Universitat de les Illes Balears

TREBALL FI DE GRAU

APLICACIÓ DE LA INTEL·LIGÉNCIA ARTIFICIAL PER OBTENIR PARÀMETRES SOBRE L'ECOLOGIA DE XYRICHTYS NOVACULA

Antonio Ginard Matas

Grau de Biologia

Facultat de Ciències

Any Acadèmic 2019-20

APLICACIÓ DE LA INTEL·LIGÉNCIA ARTIFICIAL PER OBTENIR PARÀMETRES SOBRE L'ECOLOGIA DE *XYRICHTYS NOVACULA*

Antonio Ginard Matas

Treball de Fi de Grau

Facultat de Ciències

Universitat de les Illes Balears

Any Acadèmic 2019-20

Paraules clau del treball:

Peix, Deep Learning, Monitoreig, *Xyrichtys novacula*

Josep Alós, Instituto Mediterráneo de Estudios Avanzados IMEDEA (CSIC-UIB)

Guillem Mateu, doctor ponent de la UIB

S'autoritza la Universitat a incloure aquest treball en el Repositori Institucional per a la seva consulta en accés obert i difusió en línia, amb finalitats exclusivament acadèmiques i d'investigació	Autor		Tutor	
	Sí	No	Sí	No
	☒	☐	☒	☐

Índex

Resum 1

Abstract 1

1 Introducció 2

2 Material i mètodes 7

2.1 Obtenció d'imatges amb càmeres amb esquer submarines 7

2.2 Selecció i construcció de l'algoritme (xarxa neuronal) per identificar de manera automàtica el raor 9

2.3 Breu resum del funcionament de la xarxa: la xarxa de propostes de regió 10

2.4 Classificació dels objectes a través de la Fast R-CNN i formació de la xarxa 11

2.5 Homogeneïtzació d'imatges 12

2.6 Anotació i classificació d'imatges 13

2.7 Nombre d'imatges necessari i procés d'augmentació 13

2.8 Validació de la xarxa 14

3 Resultats 15

4 Discussió 21

5 Conclusió 24

6 Agraïments 24

7 Bibliografía 24

Resum

La intel·ligència artificial (IA) està de cada vegada més present en tot els àmbits de la ciència. En el cas de l'ecologia, la seva utilització ha suposat una gran millora en els darrers cinc anys, en l'obtenció de dades d'abundància i diversitat d'espècies amb l'objectiu de millorar-ne la seva conservació. En aquest treball s'ha desenvolupat un algoritme de *Deep Learning* (branca del entrenament de les màquines de la IA) pel seguiment automàtic de poblacions de peixos en el medi marí utilitzant càmeres submarines. El nostre cas d'estudi es centra en el raor *(Xyrichtys novacula),* una espècie d'alt interès pesquer per les Illes Balears del qual és necessari millorar-ne el seu seguiment. Per aconseguir-ho, vàrem entrenar tres versions de la xarxa de *Deep learning* anomenada *Faster R-CNN* amb la funció de detecció automàtica i no supervisada del raor: una sense augmentació i dos llindars d'augmentació per tal d'observar quina podria donar millors resultats. Després d'una validació on es comparava la classificació de les diferents versions de la xarxa amb una classificació feta visualment, vàrem observar que la xarxa amb major percentatge d'encert fou la xarxa sense augmentació amb una taxa del 74% en la detecció del raor. Aquest resultat suggereix que el nostre algoritme basat en la xarxa *Faster R-CNN* és una eina prometedora per automatitzar el seguiment de les poblacions de raors de les Illes Balears i contribuir així a la seva gestió i conservació. Aquest treball, a més, estableix les bases per la millora de l'algoritme i fer-lo extensible a altres espècies de peixos de les Illes Balears o fins i tot terrestres, amb l'objectiu de dissenyar estratègies de seguiment automàtic de la biodiversitat mitjançant la IA.

Abstract

The use of Artificial intelligence (AI) is growingin all areas of science. In the case of ecology, the use of AI has increased exponentially in the last five years, mainly to generate and analyze massive data sets of abundance and diversity of species. In this work, a Deep Learning algorithm (as a branch of AI) has been developed for the automatic monitoring of fish populations in the marine environment using underwater cameras. Our case study focuses on the raor (*Xyrichtys novacula*), a species of high fishing interest in the Balearic Islands whose monitoring needs to be improved. To achieve this, we have trained three versions of the Deep

learning network called Faster R-CNN with the function of automatic and unsupervised detection of fish: one without augmentation and two more with two augmentation thresholds in order to observe which could give the best results. After a validation, we have observed that the network with the highest percentage of success was the network without increase with a rate of 74% in the detection of the raor. This result suggests that our algorithm based on the Faster R-CNN network is a promising tool to automate the monitoring of the raor populations of the Balearic Islands, and thus contribute to their management and conservation. This work also establishes the basis for improving the algorithm and making it extensible to other fish species in the Balearic Islands or even terrestrial, with the aim of designing strategies for automatic monitoring of biodiversity through AI.

1 Introducció

La pesca recreativa és una activitat pesquera no professional, on es pesquen una sèrie d'organismes marins, en general peixos, amb uns fins d'esbarjo, esport o en ocasions turisme (Arlinghaus et al., 2019). A les Illes Balears, aquesta activitat d'esbarjo és practicada entorn del 5-10% de la població, oferint multitud de beneficis psicològics, socials i econòmics als ciutadans de les illes (Alós et al., 2016; Morales-Nin et al., 2005). La pesca recreativa és una pesca sense ànim de lucre, sent la seva principal motivació gaudir de la natura. A causa del notable creixement a escala mundial, nacional i local d'aquesta activitat, estudis recents han demostrat que de vegades la pesca recreativa explota els recursos marins a nivells que no poden ser menyspreats, requerint plans específics per la seva gestió per fer de la pesca recreativa una activitat sostenible (Cabanellas-Reboredo et al., 2016; Font & Lloret, 2014).

A una àmplia zona del litoral mallorquí es permet la pràctica de la pesca recreativa existint zones on aquesta està regulada principalment a través d'Àrees Marines Protegides (AMP) (Figura 1). Per exemple, Alós et al. (2016) ha estimat recentment que la ràtio d'explotació anual del raor, *Xyrichtys novacula* (una espècie de peix d'alt valor econòmic) a una AMP de les Illes Balears és del 55%, suggerint una reducció de la població d'aquesta espècie de fins al 45% per la pesca recreativa. Les zones on es practica la pesca recreativa, així com altres activitats recreatives, inclou zones de la Xarxa Natura 2000 (Tsiafouli et al., 2013). La Xarxa Natura 2000 és una xarxa ecològica europea d'àrees de conservació de la

biodiversitat que consta d'una sèrie d'àrees especials: Zones Especials de Conservació (ZEC) establides d'acord amb la Directiva d'hàbitat i Zones d'Especial Protecció per Aus (ZEPA). La seva finalitat és assegurar la supervivència a llarg termini de les espècies i tipus d'hàbitats arreu d'Europa i així aturar la pèrdua de biodiversitat (Evans, 2012). La Xarxa Natura 2000 és el principal instrument per a la conservació de la naturalesa a la Unió Europea (Blicharska et al., 2016).

Monitoritzar la pesca recreativa, així com el de qualsevol activitat recreativa que és realitza dintre de la Xarxa Natura 2000 és fonamental per poder complir amb els seus objectius de conservació d'aquesta figura de protecció (Balaguer et al., 2011; Farías Torbidoni, 2011). No només per la Xarxa Natura 2000 és important monitoritzar la pesca recreativa per garantir la seva sostenibilitat, sinó que la Política Pesquera Comú (PPC) cita textualment que; "*Donat que la pesca recreativa pot tenir un impacte significatiu en els recursos pesquers, els Estats membres deuen assegurar que es realitza de forma compatible amb els objectius de la política pesquera comú*" (Reglament UE Nº1224/2009) (*EUR-Lex - 32009R1224 - EN - EUR-Lex*, 2009). Per aquest motiu en el seu Art. 55 (Zeller et al., 2008) estableix que els seus estats membres han d'assegurar-se del control i vigilància de la pesca recreativa i del monitoreig de la salut de les poblacions d'espècies objectiu en el seu territori i a les seves aigües comunitàries (Holden & Garrod, 1996). És per tant fonamental dissenyar estratègies de seguiment de la pesca recreativa tant per establir estratègies de conservació de la biodiversitat, com per assegurar el futur d'una activitat que aporta un elevat valor social a les Illes Balears.

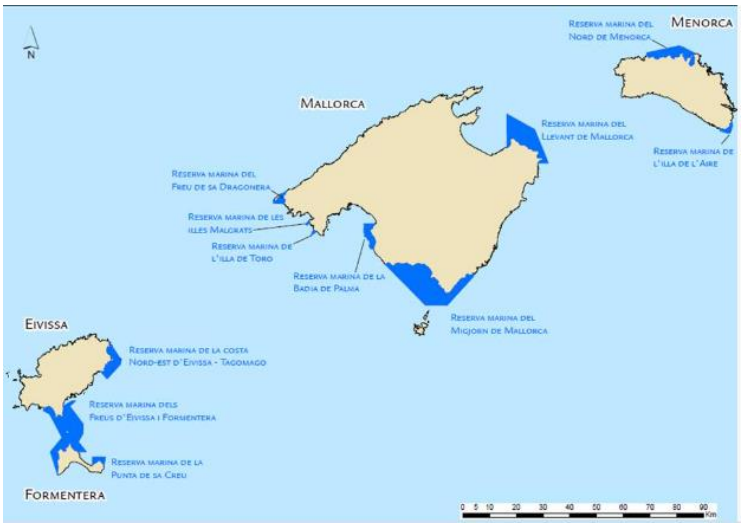

Figura 1. Àrees Marines Protegides (AMP) de les Illes Balears, on trobem a Menorca la Reserva marina del nord de Menorca i la Reserva marina de l'Illa de l'aire. A Mallorca; la Reserva marina del llevant de Mallorca, la

Reserva marina del Freu de sa Dragonera, la Reserva marina de les Illes Malgrats, la Reserva marina de l'Illa del Toro, La reserva marina de la Badia de Palma i la Reserva Marina del Migjorn de Mallorca. A Eivissa i Formentera trobem la Reserva marina nord-est d'Eivissa i Tagomago, la Reserva marina dels Freus d'Eivissa i Formentera i la Reserva marina de la punta de la creu. Totes elles formen part de la Xarxa Autonatura2000 (*Les Reserves Marines a Les Illes Balears-Les Reserves Marines a Les Illes Balears*, 2019).

Malauradament la recopilació i la gestió de dades procedents de la pesca recreativa és una qüestió extremadament complexa (Ryan et al., 2016; Zeller et al., 2008). La gran magnitud social de la pesca recreativa, l'elevada heterogeneïtat, les dificultats del seu control i del monitoreig espacial, fa que sigui uns dels majors reptes de la PPC i de la gestió de les pesqueries a escala global (Arlinghaus et al., 2019). Una de les grans necessitats per poder aconseguir aquests objectius seria desenvolupar noves estratègies i metodologies per la caracterització i l'avaluació de la pesca recreativa a escala europea. Malgrat els esforços de la comunitat científica, en moltes ocasions recolzades per l'administració local, per millorar la recopilació y gestió de dades procedents de la pesca recreativa, aquest procés segueix sent uns dels majors reptes a nivell científic per a l'aplicació de la PPC. El repte és especialment complex pel que fa a la monitorització de les poblacions de peixos explotats per la pesca recreativa (biomassa, nombre d'individus distribució de talles, diversitat d'espècies...) a causa de l'absència de sistemes automàtics de monitoreig a curt i a llarg termini (Post et al., 2011). Obtenir aquest tipus de dades de les poblacions de peixos és un aspecte clau per aconseguir els objectius del Fons Europeu Marítim de Pesca (FEMP) i de la PPC per aconseguir una activitat pesquera socialment i mediambientalment sostenible, així com per garantir el manteniment i potenciació de la biodiversitat i els serveis ecosistèmics en la Xarxa Natura 2000.

Fins fa només uns anys l'única manera d'obtenir aquest tipus de dades era amb censos visuals amb escafandre autònom (Abelson & Shlesinger Abelson, 2002; Naga et al., 1994.) metodologia que resulta poc cost-efectiva, ja que es requereix una infraestructura important pel volum de dades que podies arribar a obtenir (Brock, 1982). Un altre inconvenient és la pertorbació de l'ecosistema per part del mostrejador, ja que les espècies objectiu poden alterar el seu comportament per la seva presencia (Edgar et al., 2004). Amb l'aparició de les primeres càmeres submarines i la seva utilització per primera vegada en ecologia marina l'any 1950 (Barnes, 1952) es va trobar una nova eina de treball digital dins aquesta disciplina. En els anys 60, es van desenvolupar diferents sistemes que eren capaços

d'enregistrar a la columna d'aigua des de la superfície al fons amb una càmera que recorria uns raïls i que anava enregistrant contínuament (Lafond, 1961). Un altre sistema que va sorgir fou *el Baited Remote Underwater Video* que consisteix en posar una o dues càmeres davant un esquer atraient per les espècies de peixos objectiu, a una distància entre 0.5 i 1.5 metres (Ellis & DeMartini, 1995; Heagney et al., 2007; Willis & Babcock, 2000) (Figura 2).

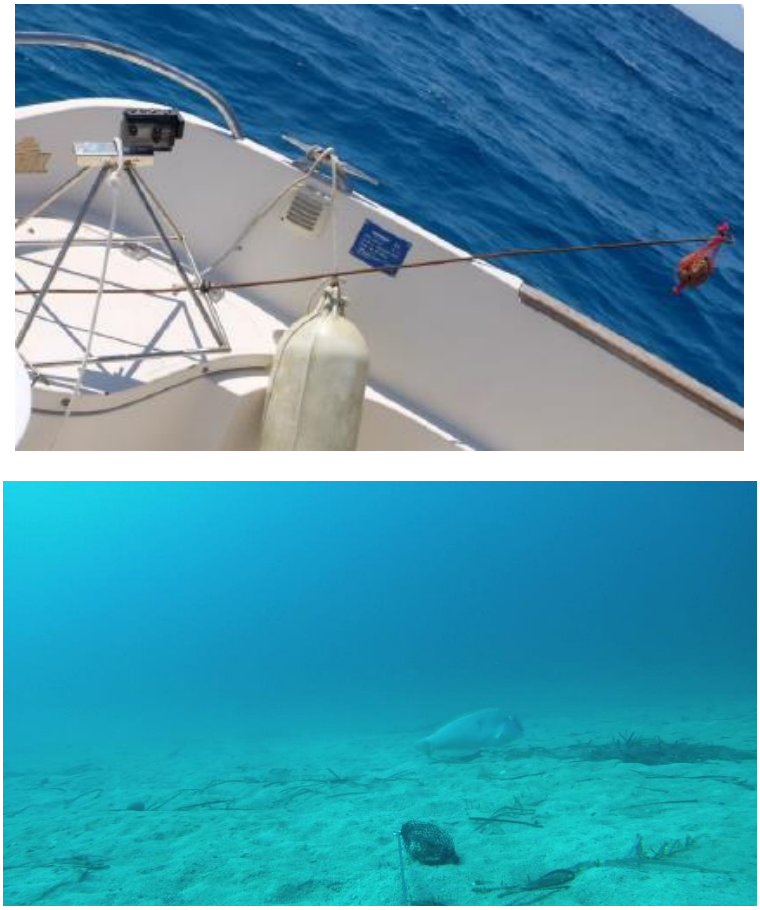

Figura 2. A la part superior mostra l'estructura de les *Baited Remote Underwater Video Cameras* on es van enregistrar els vídeos per a l'experiment. La figura inferior podem veure el plànol del fons marí obtingut amb la càmera, i podem identificar un exemplar de raor.

El seguiment de la pesca en general i de les poblacions de peixos explotades s'ha vist recentment beneficiat per aquests avanços tecnològics basats amb l'adquisició i anàlisi d'imatges i vídeos d'alta definició (Álvarez-Ellacuría et al., 2019). La utilització de càmeres submarines ens permet monitoritzar les poblacions de peixos amb objectius de sostenibilitat pesquera, facilitant també per l'aparició i abaratiment de càmeres d'acció de tipus *go-pro,* ha motivat la generació de gran quantitat de material audiovisual. Aquesta adquisició de grans quantitats d'imatges de peixos facilita molt l'aprenentatge de màquines (*Machine learning*) que permet, una vegada "entrenat", als ordinadors classificar automàticament peixos per

obtenir dades com l'abundància o la mida a través de l'anàlisi d'imatges (Marini et al., 2018; Siddiqui et al., 2018). Aquest fenomen ha permès augmentar considerablement l'esforç de mostreig, ja que permet mesurar de manera automàtica i no supervisada les poblacions de peixos marins. Dintre d'aquest camp en els últims anys el *Deep Learning* (DL) està generant una revolució pel seu gran grau de precisió i classificació.

El DL ha irromput en la vida quotidiana i en moltes disciplines científiques (Goodfellow et al., 2016). Aquesta eina analítica aporta un alt rendiment i flexibilitat per a infinitat de processos com per exemple la construcció de vehicles autònoms. Els dos fets principals que donen el gran potencial al DL per resoldre problemes quotidians són: i) la era *big data* en la qual vivim on disposem d'una gran capacitat d'informació (com en el cas de boom de l'ús de les càmeres per monitoritzar poblacions de peixos) i ii) un augment potent de la capacitat informàtica dels equips en què es fa feina (M. W. Mathis & Mathis, 2020). Inicialment el DL es trobava de forma esporàdica en la disciplina de l'ecologia, però en els darrers cinc anys l'aplicació del DL a ecologia ha tingut un creixement exponencial (Christin et al., 2019). S'ha utilitzat en estudis molt diferents com per a la identificació d'imatges i classificació, estimacions de densitat i diversitat, conservació ambiental, recursos de gestió i estimacions de la mida de l'animal (Humphries et al., 2018). Cal destacar que és possible l'ús d'aquesta eina tant al laboratori com al camp obtenint una gran quantitat de dades de forma automàtica. Durant els darrers anys, aquesta eina està sent molt utilitzada en ecologia del comportament per totes els avantatges anteriorment esmentats, i pel fet que permet generar dades massives de comportament sense gairebé esforç per part de l'investigador (Owens & Rennison, 2017).

Aquest treball té com a objectiu estendre les aplicacions del DL en l'ecologia mitjançant el desenvolupament i la validació d'un sistema de seguiment automàtic d'espècies explotades per la pesca recreativa a la Xarxa Natura 2000 marina utilitzant càmeres d'alta definició i un algoritme d'intel·ligència artificial de DL. L'objectiu específic d'aquest treball és crear una xarxa neuronal artificial (DL) que ens permeti detectar i identificar de manera automàtica espècies de peixos marins explotats per la pesca recreativa utilitzant càmeres submarines per a obtenir de manera no-supervisada estimes de densitat d'individus dintre de la Xarxa Natura 2000. Per l'estudi ens hem centrat en el cas de les poblacions de raors, *Xyrichthys novacula* – l'espècie més preuada de la pesca recreativa a les Illes Balears (Alós et al., 2016). La informació generada de manera automàtica per la nova metodologia permetrà

conèixer l'estat de les poblacions de les Illes i generar tot un conjunt d'informació necessària per poder assolir els objectius de sostenibilitat impulsats per la PPC i l'efectivitat de la Xarxa Natura 2000.

2 Material i mètodes

El DL aplicat al seguiment de les poblacions de peixos es basa en un algoritme d'anàlisi d'imatges (una xarxa neuronal artificial) que imita la interacció d'una xarxa neuronal del cervell humà i que ens permet dur a terme diverses funcions, en el nostre cas identificar i classificar un peix de manera automàtica. El sistema intenta simular el procés d'una xarxa neuronal d'un cervell humà en període d'aprenentatge (A. Mathis et al., 2018). Una neurona artificial és una funció matemàtica (algoritme) que s'alimenta d'una o més entrades i la seva suma produeix una sortida d'informació cap a una altre neurona a través de capes. Les connexions entre neurones són transportades per inputs, els quals son valors numèrics que influeixen en la transmissió de la informació. Aquests inputs fan que l'algoritme recopili aquesta informació i pugui "aprendre" de, per exemple, imatges etiquetades de peixos ja classificats. Quan una d'aquestes xarxes aprèn, cerca un cert conjunt d'inputs que li permet realitzar una funció específica, en el nostre cas classificar i identificar espècies de peixos marins. En el context d'aquest treball, una entrada (una imatge) s'analitza a través de diferents capes neuronals y genera una capa de sortida que prediu si trobem un peix o no, i de quina espècie es tracta (en el nostre cas un raor). Amb el desenvolupament d'aquesta eina de DL, es pretén convertir el monitoreig de les poblacions de peixos en un procés automàtic i sense supervisió humana.

2.1 Obtenció d'imatges amb càmeres amb esquer submarines

Amb l'objectiu de generar un algoritme de DL per classificar de manera automàtica peixos marins utilitzant càmeres submarines, el primer punt és la recopilació d'imatges per entrenar l'algoritme. Des de l'any 2011 el Grup d'Ecologia de Peixos de l'IMEDEA (UIB-CSIC) utilitza càmeres de vídeo submarines amb esquer (*Baited cameras*) per obtenir dades sobre la distribució i concentració de peixos vulnerables per a la pesca recreativa. Les càmeres estan compostes per una estructura metàl·lica on trobem una càmera amb una carcassa subaquàtica i un màstil metàl·lic on es troba una xarxa on es diposita l'esquer per atreure els

peixos dels voltants (Figura 2 i 3). Amb aquesta estructura l'esquer queda visible en tot moment en el marc de la gravació de la càmera, de manera que els peixos se senten atrets per l'olor i la vista de l'esquer i d'aquesta manera es poden quantificar. Aquest mètode és un mètode no invasiu que permet generar dades com índex d'abundància relativa i diversitat dels peixos amb poques pertorbacions al seu medi natural (Alós et al., 2016).

Figura 3. Peixos de l'espècie *Balistes carolinensis* atrets per l'esquer i enregistrats per una càmera submarina típica per avaluar la densitat de peixos dintre de la Xarxa Natura 2000.

Amb la col·lecció dels vídeos obtinguts entre els anys 2011 i 2019 pel Grup d'Ecologia de Peixos del IMEDEA, es va realitzar una base de dades de vídeos submarins (exemples, Figures 2 i 3). Aquesta base de dades conté 932 vídeos de duració d'una hora de mitja, els quals varen ser analitzats visualment per determinar el nombre de peixos i espècies presents a cada gravació. Aquesta base s'utilitzarà per elegir a partir de quins vídeos s'extrauran imatges que s'utilitzaran per entrenar les xarxes de DL i les seves posteriors validacions una vegada la màquina detecti automàticament el nostre peix objectiu. Per entrenar la nostra xarxa , es van capturar captures de pantalla de vídeos mitjançant el reproductor multimèdia VLC (*VLC - Features - VideoLAN*, 2020). Es van utilitzar 25 vídeos de la base de dades per generar 9,208 imatges on hi apareixien raors, que varen servir per entrenar la xarxa de DL que volem generar (veure secció entrenament de la xarxa).

A partir de les imatges obtingudes de les càmeres subaquàtiques es va crear una base de dades formada per imatges on apareix un raor (*Xyrichtys novacula*). El raor és un peix molt popular a les Illes Balears, sobretot entre els pescadors recreatius. Aquests peixos es troben a zones arenoses i d'aigües temperades. La nostra zona d'estudi compleix aquestes característiques, es tracta d'una àrea natural protegida (AMP) de la Badia de Palma i que

pertany a la Xarxa Natura 2000 marina. És una zona especialment caracteritzada per l'abundància de raors, ja que és una zona amb les característiques ambientals idònies, i a part es tracta d'una AMP on es regula la pesca recreativa (Alós et al., 2016) (Figura 4).

Figura 4. Distribució de la Reserva Marina de la Badia de Palma que forma part de la Xarxa Natura 2000 on es va desenvolupar aquest Treball de Fi de Grau.

2.2 Selecció i construcció de l'algoritme (xarxa neuronal) per identificar de manera automàtica el raor

El segon punt en la creació de l'algoritme de DL per monitoritzar les poblacions de peixos és dissenyar la nostra xarxa neuronal que servirà per classificar imatges. Es va utilitzar una xarxa anomenada *Faster R-CNN* que prové d'una biblioteca de codis oberta i l'entrenarem utilitzant *Python 3* (Ren et al., 2017). Amb aquesta acció disposem d'una xarxa neuronal convolucional pre-entrenada, això significa que aquesta xarxa ja ha estat entrenada amb un gran conjunt de dades (un paquet COCO (Redmon & Farhadi, 2016)) abans de la nostra utilització pel nostre cas d'estudi. Aquest fet ens permet usar menys dades per a l'entrenament, el que fa que ens faciliti l'esforç davant la creació d'una xarxa òptima pels nostres objectius.

El nostre objectiu d'entrenament de la xarxa és determinar les caixes de delimitació al voltant dels objectes que ens interessen quan apareixen a la imatge (Figura 2 i 3). La manera amb què es delimita aquest quadre consisteix a establir quatre punts de límit en forma de coordenades de píxels de la imatge extreta (Figura 5). Al DL sempre parlem de probabilitats, una xarxa DL ens dona una probabilitat del resultat de la funció que hem entrenat, en el nostre cas, la presència d'un peix. Al final el resultat és un mapa de característiques que tenen unes

dimensions espacials més petites que la imatge original però tenen una major profunditat. Un detall important és que la xarxa manté la ubicació dels objectes que ha codificat en relació amb la imatge original.

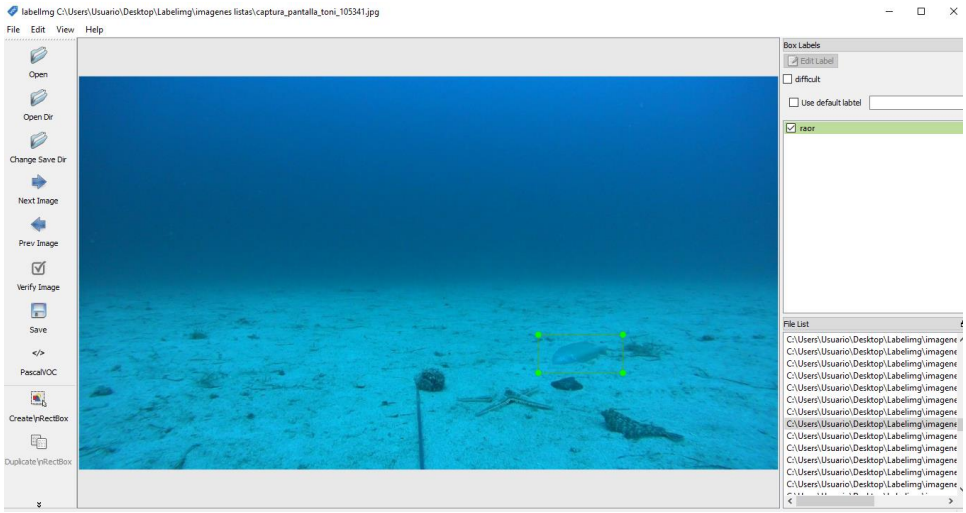

Figura 5 . Procés d'etiquetatge de les imatges amb el programa LambelImg.

2.3 Breu resum del funcionament de la xarxa: la xarxa de propostes de regió

La xarxa base genera un mapa de funcions per treballar. El nostre objectiu en general és determinar les caixes de delimitació al voltant dels objectes que ens interessen quan apareixen a la imatge. La manera amb què es delimita aquest quadre consisteix a establir quatre punts de límit en forma de coordenades de píxels de la imatge extreta com ja s'ha comentat a l'apartat 2.2 (Figura 5).

La xarxa de propostes de regió, juntament amb els ancoratges, és la gran novetat del *Faster R-CNN* comparant-ho amb els seus predecessors.

El *RPN* (*Region Proposal Network*) són un conjunt de regions propostes per a la detecció de l'objecte i utilitza el mapa de funcions generat amb la xarxa base per trobar entre tots els ancoratges alguns on puguem trobar un objecte. Durant l'entrenament, l'*RPN* agafa tots els ancoratges i els divideix en dues categories, les que podem trobar més probablement al primer pla i les que probablement les trobem al fons. Llavors obtindrem una llista d'objectes rellevants, cadascun amb una proposta de coordenades i una "objectivitat". Per poder arribar a classificar les opcions trobades al pas anterior es necessita un nou mapa de funcions, extret de les regions de la imatge que conté cada caixa de delimitació que la xarxa ens proposa per a cada objecte (Figura 6). Si tenim en conta la gran quantitat de propostes que tenim per cada

imatge ho convertiria un procés molt lent. Denominem captures de la Comunicació de la Regió d'Interès (ROI) a les regions que tenen una mida fixa que correspon a les caixes de delimitació de cada proposta i les envia al següent cas on es produeix la veritable classificació.

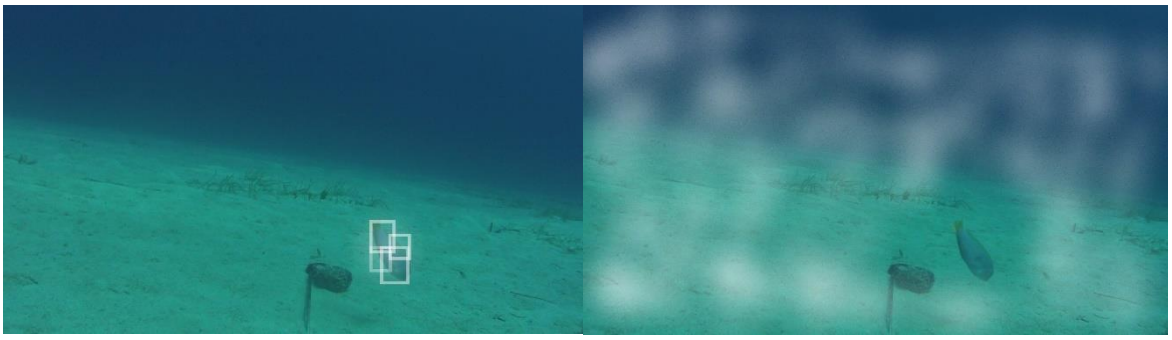

Figura 6. A la imatge de l'esquerra podem veure les caixes de proposta aplicada al raor que s'ha enregistrat. I a la imatge de la dreta veiem una proposta de fons proposada per la xarxa.

2.4 Classificació dels objectes a través de la Fast R-CNN i formació de la xarxa

La xarxa Neural Convolucional és l'últim pas de la *Fater R-CNN.* Aquesta crea un mapa de funcions a partir d'una imatge amb la xarxa que prèviament hem entrenat amb la versió d'*RPN*. La *Faster R-CNN* és una xarxa que utilitza funcions calculades anteriorment i que pugin trobar-s'hi objectes que ens interessin; per tant té dos objectius finals: assignar una classe a cada proposta i ajustar el límit de les caixes per a les propostes segons la classe prevista. La *Faster R-CNN* pren les propostes i els objectes de veritat fonamental i les compara amb un valor. Si la proposta té un valor superior a 0.6 s'assignaran a la seva veritat de què és l'objecte, detecta el peix, metres que les propostes entre 0.1 i 0.5 les considera com a fons. Per això, divideix les propostes amb dos grups, un on només conté propostes de fons i l'altre on només conté propostes de primer pla. A partir d'aquests dos grups crea un conjunt de dades compostes per un 25% per propostes de primer pla i un 75% de propostes de fons (Figura 6). Al final de tot el procés de la *Faster R-CNN*, acabem amb una llista d'objectes detectats en una imatge, cadascun amb la probabilitat de pertànyer a una classe d'objectes i amb una caixa delimitadora localitzant-la. En el nostre cas vam establir la xarxa neuronal per cercar només un objecte (un peix) i assignar-li només una classe ("*Xyrichtys novacula*") (Figura 7).

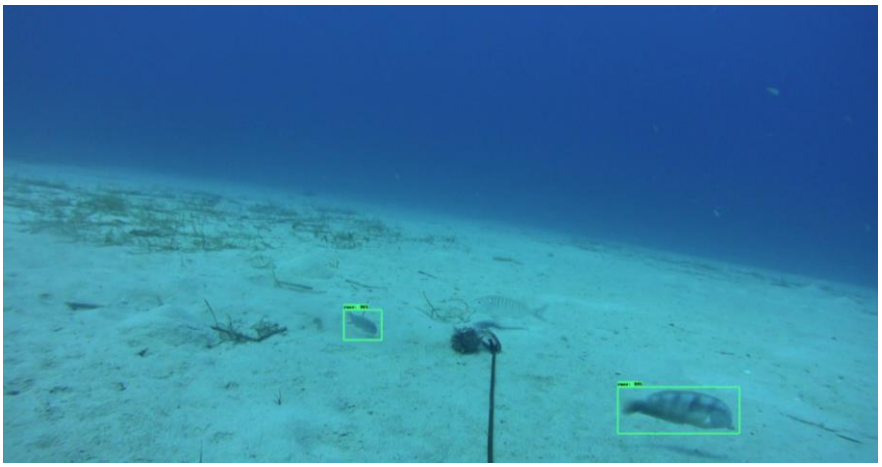

Figura 7 . Detecció i assignació de la classe de la xarxa neuronal *Faster R-CNN* entrenada per detectar el raor dintre de la Xarxa Natura 2000.

La *Faster R-CNN* està formada per tres xarxes diferents. La primera és la xarxa base CNN pre-entrenada, on l'objectiu d'aquesta xarxa és disminuir el nostre esforç, el temps de dedicació i la potència informàtica utilitzant dades calculades amb anterioritat (Hendrycks et al., 2019). Hem de considerar que a les nostres imatges només hi trobem objectes que ens interessen per detectar i classificar, en el nostre cas peixos, i que són completament diferents dels objectes utilitzats per a la formació prèvia de la xarxa, que no necessàriament són peixos. La segona i la tercera xarxa corresponen a RPN i la *Faster R-CNN*. Aquestes són les xarxes que realment hem format. La *Fast R-CNN* original es va formar pas a pas mitjançant l'aproximació, per això les dades es van fusionar abans d'un darrer pas d'entrenament complet (Ren et al., 2017). Ara podem saber que la formació conjunta condueix a millors resultats.

Per això, totes les imatges que utilitzarem per a la formació de la xarxa passaran primer per RPN i després per el *Faster R-CNN*. Vam establir la xarxa neuronal amb tres taxes d'aprenentatge diferents a tres moments de la formació. Una sense llindar d'augmentació i dues més amb un llindar d'augmentació diferent.

2.5 Homogeneïtzació d'imatges

En el context d'aquest Treball de Fi de Grau, la xarxa de DL pre-entrenada *Faster R-CNN* es va entrenar amb les imatges de raor generades al punt 2.1. Per entrenar les xarxes de DL es necessita un conjunt d'imatges extretes dels vídeos on es localitzin els diferents peixos objectiu. Aquestes imatges han de tenir una mateixa mida i un format de nomenclatura seqüencial. Es varen convertir totes les captures de pantalla dels vídeos a format JPG. La mida

idònia per a la selecció i entrenament de les xarxes és un ample de 1,280 píxels x 1,080 píxels (les figures 2 i 3 són exemples de les imatges generades).

2.6 Anotació i classificació d'imatges

Una vegada totes les imatges dels vídeos es varen homogeneïtzar es va procedir a l'anotació i la classificació dels peixos que ens interessen de les imatges (Figura 7). Amb una xarxa de DL podem usar el model de classificació, on ens informa automàticament de la presència d'un objecte a la imatge o la detecció d'objectes. En l'últim cas la xarxa classifica i localitza l'objecte. Aquest tipus de xarxes són les més interessants per nosaltres perquè volem saber quin peix apareix al vídeo i quin moviment realitza (Reithaug, 2018). La part d'etiquetatge del procés d'aprenentatge és un procés manual. S'ha de obrir cada imatge, etiquetar i anomenar l'objecte que ens interessa. Aquest procés es duu a terme amb la finalitat de què l'ordinador pugui "aprendre". Per arribar a aconseguir-ho es necessiten un gran volum d'imatges de l'objecte que ens interessa.

Per a l'etiquetatge vàrem utilitzar un programa gratuït *LabelImg* desenvolupat per *Python* (Figura 5). El llenguatge de *Phyton* és un llenguatge de programació lliure i molt versàtil. El programa obre cada imatge i permet realitzar anotacions de manera ràpida i en diferents formats. La xarxa de DL que vàrem utilitzar (*Faster R-CNN*) necessita les anotacions guardades en XML, per tant vàrem guardar las imatges en aquest format i amb la seva etiqueta dels peixos que hi apareixen, el seu nom, el nombre de la imatge i la posició dels peixos dins la imatge (Figura 5).

2.7 Nombre d'imatges necessari i procés d'augmentació

El nombre d'imatges utilitzades per a la formació de la xarxa és un paràmetre molt important en el DL (M. W. Mathis & Mathis, 2020) per tant, és un paràmetre a considerar en el desenvolupament de la nostra xarxa. Observant els canvis, en les prediccions, que sofreix la xarxa neuronal en relació al nombre d'imatges ens pot donar informació molt valuosa sobre el seu funcionament i de com la podem millorar (M. W. Mathis & Mathis, 2020). No obstant l'augment de dades d'imatges (procés anomenat augmentació) és un mètode que s'utilitza per ampliar artificialment la mida d'un conjunt de dades d'entrenament creant versions modificades de les imatges originals d'un conjunt de dades (Perez & Wang, 2017). Per aquest treball es va avaluar l'efecte de l'augmentació i es va establir una xarxa neuronal utilitzant tres

tipus de modificacions que ens poden ser útils per detectar un peix. Aquestes tres modificacions són: *Random Horizontal Flip*, per crear imatges mirall que parteixen de punts aleatoris al llarg d'una línia horitzontal; color de distorsió aleatòria, on a l'atzar se canvia els colors de la imatge i s'ajusta de manera aleatòria la lluminositat per modificar la brillantor de les imatges; i la darrera modificació és la imatge de retall aleatori, que consisteix en retallar una part d'una imatge amb els seus quadres de delimitació (*Tensorflow-Models/Preprocessor.Proto at Master · Rwightman/Tensorflow-Models · GitHub*, 2020). L'augment de dades és una tècnica molt habitual en anàlisi d'imatges amb DL, ja que a més de reduir el nostre esforç s'ha demostrat que augmenta substancialment la precisió del model (Wong et al., 2016).

2.8 Validació de la xarxa

Per avaluar la precisió de les prediccions de la nostra xarxa de DL en la tasca d'identificar correctament el raor en imatges submarines, hem realitzat una validació a cadascuna de les tres xarxes entrenades; una primera sense augmentació, i dues més amb augmentació (en una d'elles la probabilitat de detecció de l'objecte serà major a 0.5 i l'altre serà major a 0.7). El que ens permet el llindar és decidir a partir de quina probabilitat de detecció de la xarxa volem que ens doni la seva detecció com a certa (presència o absència del peix). D'acord amb aquest argument, la xarxa sense augmentació ens informa dels objectes que ha trobat amb una possibilitat de què fossin un raor major que un 95%, la segona xarxa en canvi té un valor de detecció del 50% i la tercera xarxa té un valor de detecció del 70%. La validació va consistir a analitzar els mateixos vídeos manualment i utilitzar l'algoritme i comparar els dos resultats. Per a la validació, es van seleccionar un total de vuit vídeos.

És important tenir present que les imatges utilitzades per a la validació ha de ser completament noves per a l'algoritme, no poden haver estat utilitzades per entrenar la xarxa. Per tant, els vídeos emprats en la validació no es van fer servir en el procés d'entrenament de la xarxa.

Per fer la validació, els vídeos seleccionats es van analitzar de manera manual. L'anàlisi manual es va fer veient els vídeos i prenent nota de la presència o absència del raor per segon. Es van generar una sèrie temporal on 0 significava l'absència del raor i en cas de la presència d'un o varis raors s'anotava el nombre de raors que apareixen alhora (Figura 6 i 7). La nostra xarxa va analitzar de manera automàtica els vídeos per comparar l'efectivitat del nostre

algoritme en la identificació del raor, generant un marc de probabilitat de presència d'un o varis raors a cada fotograma. La probabilitat de la presència d'un peix és d'un valor entre 0 i 1, i es va fixar un llindar a partir del qual es considera que apareix un peix. Per a cada validació es va utilitzar els llindars de 0.5 i 0.7, per seleccionar la que ens proporciona la predicció més exacta i entendre millor la distribució de probabilitats de la sortida.

El procés de validació general va generar dues sèries temporals per vídeo, una manual i un altre resultant de la xarxa de DL. Per cada una de les validacions i les vam comparar per calcular-ne el percentatge coincident (% d'èxit d'absència o presència de raor). Per a cada vídeo, en primer lloc vàrem comparar el temps, les sèries de temps es divideixen en dues: les parts on hi trobem peixos i les parts on no n'hi ha. Així podem visualitzar fàcilment si l'algoritme té més problemes per predir la presència d'un o varis peixos quan estan presents, o l'absència d'aquests quan està absent. És a dir, volem saber la quantitat de falsos positius i negatius que ens pot donar la nostra xarxa.

Nosaltres vam analitzar vuit vídeos d'una durada mitjana de 40 minuts, generant aproximadament 27,800 fotogrames analitzats amb l'absència / presència dels raors, i la quantitat en cas de què n'aparegui més d'un. Totes les transformacions de dades i comparacions es van realitzar amb el programa *Python* (Kiusalaas, 2013).

3 Resultats

Després de la validació de les tres xarxes entrenades obtenim una taula de probabilitats on hi apareix la probabilitat d'encert de la xarxa enfront de la validació manual en cada vídeo.

Vam analitzar un total de vuit vídeos, i a continuació comentarem els resultats obtinguts en cadascun d'ells:

Els resultats de la validació del vídeo 58A mostren, en general, un bon funcionament de la xarxa (Figura 8). En general trobem un individu al transcurs del vídeo menys a un moment proper a l'inici on podem observar dos mentre que les xarxes en detecten fins a tres al voltant del segon 1500 (Figura 8). La xarxa sense augmentació té un percentatge d'encert de 82.96% mentre que l'augmentació de 0.5 de probabilitat en té un de 35.54% i l'augmentació de probabilitat 0.7 té un 59.25% d'encert.

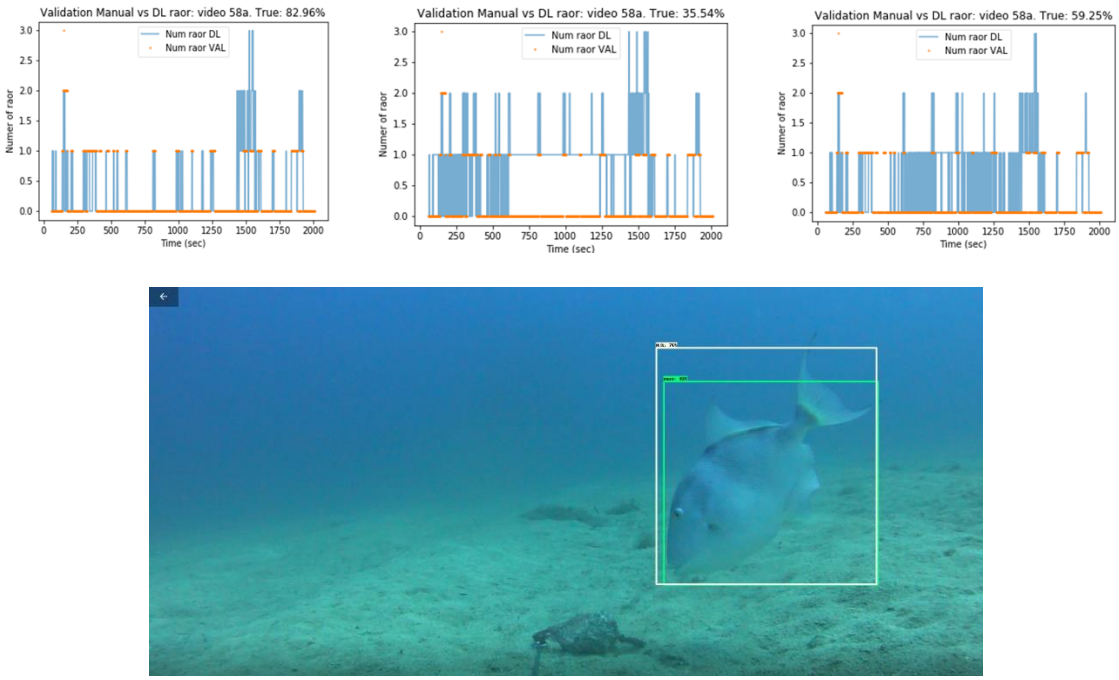

Figura 8. A la part superior, d'esquerra a dreta les gràfiques de la primera xarxa sense validació, de la segona xarxa amb una augmentació de probabilitat 0.5 i la tercera xarxa amb l'augmentació de probabilitat 0.7 del vídeo 58A. A l'eix de les X veiem el nombre de raors mentre que a l'eix de les Y el temps en segons transcorreguts durant la gravació. En color taronja la validació manual i en color blau la detecció d'un raor per part de les xarxes. A la part superior de les gràfiques podem veure el percentatge de coincidència entre la validació i la detecció del DL que ens diu el percentatge d'encert. A la part inferior veiem un fotograma del segon 1500 on veiem la causa de la detecció de 2 raors.

Els resultats de la validació del vídeo 58B també mostren un bon funcionament de la xarxa (Figura 9). Al transcurs del vídeo veiem en molts de moments l'absència de raor, a la meitat hi trobem l'aparició de fins a dos individus, la xarxa s'ajusta a la validació. La xarxa sense validació té un percentatge d'encert de 92.47% mentre que l'augmentació de probabilitat 0.5 en té un de 89.61% i l'augmentació de probabilitat 0.7 té un 89.61% d'encert.

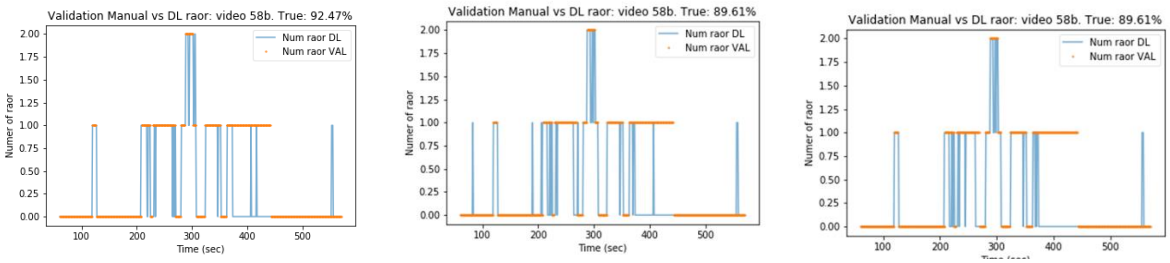

Figura 9. D'esquerra a dreta les gràfiques de la primera xarxa sense validació, de la segona xarxa amb una augmentació de probabilitat 0.5 i la tercera xarxa amb l'augmentació de probabilitat 0.7 del vídeo 58B. A l'eix de les X veiem el nombre de raors mentre que a l'eix de les Y el temps en segons transcorreguts durant la gravació. En color taronja la validació manual i en color blau la detecció d'un raor per part de les xarxes. A la

part superior de les gràfiques podem veure el percentatge de coincidència entre la validació i la detecció del DL que ens diu el percentatge d'encert.

Respecta al vídeo 69A, a l'inici del vídeo podem veure dos raors però aviat desapareixen de la imatge, transcorregut el vídeo un torna aparèixer però les xarxes no el detecten (Figura 10). La xarxa sense augmentació té un percentatge d'encert de 67.27% mentre que l'augmentació de probabilitat 0.5 en té un de 67.44% i l'augmentació de probabilitat 0.7 té un 67.36% d'encert.

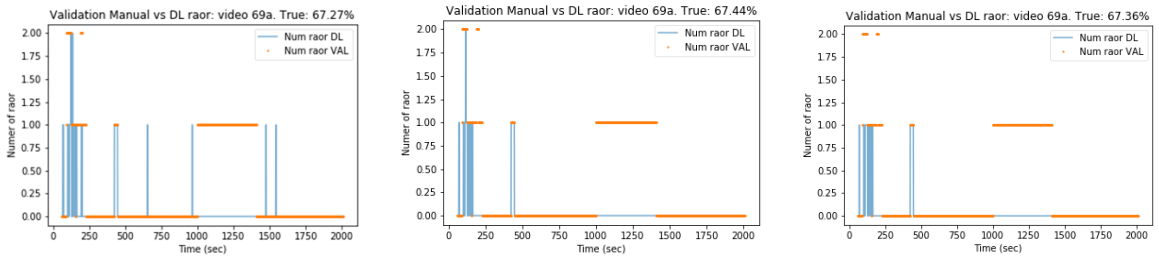

Figura 10. D'esquerra a dreta les gràfiques de la primera xarxa sense validació, de la segona xarxa amb una augmentació de probabilitat 0.5 i la tercera xarxa amb l'augmentació de probabilitat 0.7 del vídeo 69A. A l'eix de les X veiem el nombre de raors mentre que a l'eix de les Y el temps en segons transcorreguts durant la gravació. En color taronja la validació manual i en color blau la detecció d'un raor per part de les xarxes. A la part superior de les gràfiques podem veure el percentatge de coincidència entre la validació i la detecció del DL que ens diu el percentatge.

Els resultats de la validació del vídeo 106A mostren un funcionament de la xarxa menor que els altres (Figura 11). Aquest vídeo és un dels més complexos, hi apareixen fins a quatre raors amb moltes interaccions d'entrada i sortida de pantalla a més d'aparèixer altres espècies, les xarxes detecten fins a cinc individus quan la validació ens diu que n'hi ha quatre. La xarxa sense augmentació té un percentatge d'encert de 55.21% mentre que l'augmentació de probabilitat 0.5 en té un de 59.22% i l'augmentació de probabilitat 0.7 té un 59.29% d'encert.

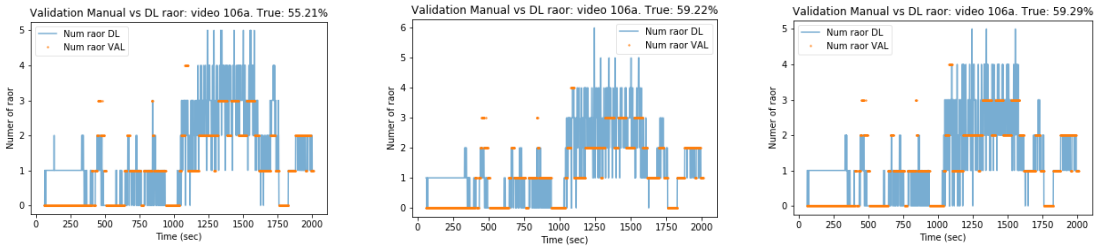

Figura 11. D'esquerra a dreta les gràfiques de la primera xarxa sense validació, de la segona xarxa amb una augmentació de probabilitat 0.5 i la tercera xarxa amb l'augmentació de probabilitat 0.7 del vídeo 106A. A l'eix de les X veiem el nombre de raors mentre que a l'eix de les Y el temps en segons transcorreguts durant la

gravació. En color taronja la validació manual i en color blau la detecció d'un raor per part de les xarxes. A la part superior de les gràfiques podem veure el percentatge de coincidència entre la validació i la detecció del DL que ens diu el percentatge d'encert.

Els resultats de la validació del vídeo 112A mostren un bon funcionament de la xarxa (Figura 12). Durant el transcurs del vídeo hi trobem un sol raor que apareix i desapareix juntament amb altres peixos, les xarxes detecten un màxim de tres individus. La xarxa sense validació té un percentatge d'encert de 85.70% mentre que l'augmentació de probabilitat 0.5 en té un de 71.01% i l'augmentació de probabilitat 0.7 té un 72.66% d'encert.

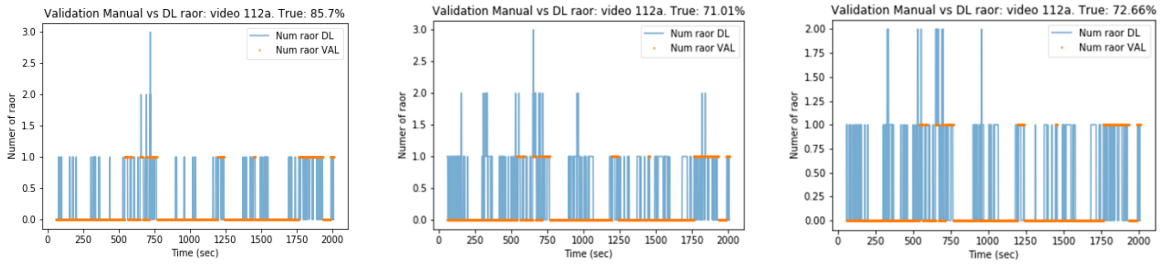

Figura 12. D'esquerra a dreta les gràfiques de la primera xarxa sense validació, de la segona xarxa amb una augmentació de probabilitat 0.5 i la tercera xarxa amb l'augmentació de probabilitat 0.7 del vídeo 112A. A l'eix de les X veiem el nombre de raors mentre que a l'eix de les Y el temps en segons transcorreguts durant la gravació. En color taronja la validació manual i en color blau la detecció d'un raor per part de les xarxes. A la part superior de les gràfiques podem veure el percentatge de coincidència entre la validació i la detecció del DL que ens diu el percentatge d'encert.

Els resultats de la validació del vídeo 112B mostren molt bon funcionament de la xarxa (Figura 13). Al vídeo trobem una aparició esporàdica d'un raor la qual les xarxes detecten. La xarxa sense augmentació té un percentatge d'encert de 96.70% mentre que l'augmentació de probabilitat 0.5 en té un de 97.4% i l'augmentació de probabilitat 0.7 té un 97.64% d'encert.

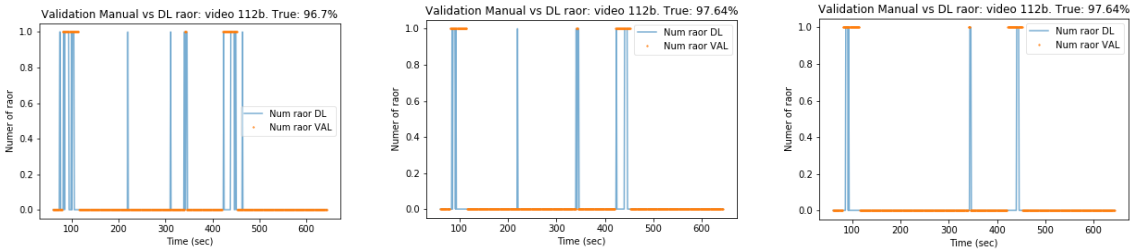

Figura 13. D'esquerra a dreta les gràfiques de la primera xarxa sense validació, de la segona xarxa amb una augmentació de probabilitat 0.5 i la tercera xarxa amb l'augmentació de 0.7 probabilitat del vídeo 112B. A l'eix de les X veiem el nombre de raors mentre que a l'eix de les Y el temps en segons transcorreguts durant la gravació. En color taronja la validació manual i en color blau la detecció d'un raor per part de les xarxes. A la

part superior de les gràfiques podem veure el percentatge de coincidència entre la validació i la detecció del DL que ens diu el percentatge d'encert.

Els resultats de la validació del vídeo 113A mostren, en general, un funcionament de la xarxa pitjor que la mitja (Figura 14). Vídeo amb molta presència d'individus de diverses espècies diferents, amb un màxim de quatre raors simultàniament en pantalla on les xarxes n'arriben a detectar cinc. La xarxa sense augmentació té un percentatge d'encert de 60.10% mentre que l'augmentació de probabilitat 0.5 en té un de 56.28% i l'augmentació de probabilitat 0.7 té un 52.31% d'encert.

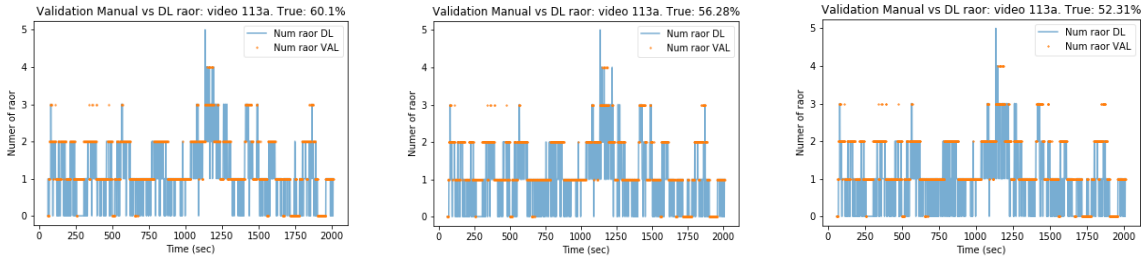

Figura 14. D'esquerra a dreta les gràfiques de la primera xarxa sense validació, de la segona xarxa amb una augmentació de probabilitat 0.5 i la tercera xarxa amb l'augmentació de probabilitat 0.7 del vídeo 113A. A l'eix de les X veiem el nombre de raors mentre que a l'eix de les Y el temps en segons transcorreguts durant la gravació. En color taronja la validació manual i en color blau la detecció d'un raor per part de les xarxes. A la part superior de les gràfiques podem veure el percentatge de coincidència entre la validació i la detecció del DL que ens diu el percentatge d'encert.

Finalment, els resultats de la validació del vídeo 113B mostra també un funcionament regular de la xarxa (Figura 15). En aquest vídeo, apareixen un màxim de dos raors i en general en trobem només un en pantalla, la xarxa també en detecta com a màxim dos. La xarxa sense augmentació té un percentatge d'encert de 57.21% mentre que l'augmentació de probabilitat 0.5 en té un de 54.71% i l'augmentació de probabilitat 0.7 té un 49.85% d'encert.

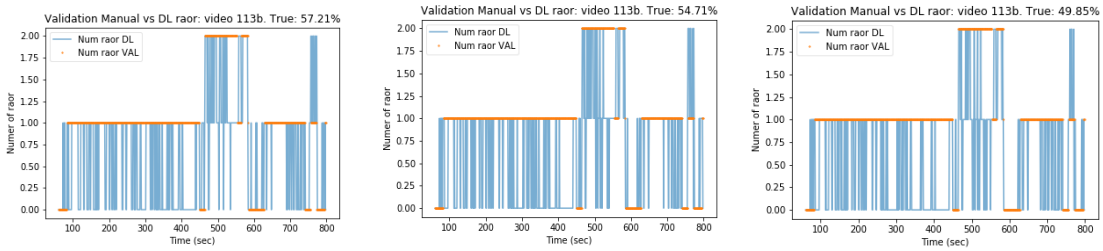

Figura 15. D'esquerra a dreta les gràfiques de la primera xarxa sense validació, de la segona xarxa amb una augmentació de probabilitat 0.5 i la tercera xarxa amb l'augmentació de probabilitat 0.7 del vídeo 113B. A l'eix de les X veiem el nombre de raors mentre que a l'eix de les Y el temps en segons transcorreguts durant la

gravació. En color taronja la validació manual i en color blau la detecció d'un raor per part de les xarxes. A la part superior de les gràfiques podem veure el percentatge de coincidència entre la validació i la detecció del DL que ens diu el percentatge d'encert.

Nom del Vídeo	Sense Augmentació	Primera Augmentació (0'5)	Segona augmentació (0'7)
58 A	82.96%	35.54%	59.25%
58B	92.47%	89.61%	89.61%
69A	67.27%	67.44%	67.36%
106A	55.21%	59.22%	59.29%
112A	85.70%	71.01%	72.66%
112B	96.70%	97.64%	97.64%
113A	60.10%	56.28%	52.31%
113B	57.21%	54.73%	49.85%

Taula 1. Taula general on es troben tots els percentatges (%) d'encert dels 9 vídeos analitzats en les tres validacions realitzades (xarxa sense augmentació, xarxa amb augmentació amb un llindar de probabilitat 0.5 i xarxa amb augmentació amb un llindar de probabilitat 0.7). Les files corresponen als diferents vídeos analitzats (nom assignat vídeo) i les columnes el tipus de xarxa aplicada. El percentatge (%) es refereix a l'encert amb la classificació del raor en el vídeo.

De manera general, els valors de la xarxa sense augmentació es troben en un interval entre 55.21% a 99.08% d'encert, amb una mitja de 74.62% i una variància de 0.0309. La primera augmentació (0.5) trobem un interval d'entre 35.54% a 97.64% obtenint una mitja de 66,43% i una variància de 0.0397. La segona augmentació (0.7) trobem un interval d'entre 49.85% i 97.64%, amb una mitja de 71.43% i una variància de 0.0339, d'encert de detecció del raor (Taula 1) (Figura 16).

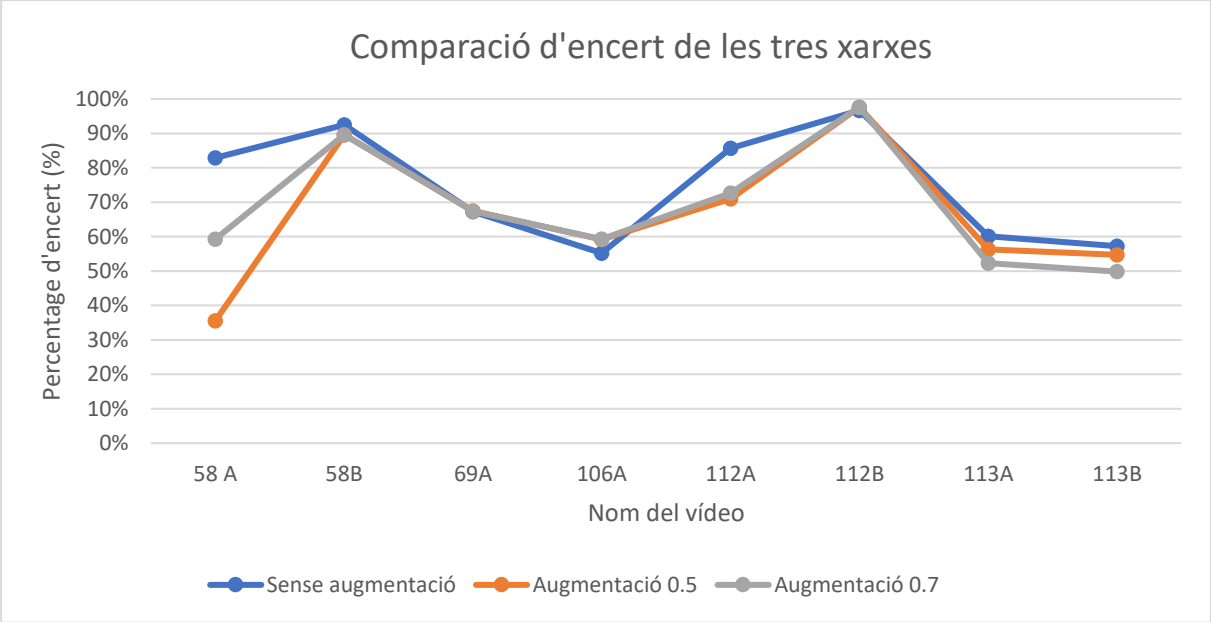

Figura 16. Comparació d'encert de les tres xarxes entrenades. A l'eix de les Y trobem el percentatge d'encert aconseguit per cada xarxa i a l'eix de les X les diferents xarxes (xarxa sense augmentació, xarxa amb augmentació amb un llindar 0.5 i xarxa amb augmentació amb un llindar 0.7) per a la detecció del raor. El 100% suposaria que la xarxa encerta tots els moments on apareix un raor en el vídeo.

4 Discussió

Les grans dificultats per obtenir dades de poblacions de peixos ens ha fet cercar noves alternatives per a poder obtenir dades d'abundància i diversitat d'espècies, com per exemple, l'aplicació del *Deep Learning* (DL) (Álvarez-Ellacuría et al., 2019). L'entrenament de les xarxes d'intel·ligència artificial de DL per realitzar tasques de seguiment de poblacions de peixos ens permet obtenir una gran quantitat de dades amb poc esforç (M. W. Mathis & Mathis, 2020). En el nostre cas, hem entrenat tres versions de la xarxa pre-entrenada *Faster R-CNN* per identificar raors en imatges submarines de manera automàtica. Els resultats de la validació mostren, en la millor de les versions de la xarxa, una taxa d'encert del 74.62%. Amb aquests resultats podem concloure que la nostra xarxa és eficient en la tasca de detecció automàtica de raors (Xu & Matzner, 2018), encara que requereix d'algun refinament. El nostre treball suggereix que l'aplicació del DL és una eina prometedora pel seguiment de les poblacions de peixos de les Illes Balears, contribuint a diferents aspectes de la Política Pesquera Comú (PPC) amb dades de espècies explotades per la pesca, o per avaluar l'efectivitat de la Xarxa Natura 2000. Aquest Treball suposa ser la primera passa per desenvolupar una nova metodologia automàtica i no supervisada del seguiment de la biodiversitat marina per afavorir-ne la conservació.

En aquest estudi hem emprat la xarxa *Faster R-CNN* per entrenar-la per a la detecció del raor. Aquesta és una xarxa entrenada per nosaltres amb 9208 imatges més. Aquesta xarxa no només ha estat utilitzada per identificar especies com peixos si no altres objectes, com el cas de l'estudi per la identificació de cotxes (Miralles, 2015). En aquets estudis la seva eficiència ha estat d'un 99.99%. En general podem concloure que la *Faster R-CNN* és una bona xarxa pre-entrenada per la detecció de peixos marins, ja que ha demostrat que amb el seu correcte entrenament pot donar uns valors d'encert molt alts.

L'efectivitat de la nostra xarxa ha estat de manera global de 74%. Altres estudis d'aplicació del DL en peixos en el fons marí han obtingut encerts del 54% (Xu & Matzner, 2018) i d'altres superiors de 81% (Li et al., 2016), 82% (Mandal et al., 2018) i 90% (Norouzzadeh et al., 2018; Villon et al., 2018), suggerint que aquesta eina pot contribuir notablement a la conservació de les espècies de peixos marins i fins i tot d'animals terrestres. Per a intentar millorar els percentatges d'encert de les xarxes s'ha fet una visualització de cada vídeo aplicant la xarxa

per veure on s'equivocava i perquè (Ching et al., 2018). A la xarxa sense augmentació hem obtingut un percentatge d'encert més gran, és un percentatge alt tenint en compte la dificultat i l'aplicació recent d'aquesta tecnologia a l'ecologia (Reithaug, 2018). Per tant, en el nostre cas estudi, el procés d'augmentació no ha contribuït a millorar l'eficiència de la xarxa. Altres estudis com Reithaug (2018) han demostrat que la augmentació millora l'eficiència de les xarxes de DL. En el nostre cas, la falta de millora per part de l'augmentació pot ser motivat per l'ampliació del rang de detecció d'objectes, cosa que fa que tingui més probabilitat d'error (Reithaug, 2018).

En general, els factors que expliquen els nostres percentatges solen ser artefactes dins els fons natural on s'enregistren les imatges, detecció del la xarxa de l'esquer i confusió amb altres espècies que conviuen amb el raor. Els vídeos 58B i 112B són vídeos amb un encert major a 90% amb la xarxa sense augmentació. Són vídeos amb poca presència de raors o directament sense presència d'altres espècies i amb fons molt homogenis, fet que provoca un error mínim a la xarxa. Els vídeos 58A i 112A són vídeos amb un encert major a un 80% amb la xarxa sense augmentació. Són vídeos on la xarxa comet erros amb confusió amb altres espècies, al 58A amb *Ballistes caronilensis* i el 112A amb *Bothus podas* a més de petits intervals de detecció de la xarxa de l'esquer. Aquestes xarxes amb les augmentacions augmenten aquests problemes i fa que empitjori el percentatge d'encert (Villon et al., 2018).

Un altre problema de la nostra xarxa poden ser els presents als vídeos 69A i 113A. Són vídeos amb un encert per damunt del 60% amb la xarxa sense augmentació. Els errors com són en aquest cas és a causa de la distància del peix a la càmera, la xarxa està entrenada per detectar els raors pròxims a la càmera i quan estan més lluny ja no es detecta, cosa que amb la validació manual si que podem detectar. Per tant, podem pensar que és un error amb la validació manual, ja que es van comptabilitzar raors que no entraven dins el rang de detecció de la xarxa i que per tant no podia detectar. Els vídeos 106B i 113B són vídeos d'entre un 50 i un 60% d'encert amb la xarxa sense augmentació. Són vídeos on la xarxa comet tots els errors anteriors de manera esporàdica però que en conjunt creen un gran error.

En general, la taxa d'encert baixa per la detecció errònia de l'esquer, la confusió amb altres especies (ja que són vídeos amb una gran quantitat d'espècies de peixos diferents) i amb la distància de validació manual.

A priori, la tècnica de l'augmentació de les xarxes per així crear més variabilitat el que fa és incrementar els percentatges d'errors anteriorment comentats o en tot cas mantenir els que tenia la primera xarxa sense augmentació. Tots aquests errors es podrien corregir entrenat la xarxa per a la detecció de les espècies més habituals en l'ecosistema del raor amb el mateix procediment que nosaltres hem dut a terme (Villon et al., 2018).A més, podem entrenar la xarxa per a la detecció de l'esquer per així poder discriminar-la i utilitzar un millor patró de validació amb un rang de distància conegut per així no comptabilitzar els peixos que estan massa lluny per a la seva detecció (Álvarez-Ellacuría et al., 2019). Per tant per millor el funcionament de la xarxa en un futur és pot millorar entrenant-la per a la detecció i classificació de les imatges convivents amb el raor així com de la xarxa de l'esquer amb la mateixa metodologia que s'ha fet en aquest Treball de Fi de Grau.

Deixant de banda les limitacions de la nostra xarxa, la versió que funciona millor és la xarxa sense augmentació que produeix un 74.7% d'encert de mitja. Això la converteix en una eina prometedora pel monitoreig automàtic de les poblacions de raors de les Illes Balears. Aquesta espècie és molt valorada per la pesca local i les poblacions es veuen afectades per la gran pressió que sofreix aquesta espècie (Alós et al., 2016). En l'actualitat no existeix una metodologia vàlida per monitoritzar aquesta espècie, i aquest treball posa la primera passa per posar en funcionament un programa de seguiment de les poblacions d'aquesta espècie. Això fa que aquest treball no només sigui un avanç important, sinó que també suposi tenir un component aplicat molt gran en quant a la gestió de la pesca. La gran quantitat de dades obtingudes amb aquesta metodologia sobre les poblacions del raor poden ser de gran ajuda per assolir objectius de la PPC per a aconseguir una activitat pesquera socialment i mediambientalment sostenible. També, l'objectiu de la Xarxa Natura 2000 de garantir el manteniment i potenciació de la biodiversitat i els serveis ecosistèmics es pot veure beneficiat pel nostre treball. Per tant, aquest treball posa de manifest la utilitat de la intel·ligència artificial i la conservació de la biodiversitat marina.

5 Conclusió

- La *Faster R-CNN* entrenada amb el raor és eficient per a la seva detecció, classificació i comptabilització de raors.
- L'eficàcia de la xarxa més eficient és la xarxa sense augmentació i detecta el raor amb una mitja de 74.7% d'encert.
- Els errors que comet la xarxa són provocats per la interferència d'altres espècies de peixos, de la xarxa de l'esquer o d'artefactes presents al fons marí.
- La xarxa funciona millor quan els fons on s'enregistren són homogenis.
- Per millorar la xarxa seria adient seguir entrenant-la amb un conjunt major d'imatges del raor i entrenar-la amb espècies convivents per evitar que es confongui.
- Aquesta xarxa suposa una nova eina cost-efectiva pel seguiment de les poblacions de raor en estat salvatge i amb la possibilitat d'entrenar-la per a la detecció d'altres espècies.

6 Agraïments

Aquest treball va ser finançat pel Projecte R&D CLOCKS (subvenció núm. PID2019-104940GA-I00) finançat pel Ministeri espanyol de ciència i innovació (MICINN) i l'Agència Estatal d'Investigació i el projecte de recerca intramural JSATS (subvenció núm. PIE202030E002) finançat pel MICINN i el Consell Nacional de Recerca Espanyol (CSIC).

També agrair per la col·laboració a tot el Grup d'ecologia de peixos de l'IMEDEA, a la doctora Arancha Lana i en concret al meu tutor Josep Alós ; també a Guillem Mateu per l'enllaç entre la UIB i l'IMEDEA.

7 Bibliografía

Abelson, A., & Shlesinger Abelson, Y. (2002). Comparison of the development of coral and fish communities on rock-aggregated artificial reefs in Eilat, Red Sea. *ICES Journal of Marine Science*, *59*, 122–126. https://doi.org/10.1006/jmsc.2002.1210

Alós, J., Palmer, M., Rosselló, R., Reports, R. A.-S., & 2016, U. (2016). Fast and behavior-selective exploitation of a marine fish targeted by anglers. *Nature.Com*. https://www.nature.com/articles/srep38093

Álvarez-Ellacuría, A., Palmer, M., Catalán, I. A., & Lisani, J.-L. (2019). Image-based,

unsupervised estimation of fish size from commercial landings using deep learning. *ICES Journal of Marine Science*. https://doi.org/10.1093/icesjms/fsz216

Arlinghaus, R., Abbott, J. K., Fenichel, E. P., Carpenter, S. R., Hunt, L. M., Alós, J., Klefoth, T., Cooke, S. J., Hilborn, R., Jensen, O. P., Wilberg, M. J., Post, J. R., & Manfredo, M. J. (2019). Governing the recreational dimension of global fisheries. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 116, Issue 12, pp. 5209–5213). National Academy of Sciences. https://doi.org/10.1073/pnas.1902796116

Balaguer, P., Diedrich, A., Sardá, R., Fuster, M., Cañellas, B., & Tintoré, J. (2011). Spatial analysis of recreational boating as a first key step for marine spatial planning in Mallorca (Balearic Islands, Spain). *Ocean and Coastal Management*, *54*(3), 241–249. https://doi.org/10.1016/j.ocecoaman.2010.12.002

Blicharska, M., Orlikowska, E. H., Roberge, J., & Grodzinska-jurczak, M. (2016). Contribution of social science to large scale biodiversity conservation : A review of research about the Natura 2000 network. *BIOC*, *199*, 110–122. https://doi.org/10.1016/j.biocon.2016.05.007

Brock, R. E. (1982). A CRITIQUE OF THE VISUAL CENSUS METHOD FOR ASSESSING CORAL REEF FISH POPULATIONS. In *BULLETIN OP MARINE SCIENCE* (Vol. 32, Issue 1).

Cabanellas-Reboredo, M., Palmer, M., … J. A.-F., & 2017, U. (2016). Estimating harvest and its uncertainty in heterogeneous recreational fisheries. *Elsevier*. https://www.sciencedirect.com/science/article/pii/S0165783616304180

Ching, T., Do, B. T., Himmelstein, D. S., Beaulieu-jones, B. K., Kalinin, A. A., Way, G. P., Ferrero, E., Agapow, P., Carpenter, A. E., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lanchantin, J., Xu, J., Woloszynek, S., Lu, Z., Harris, D. J., Peng, Y., & Wiley, L. K. (2018). *Opportunities and obstacles for deep learning in biology and medicine Authors*.

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, *10*(10), 1632–1644. https://doi.org/10.1111/2041-210X.13256

Demartini, E. E. (n.d.). *Evaluation of a video camera technique for indexing abundances of juvenile pink snapper, Pristipomoides filamentosus, and other Hawaiian insular shelf fishes*. Retrieved June 10, 2020, from

https://www.researchgate.net/publication/279568979

Edgar, G. J., Barrett, N. S., & Morton, A. J. (2004). Biases associated with the use of underwater visual census techniques to quantify the density and size-structure of fish populations. *Journal of Experimental Marine Biology and Ecology*, *308*(2), 269–290. https://doi.org/10.1016/j.jembe.2004.03.004

EUR-Lex - 32009R1224 - EN - EUR-Lex. (2009). https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32009R1224

Evans, D. (2012). *Building the European Union ' s Natura 2000 network*. *26*, 11–26. https://doi.org/10.3897/natureconservation.1.1808

Farías Torbidoni, E. I. (2011). Managing for recreational experience opportunities: The case of hikers in protected areas in Catalonia, Spain. *Environmental Management*, *47*(3), 482–496. https://doi.org/10.1007/s00267-010-9606-z

Font, T., & Lloret, J. (2014). Biological and ecological impacts derived from recreational fishing in Mediterranean coastal areas. In *Reviews in Fisheries Science and Aquaculture* (Vol. 22, Issue 1, pp. 73–85). Taylor and Francis Inc. https://doi.org/10.1080/10641262.2013.823907

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. https://books.google.es/books?hl=es&lr=&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=Goodfellow+I,+Bengio+Y,+Courville+A.+Deep+Learning.+2016:785&ots=MMV-imsERT&sig=RAY8kZe_yEO-eIIjLSBLqEi8R6g

Heagney, E., Lynch, T., Babcock, R., & Suthers, I. (2007). Pelagic fish assemblages assessed using mid-water baited video: standardising fish counts using bait plume size. *Marine Ecology Progress Series*, *350*, 255–266. https://doi.org/10.3354/meps07193

Hendrycks, D., Lee, K., & Mazeika, M. (2019). *Using Pre-Training Can Improve Model Robustness and Uncertainty*. http://arxiv.org/abs/1901.09960

Holden, M., & Garrod, D. (1996). The Common Fisheries Policy: origin, evaluation and future. *The Common Fisheries Policy: Origin, Evaluation and Future.*, *Ed. 2*.

Humphries, G., Magness, D., & Huettmann, F. (2018). *Machine learning for ecology and sustainable natural resource management*.

https://link.springer.com/content/pdf/10.1007/978-3-319-96978-7.pdf

Kiusalaas, J. (2013). *Numerical methods in engineering with Python 3*. https://books.google.com/books?hl=en&lr=&id=YlkgAwAAQBAJ&oi=fnd&pg=PP10&dq=python+3&ots=-45s52Kw-h&sig=xEd5BUbu1oOCsGcvuUZ8POqgSqQ

Lafond, E. C. (1961). *The Isotherm Follower 1*.

Les reserves marines a les Illes Balears-Les reserves marines a les Illes Balears. (2019). https://www.caib.es/sites/reservesmarines/ca/les_reserves_marines_a_les_illes_balears-850/

Li, X., Shang, M., Qin, H., & Chen, L. (2016). *Fast Accurate Fish Detection and Recognition of Underwater Images with Fast R-CNN*.

Mandal, R., Connolly, R. M., Schlacher, T. A., & Stantic, B. (2018). *Assessing fish abundance from underwater video using deep neural networks*.

Marini, S., Fanelli, E., Sbragaglia, V., Azzurro, E., Del Rio Fernandez, J., & Aguzzi, J. (2018). Tracking Fish Abundance by Underwater Image Recognition. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-018-32089-8

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Mathis, M. W., & Mathis, A. (2020). Deep learning tools for the measurement of animal behavior in neuroscience. *Current Opinion in Neurobiology*, *60*, 1–11. https://doi.org/10.1016/j.conb.2019.10.008

Miralles, L. (2015). *Reconocimiento de dígitos escritos a mano mediante métodos de tratamiento de imagen y modelos de clasificación Machine learning View project Human activity recognition View project*. https://www.researchgate.net/publication/278157473

Morales-Nin, B., Moranta, J., García, C., Tugores, M. P., Grau, A. M., Riera, F., & Cerdà, M. (2005). The recreational fishery off Majorca Island (western Mediterranean): some implications for coastal resource management. *ICES Journal of Marine Science*, *62*(4), 727–739. https://doi.org/10.1016/j.icesjms.2005.01.022

Naga, J. B.-, Quarterly, T. I., & 1994, undefined. (n.d.). Marine reserves: they enhance fisheries, reduce conflicts, and protect resources. *Aquaticcommons.Org*. Retrieved June 10, 2020, from http://aquaticcommons.org/9095/1/na_2198.pdf

Nature, H. B.-, & 1952, undefined. (n.d.). Under-water television and marine biology. *Springer*. Retrieved June 10, 2020, from https://link.springer.com/content/pdf/10.1038/169477a0.pdf

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Owens, G. L., & Rennison, D. J. (2017). Evolutionary ecology of opsin gene sequence, expression and repertoire. *Molecular Ecology*, *26*(5), 1207–1210. https://doi.org/10.1111/mec.14032

Perez, L., & Wang, J. (2017). *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*. http://arxiv.org/abs/1712.04621

Post, J. R., Sullivan, M., Cox, S., Lester, N. P., Walters, C. J., Parkinson, E. A., Paul, A. J., Jackson, L., & Shuter, B. J. (2011). Canada's Recreational Fisheries: The Invisible Collapse? *Changed Publisher: Wiley*. https://doi.org/10.1577/1548-8446(2002)027<0006:CRF>2.0.CO;2

Redmon, J., & Farhadi, A. (2016). *YOLO9000: Better, Faster, Stronger*. http://pjreddie.com/yolo9000/

Reithaug, A. (2018). Employing Deep Learning for Fish Recognition. *Western Norway University of Applied Science*, *June*.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Ryan, K. L., Trinnie, F. I., Jones, R., Hart, A. M., & Wise, B. S. (2016). Recreational fisheries data requirements for monitoring catch shares. *Fisheries Management and Ecology*, *23*(3–4), 218–233. https://doi.org/10.1111/fme.12151

Siddiqui, S. A., Salman, A., Malik, M. I., Shafait, F., Mian, A., Shortis, M. R., & Harvey, E. S. (2018). Automatic fish species classification in underwater videos: exploiting pre-trained deep neural network models to compensate for limited labelled data. *ICES Journal of Marine Science*, *75*(1), 374–389. https://doi.org/10.1093/icesjms/fsx109

tensorflow-models/preprocessor.proto at master · rwightman/tensorflow-models · GitHub. (2020). https://github.com/rwightman/tensorflow-models/blob/master/research/object_detection/protos/preprocessor.proto

Tsiafouli, M. A., Apostolopoulou, E., Mazaris, A. D., Kallimanis, A. S., Drakou, E. G., & Pantis, J. D. (2013). Human activities in Natura 2000 sites: A highly diversified conservation network. *Environmental Management*, *51*(5), 1025–1033. https://doi.org/10.1007/s00267-013-0036-6

Villon, S., Mouillot, D., Chaumont, M., Darling, E. S., Subsol, G., Claverie, T., & Villéger, S. (2018). A Deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological Informatics*, *48*(August), 238–244. https://doi.org/10.1016/j.ecoinf.2018.09.007

VLC - Features - VideoLAN. (2020). https://www.videolan.org/vlc/features.html

Willis, T. J., & Babcock, R. C. (2000). A baited underwater video system for the determination of relative density of carnivorous reef fish. *Marine and Freshwater Research*, *51*(8), 755–763. https://doi.org/10.1071/MF00010

Wong, S. C., Gatt, A., Stamatescu, V., & Mcdonnell, M. D. (2016). Understanding data augmentation for classification: when to warp? In *ieeexplore.ieee.org*. https://ieeexplore.ieee.org/abstract/document/7797091/

Xu, W., & Matzner, S. (2018). *Underwater Fish Detection using Deep Learning for Water Power Applications*. https://github.com/wenweixu/keras-yolo3.

Zeller, D., Darcy, M., Booth, S., Lowe, M. K., & Martell, S. (2008). What about recreational catch?. Potential impact on stock assessment for Hawaii's bottomfish fisheries. *Fisheries Research*, *91*(1), 88–97. https://doi.org/10.1016/j.fishres.2007.11.010