔡孟汎總經理指導，識別出各種鯨魚聲紋的型態，以供後續深度學習模型訓練與識別。



圖 3 工作甘特圖

第二章 研究方法及過程

第一節 資料集標記

由於台灣海域的鯨豚聲音資料集極為難以收集與分析，故本研究以 Kaggle Whale Sound Dataset 混合台灣大學工科海洋系陳琪芳特聘教授的研究團隊耗時數年所收集到的音檔作為深度學習模型所需要的訓練與測試資料集。Kaggle Whale Sound Dataset 是公開的海洋哺乳動物聲學資料庫，主要收錄北大西洋露脊鯨的叫聲錄音。有標籤的訓練資料筆數為 30,436 筆。無標籤的測試資料筆數為 54,503 筆。使用 Raven lite 作為聲音的分析軟體，由康乃爾大學鳥類學實驗室所開發。

根據圖 2，將鯨魚聲紋的六種類別資料與兩種非鯨魚聲紋(船舶與雜訊)定義於表 1(label.csv)

表 1 特徵分類筆數

名稱	定義	資料筆數
no_signal	背景音	43
ship	船舶引擎	45
whale_constant	恆定頻率	43
whale_upsweep	上升頻率	135
whale_downsweep	下降頻率	131
whale_concave	凹形變化	131
whale_convex	凸形變化	132
whale_sinusoidal	正弦波狀	135

利用 Raven lite 讀入聲音檔(Wav file)，調整相關參數後進行頻譜分析而產生頻譜圖，並且能對聲音的頻譜標記或是做濾波分析，如圖 4 所示。分類與標記的結果如圖 5 及圖 6 所示。



圖 4 Raven Lite 操作環境

ST600HF-SN(7494)/SEN = -174.4 dB
2023/03/17 15:34:33

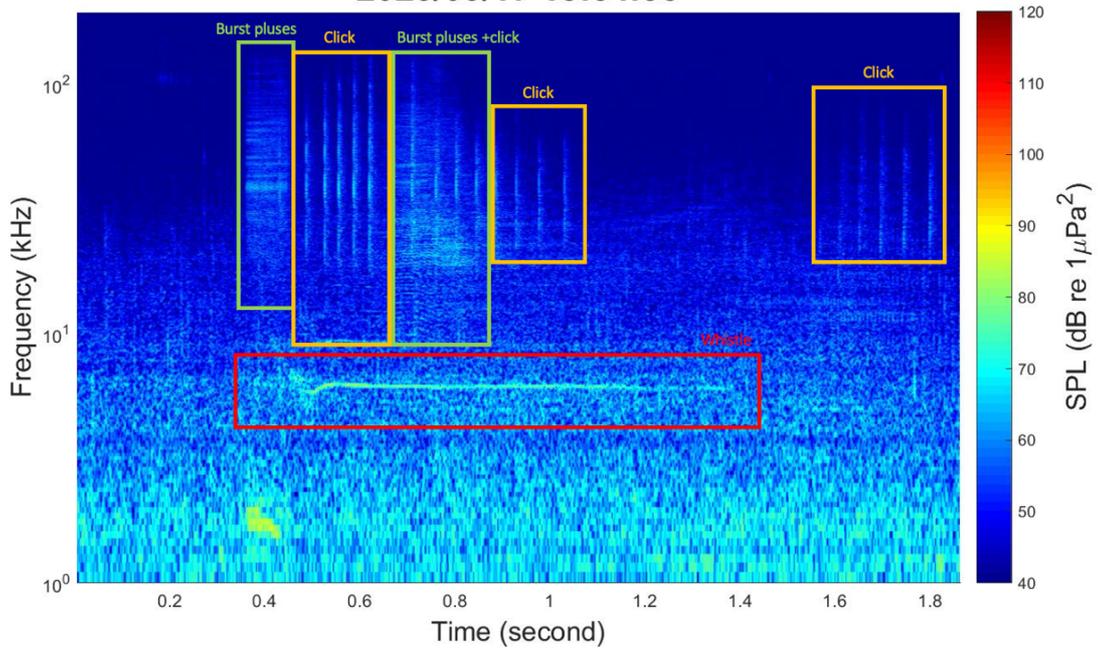


圖 5 標記聲紋種類

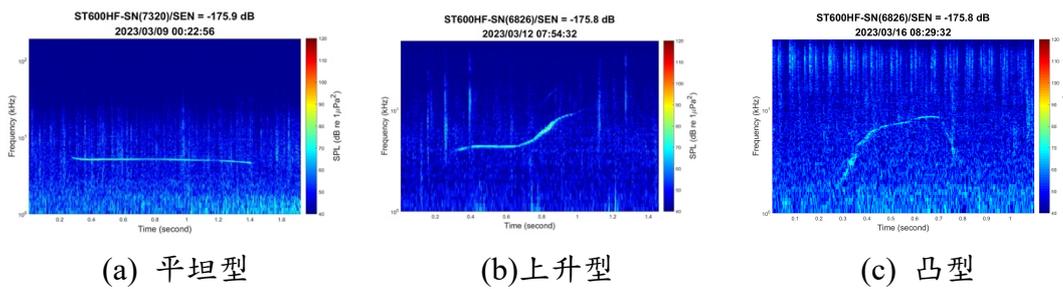


圖 6 標記哨聲聲紋中不同的種類

第二節 Faster Region-Based Convolutional Neural Networks (Faster R-CNN)

深度學習物件檢測演算法中，Faster Region-Based Convolutional Neural Networks (Faster R-CNN) [22]與 You Only Look Once (YOLO) [23]是兩種最具代表性的架構。兩者雖同屬於目標檢測領域，均致力於解決物體定位與分類的問題，但其基礎設計理念、模型架構以及運作方式有顯著差異。Faster R-CNN 是 Region-Based Convolutional Neural Networks 系列的第三代改進版本，其架構可分為三個主要模組：

1. 卷積特徵提取模組(Backbone)：Faster R-CNN 的第一步是通過卷積神經網絡(如 VGG16 或 ResNet)提取輸入影像的特徵圖。這一步為後續模組提供了高層次語義信息。
2. 區域候選生成網路 (Region Proposal Network, RPN)：Faster R-CNN 中的創新在於引入 RPN，負責從特徵圖中生成一組潛在的候選框。
 - Anchor 機制：RPN 使用預設的 Anchor 框（多種比例與尺度）來預測目標的邊界框。
 - Bounding Box Regression：同時學習調整框的位置信息以匹配真實目標。
 - 二分類任務：RPN 判定每個候選框是否包含物體（前景）或背景。
3. Fast R-CNN 偵測器 (RoI Pooling)：將 RPN 提供的候選框進一步細化：
 - 通過 RoI Pooling，將候選框映射到固定大小的特徵圖。
 - 應用全連接層進行分類（目標類別）及邊框回歸。

Faster R-CNN 的特點在於準確性高，由於透過兩階段的處理，能細緻地調整候選框並分類，適合需要高準確率的任務。然而速度相對慢，因需執行多步驟運算，尤其是 RPN 的候選框生成與篩選階段。與之相比，YOLO (You Only Look Once) 屬於一階段目標檢測模型，旨在提高目標檢測的速度。其主要架構如下：

1. 端到端的單一網路：YOLO 將目標檢測視為單一的回歸問題，直接從影像到目標的邊界框與類別預測。整個模型不需要額外的候選區域生成步驟。
2. 網路結構：
 - 網格劃分：將輸入影像分為 $S \times S$ 的網格，每個網格負責預測其範圍內的目標。

- 多任務輸出：每個網格輸出邊界框參數（中心點、寬、高）、置信度分數以及類別概率分布。

3. 損失函數：YOLO 的損失函數綜合了邊界框回歸損失、分類損失與置信度損失，以確保模型同時考慮位置精度與分類準確性。

表 1 列出 Faster R-CNN 與 YOLO 的特性比較。YOLO 的特點在於速度快，由於單步處理流程顯著縮短推論時間，適合即時應用場景。然而精度相對較低，對於小物體或重疊目標的檢測表現有限。若需要處理大量小物體或重疊的物體（如聲景或密集物體分佈），Faster R-CNN 的表現會更佳。分析本研究所應用的領域，對於準確性的重要性高於實時性，因此使用 Faster R-CNN 更為適合。

表 2 Faster R-CNN 與 YOLO 的特性比較

項目	Faster R-CNN	YOLO 系列
架構類型	兩階段 (Region Proposal + Detection)	一階段 (直接偵測)
速度	慢(需多步驟處理)	快(單次前向傳播即可完成)
準確性	高(特別是在小目標檢測上)	相對較低，尤其是小物體檢測
適用場景	精度要求高的場景，例如醫學、聲紋影像分析	實時應用，例如自駕車、監控系統
訓練成本	高(需要訓練 RPN 和分類模組)	相對較低，單一網絡訓練即可完成

第三節 Focal Loss 損失函數

透過降低大量易分類樣本(如 no_signal)的權重，Focal Loss 讓模型能夠更專注於學習那些難以區分或數量稀少的樣本(如特定鯨魚叫聲)，從而有效提升在極端不平衡資料集上的表現。Focal Loss 的基本架構如下：

以分類問題的真实類別為 y ，模型對「正類」的預測機率為 p 。定義：

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & y = 0 \end{cases}$$

傳統二元交叉熵是 $-\log p_t$ ，**Focal Loss** 在此基礎上加入兩個關鍵因子：

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t)$$

其中 $\gamma \geq 0$, focusing parameter，控制「壓抑容易樣本」的力度， $\alpha_t \in [0,1]$ ，**class-balancing weight**，處理類別不平衡（常對稀有/正類給較大的 α ）。

當樣本被分類得很準時， $p_t \rightarrow 1$ ，則 $(1 - p_t)^\gamma$ 會趨近 0，讓這些「容易樣本」的損失與梯度被大幅下調；而對困難樣本（ p_t 小），調降幅度很小，使其被聚焦學習。與其他不平衡對策的比較，具有以下特性：

- 重採樣 (over/under-sampling): 可改善不平衡, 但可能導致過擬合或丟資訊; Focal Loss 不改變資料分佈, 以損失函數直接重加權, 更穩定且計算友善。
- 類別權重交叉熵 (class-weighted CE): 僅依「類別」調權重; Focal Loss 進一步依難易程度(樣本層級調整), 有更細緻的學習信號。
- OHEM: 離散挑選困難樣本; Focal Loss 是連續權重化, 可微分、易併行、實作簡單。

Focal Loss 對極度不平衡資料特別有效: 顯著降低佔大多數的容易負樣本影響, 提升少數正樣本的學習效果。對於連續化的困難樣本挖掘, 不需額外的採樣策略或離散挑選, 簡潔且可微分。計算簡單, 只是在 CE 外包一個簡單因子, 幾乎無額外參數與記憶體負擔。泛化性強, 二元、多類別、檢測、分割 (含變體如 Focal Tversky、Focal Dice) 皆可用。Focal Loss = 交叉熵 * 根據樣本難度自動調權 * 類別不平衡再權重, 在不平衡情境下, 把訓練重點放在真正學不好的樣本上, 往往能帶來更好的檢出率與整體泛化。

第三章 研究成果

第一節 資料前處理

在如此大量的數據中識別感興趣的海洋物種是海洋生態學中重要的第一階段，可為更高層次的物種多樣性、行為特徵或海洋聲景分析提供輸入。然而，即使借助目前可用的電腦輔助方法，人工標記鯨豚的聲紋也極為耗時。

選擇將音訊事件偵測視為影像分類問題，其中影像是聲譜圖，繪製在時頻軸上的聲音功率直方圖。因此，本計劃本質上是一維時間序列的海洋聲音訊號，轉換為二維視覺化格式，並利用在圖像識別領域取得巨大成功的卷積神經網路(Fast-RCNN)進行有效區分。在本計劃中，使用了 34,000 個，時長為 2 秒(副檔名為 aiff) 聲音檔案樣本，取樣率為 2kHz。繼續將音訊轉換為梅爾頻譜圖以產生 255x255 .png 影像，然後產生 Tfrecords 格式以加快 GPU 處理的速度。最後將 Tfrecords 提供給 Faster-RCNN，以對影像中是否存在鯨魚聲紋進行分類。

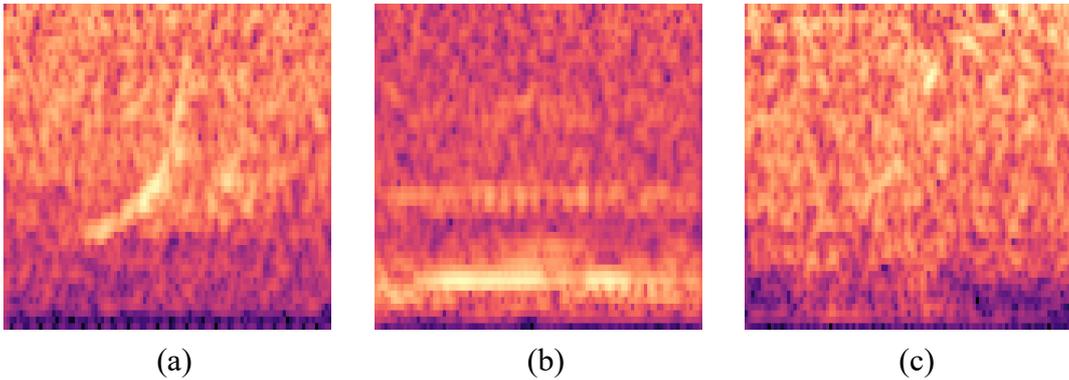


圖 7 鯨豚聲紋分類 (a)上升型 (b)平坦型 (c)非鯨豚

第二節 未使用 Focal Loss 分類結果

使用不同的指標來衡量實驗結果的性能，例如準確度(Accuracy)、精確度(Precision)、F 測量(F-Measure)和召回率(Recall)。指標的定義如下：

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = TPR = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Scope = \frac{2TP}{2TP + FP + FN} \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (5)$$

$$Specificity = TNR = \frac{TN}{TN + FP} \quad (6)$$

其中，FP, FN, TP 與 TN 分別是 False Positive、False Negative、True Positive 和 True Negative。

通過實驗對本研究所提出的方法進行了評估。對 6 種不同的鯨魚叫聲、船隻與海洋背景音進行細緻分類，結果如圖 8 所示。總體準確率為 36%，模型表現不佳，僅略優於隨機猜測，不具備實用價值。

	precision	recall	f1-score	support
no_signal	0.67	0.50	0.57	20
ship	0.35	1.00	0.52	20
whale_concave	0.21	0.25	0.23	20
whale_constant	0.40	0.10	0.16	20
whale_convex	0.30	0.35	0.33	20
whale_downsweep	0.42	0.42	0.42	19
whale_sinusoidal	0.00	0.00	0.00	12
whale_upsweep	0.00	0.00	0.00	12
accuracy			0.36	143
macro avg	0.29	0.33	0.28	143
weighted avg	0.33	0.36	0.31	143

圖 8 未使用 Focal Loss 之模型預測結果

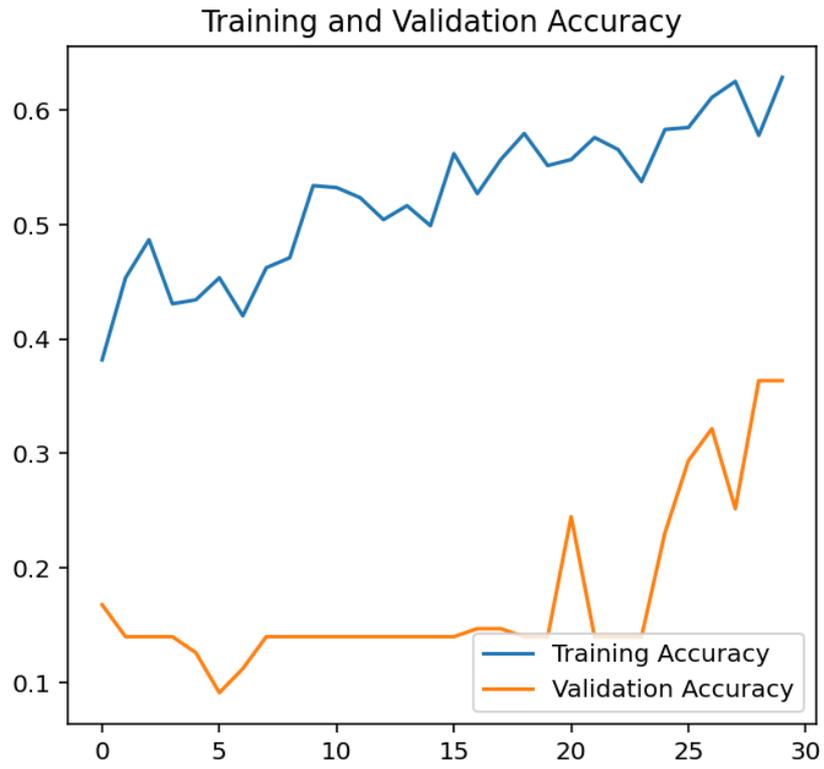


圖 9 訓練與驗證準確性

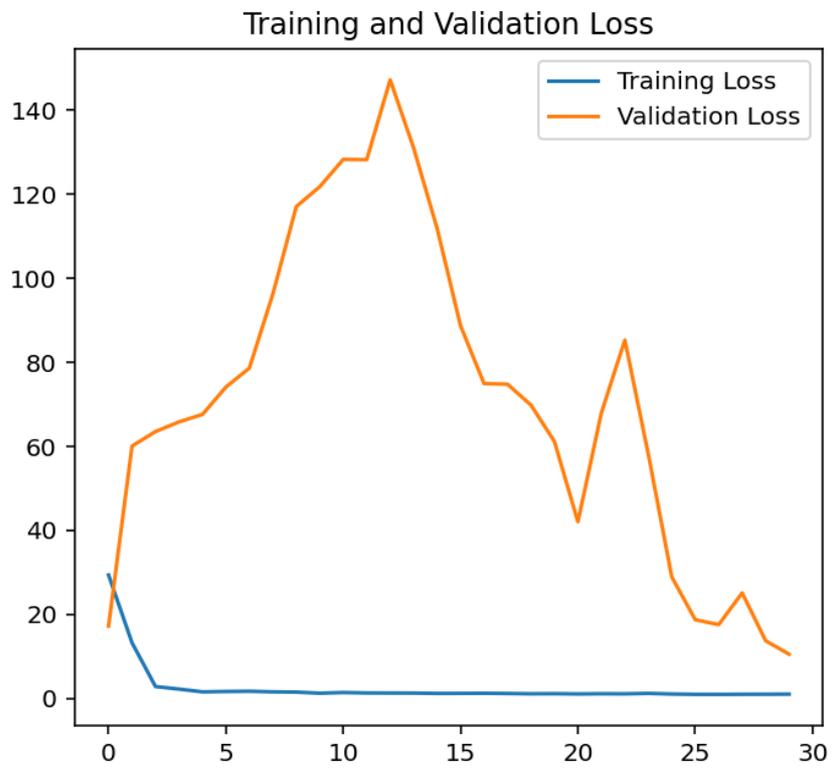


圖 10 訓練與驗證損失值

圖 11 為混淆矩陣分析，可以明顯地看出模型將大量不確定的聲音（包含多種鯨魚聲）錯誤地預測為 ship。對於樣本數最少 whale_upsweep 和 whale_sinusoidal，模型完全無法辨識(正確率為 0)。問題定義過於複雜，加上資料量不足與不平衡，導致模型訓練失敗。

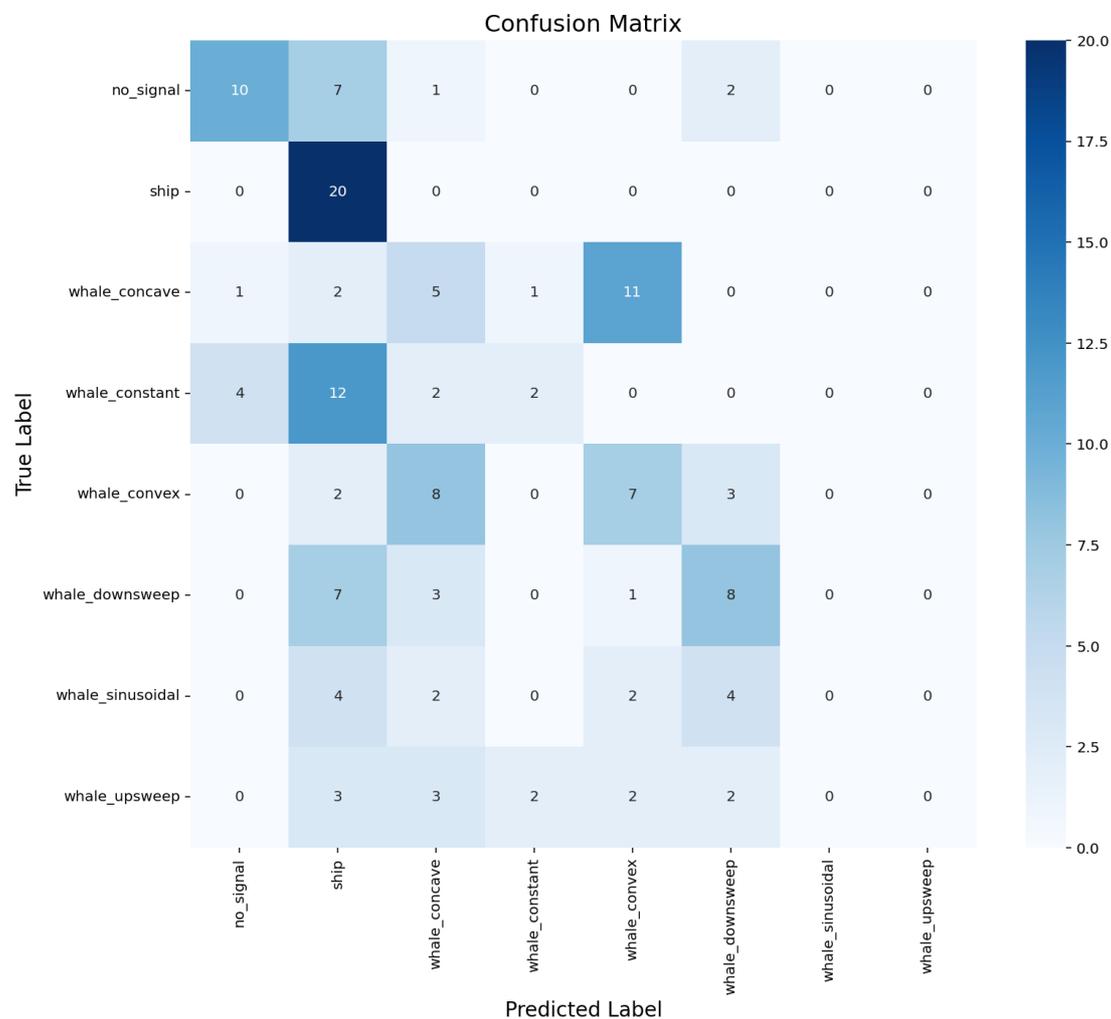


圖 11 未導入 Focal Loss 之模型混淆矩陣

第三節 使用 Focal Loss 分類結果

透過降低大量易分類樣本(如 no_signal)的權重，Focal Loss 讓模型能夠更專注於學習那些難以區分或數量稀少的樣本(如特定鯨魚叫聲)，從而有效提升在極端不平衡資料集上的表現。 γ (聚焦參數)，我們設定為 2.0。它用於調節易分類樣本的權重衰減速度。 α (類別權重)我們設定為 0.25，用於平衡正負樣本的重要性。系統 B 在更具挑戰性的八類別任務中表現穩定，測試集平均置信度達到 0.817。特別是在恆定頻率鯨魚叫聲(whale_constant)的檢測率上超過 85%，證明了其精準度分類能力。

圖 12 展示了不同批次大小（100、60 和 15）期間的分類性能。在每個批次大小，觀察訓練進度併計算所有個體評估的平均值以計算最佳結果評估。這種方法可以確保訓練盡快完成並防止模型過度擬合。結果表明，該模型對每個類別都具有良好的判別效能。所有的類別在最終準確率：約 99%（訓練集）、96%（驗證集）收斂特性：模型在約 60 個 epoch 後達到穩定，過擬合控制：驗證損失穩定，顯示良好的泛化能力。

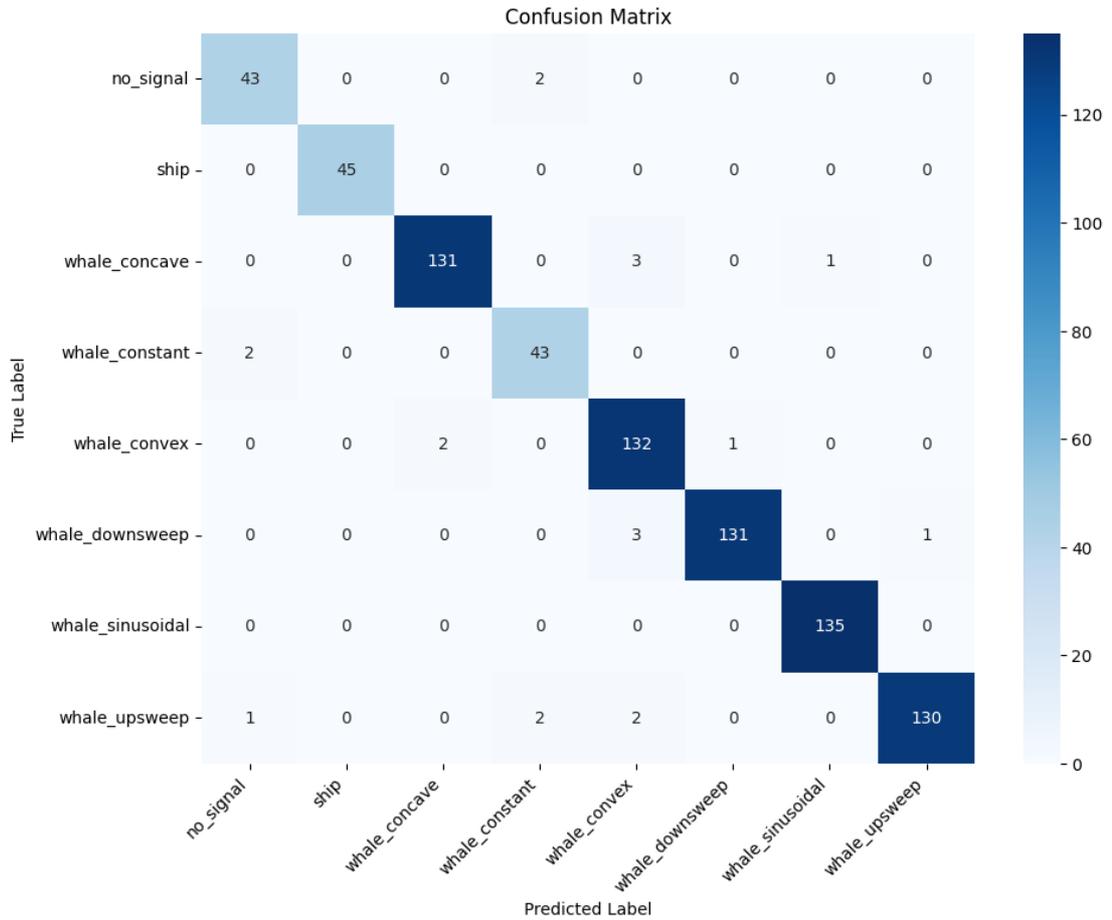


圖 12 導入 Focal Loss 之模型混淆矩陣

第四章 結論

本研究透過將聲音訊號轉換為頻譜圖，利用 Faster R-CNN 與 Focal Loss 確實能有效區分鯨魚與船舶的聲音。我們不僅實現了高準確度 (>98%) 的基礎分類，更進一步達成了鯨魚聲紋的精細識別。Focal Loss 處理不平衡資料的創新方法，克服了真實資料分佈的挑戰。

隨者 2050 年達成淨零排放的承諾越來越近，台灣努力讓人為造成的溫室氣體排放極小化，利用再生能源(風電、光電與地熱等)、負碳技術、森林碳匯等方法抵消碳排放量，達到淨零排放。然而不論是任何方法，在施工的過程不免會對環境造成一定的干擾，進而引發民眾的疑慮與抗爭。故本研究是以利用深度學習模型作為監測與分類水下聲景中鯨魚的聲紋，藉由聲紋的類型來識別鯨魚的種類、行為、數量，分析生物的生存環境是否有可能受到海岸工程之干擾。故本研究之研究成果可提供政府相關的機關作為參考。

致謝

本研究所使用之鯨豚錄音資料由台灣大學水下聲學實驗室陳琪芳特聘教授及所屬研究團隊、船舶海洋中心方銀營助理研究員與洋聲股份有限公司蔡孟汎總經理協助搜集。

參考資料

- [1] M.-N. Havlik, M. Predragovic, and C. M. Duarte, "State of play in marine soundscape assessments," *Frontiers in Marine Science*, vol. 9, pp. 919418, 2022.
- [2] T.-H. Lin, T. Akamatsu, and Y. Tsao, "Sensing ecosystem dynamics via audio source separation: A case study of marine soundscapes off northeastern Taiwan," *PLoS Computational Biology*, vol. 17, no. 2, pp. e1008698, 2021.
- [3] 經濟部能源署網站. "經濟成長力道強 2025 年再生能源裝置量目標不變."
- [4] 施仲平. "加油吧！你還能為鯨豚做更多——大中部離岸風電當今的環境課題," 10, 2025.
- [5] B. Mishachandar, S. Vairamuthu, and B. Selva Rani, "Identification and Recognition of Bio-acoustic Events in an Ocean Soundscape Data Using Fourier Analysis." pp. 275-284.
- [6] F. F. Dias, H. Pedrini, and R. Minghim, "Soundscape segregation based on visual analysis and discriminating features," *Ecological Informatics*, vol. 61, pp.

- 101184, 2021.
- [7] P. Ratilal, S. G. Seri, H. Mohebbi-Kalkhoran *et al.*, "Continental shelf-scale passive ocean acoustic waveguide remote sensing of marine ecosystems, dynamics and directional soundscapes: sensing whales, fish, ships and other sound producers in near real-time." pp. 1-7.
 - [8] B. Beckler, A. Pfau, M. Orescanin *et al.*, "Multilabel classification of heterogeneous underwater soundscapes with Bayesian deep learning," *IEEE Journal of Oceanic Engineering*, vol. 47, no. 4, pp. 1143-1154, 2022.
 - [9] B. Mishachandar, and S. Vairamuthu, "Diverse ocean noise classification using deep learning," *Applied Acoustics*, vol. 181, pp. 108141, 2021.
 - [10] L. Zeng, B. Sun, and D. Zhu, "Underwater target detection based on Faster R-CNN and adversarial occlusion network," *Engineering Applications of Artificial Intelligence*, vol. 100, pp. 104190, 2021.
 - [11] M. R. Ruiz, C. Rossi, and J. A. Esteban, "Fin whale pulse detection with deep neural networks," *Ecological Informatics*, vol. 77, pp. 102243, 2023.
 - [12] Y.-Y. Fang, C.-Y. Shih, M.-R. Zhong *et al.*, "Study of the underwater soundscape in coastal waters of Hsinta Power Plant," *The Journal of the Acoustical Society of America*, vol. 154, no. 4_supplement, pp. A275-A275, 2023.
 - [13] W. W. Au, and W. W. Au, "Characteristics of dolphin sonar signals," *The sonar of dolphins*, pp. 115-139, 1993.
 - [14] T.-H. Lin, H.-Y. Yu, C.-F. Chen, and L.-S. Chou, "Passive acoustic monitoring of the temporal variability of odontocete tonal sounds from a long-term marine observatory," *PLoS One*, vol. 10, no. 4, pp. e0123943, 2015.
 - [15] R. Ferrer-i-Cancho, and B. McCowan, "A law of word meaning in dolphin whistle types," *Entropy*, vol. 11, no. 4, pp. 688-701, 2009.
 - [16] R. A. Smolker, J. Mann, and B. B. Smuts, "Use of signature whistles during separations and reunions by wild bottlenose dolphin mothers and infants," *Behavioral Ecology and Sociobiology*, vol. 33, pp. 393-402, 1993.
 - [17] D. K. Mellinger, K. M. Stafford, S. E. Moore *et al.*, "An overview of fixed passive acoustic observation methods for cetaceans," *Oceanography*, vol. 20, no. 4, pp. 36-45, 2007.
 - [18] L. TH, "The application of passive acoustic monitoring for studying Indo-Pacific humpback dolphin behavior and habitat use off western Taiwan," National Taiwan University, 2013.
 - [19] C. Bazúa-Durán, and W. W. Au, "The whistles of Hawaiian spinner dolphins," *The Journal of the Acoustical Society of America*, vol. 112, no. 6, pp. 3064-3072, 2002.
 - [20] Y.-Y. Fang, P.-J. Sung, W.-C. Hu, and C.-F. Chen, "Underwater noise simulation of

impact pile driving for offshore wind farm in Taiwan," *Journal of Theoretical and Computational Acoustics*, vol. 28, no. 01, pp. 1950009, 2020.

- [21] W. J. Richardson, C. R. Greene Jr, C. I. Malme, and D. H. Thomson, *Marine mammals and noise*: Academic press, 2013.
- [22] S. Ren, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv preprint arXiv:1506.01497*, 2015.
- [23] J. Redmon, "You only look once: Unified, real-time object detection."