ISBN: 978-950-42-0200-4
DOI: https://doi.org/10.33414/ajea.5.719.2020

Avances en foto-identificación automatizada de fauna silvestre

Advances in automated wildlife photo-identification

Presentación: 06/10/2020

Doctorando:
Débora Pollicelli
Centro para el Estudio de Sistemas Marinos (CESIMAR), Centro Nacional Patagónico, Centro Científico Tecnológico del Consejo Nacional de Investigaciones Científicas y Técnicas (CCT CENPAT - CONICET) - Argentina
deborapollicelli@gmail.com

Director/a:
Claudio Delrieux

Co-director/a:
Mariano Coscarella

Resumen

La foto-identificación de especies silvestres es un recurso base para la obtención de información necesaria para diversas tareas de investigaciones biológicas. Hoy en día el crowdsourcing y la ciencia ciudadana están comenzando a desempeñar un rol importante en la recopilación de datos científicos. Esta fuente de datos permite aumentar considerablemente el número de registros en la base de muestreo de diferentes proyectos científicos, especialmente los relacionados con los modelos de captura-recaptura fotográfica de vida silvestre. No obstante, mientras que se aumenta la cantidad de datos recopilados de fuentes no científicas, se presenta un nuevo desafío, el procesamiento masivo de manera ágil y eficiente, que permita limpiar y seleccionar los datos relevantes para las siguientes etapas.

Este trabajo aborda la automatización de la primer etapa del proceso de foto-identificación de cetáceos, el cual se trata de la detección de la presencia o ausencia de la región de interés en la imagen (ROI). Para ello, se especializó una red neuronal convolucional de propósito general (Mask R-CNN) con imágenes de delfines de la especie *Cephalorhynchus commersonii* recolectadas en diferentes sitios de la costa patagónica durante un período de siete años.

Palabras clave: foto-identificación, red neuronal convolucional (CNN), detección de ROI, ciencia ciudadana

Abstract

The wild species photo-identification is a basic resource for obtaining the necessary information for several biological research tasks. Today crowdsourcing and citizen science are beginning to play an important role in collecting scientific data. This data source makes it possible to considerably increase the number of records in the sampling database for different scientific projects, especially those related to photographic capture-recapture models of wildlife. However, while the amount of data

collected from non-scientific sources increases, a new challenge is presented, mass processing in an agile and efficient way, which allows cleaning and selecting the relevant data for the next stages.

This work addresses the automation of the first stage of the cetacean photo-identification process, which is the detection of the presence or absence of the region of interest in the image (ROI). For this aim, a general-purpose convolutional neural network (Mask R-CNN) was specialized with dolphins images of *Cephalorhynchus commersonii* specie, collected at different sites on the Patagonian coast over a period of seven years.

Keywords: photo-identification, convolutional neural network (CNN), ROI detection, citizen science

Introducción

La identificación de individuos de especies salvajes resulta una tarea de base en estudios biológicos y de ecología de poblaciones. Una de las formas actuales para adquirir tal información es mediante el modelo de captura-recaptura fotográfica realizado a lo largo del tiempo, en amplias regiones geográficas. Al agrupar todas las imágenes que contienen el mismo individuo, es posible realizar estudios poblacionales tales como estimaciones de abundancia, de tasas de mortalidad, ciclo reproductivo, migraciones, etc.. En particular, en el caso de fauna marina, hasta al momento existen muy pocos software disponibles para asistir dicha tarea. Además, sólo son aplicables a algunas especies de cetáceos y tienen el inconveniente de requerir laboriosas tareas manuales de pre-procesamiento y verificación de emparejamiento asistido por expertos. Tal labor demanda muchísimo tiempo (horas-hombre) y exige una carga cognitiva significativa que puede ser propensa a errores operativos intra e inter-observador.

Una desventaja de los modelos de captura-recaptura es que se requiere una cantidad estadísticamente significativa de la presencia de los animales en el área de estudio (Klaich, Kinas, Pedraza, Coscarella, & Crespo, 2011). Tal requisito, implica campañas de monitoreo muy frecuentes y capacidad de identificar a un individuo dado en una vasta serie de imágenes.

Hoy en día, gracias a la ciencia ciudadana y el crowdsourcing, es posible aumentar el número de registros en las bases de datos de muestreo de diferentes proyectos científicos, especialmente los relacionados con los modelos de captura-recaptura fotográfica de vida silvestre (Berger-Wolf et al., 2017; Kosmala, Wiggins, Swanson, & Simmons, 2016; Parham, Crall, Stewart, Berger-Wolf, & Rubenstein, 2017). Esto se convierte en una excelente oportunidad para superar una de las limitaciones planteadas, pero al aumentar considerablemente la cantidad de datos recopilados de fuentes no científicas, se presenta un nuevo desafío, el procesamiento masivo de imágenes de manera ágil y eficiente. Asimismo, las fotografías que se toman en ambientes no controlados (al aire libre en la naturaleza) poseen diferencias significativas en cuanto a la calidad (ya sea por el enfoque, la iluminación, la oclusión y /o las variaciones en el punto de vista y la postura que exhiben los animales), y adicionalmente, suelen contener objetos que no son afines al caso de estudio.

Afortunadamente, las técnicas de procesamiento inteligente de imágenes, han experimentado en la última década un avance muy importante aplicado a este tipo de problemáticas, en particular , el desarrollo de redes neuronales convolucionales (CNN). Mask R-CNN (He, Gkioxari, Dollar, & Girshick, 2017) es una CNN desarrollada recientemente que se hizo rápidamente muy conocida gracias a su versatilidad para ser aplicada en la segmentación de instancias de índole general. Además Mask R-CNN esta disponible en GitHub con los pesos pre-entrenados con enormes dataset de imágenes. Esto, sirve como base para realizar un nuevo entrenamiento utilizando transferencia de conocimiento y con algunas modificaciones adaptarla a nuevas tareas específicas. Tal adaptación permite detectar la presencia / ausencia de la ROI y posteriormente segmentar la figura, o bien, descartar las imágenes con ausencia de ROI.

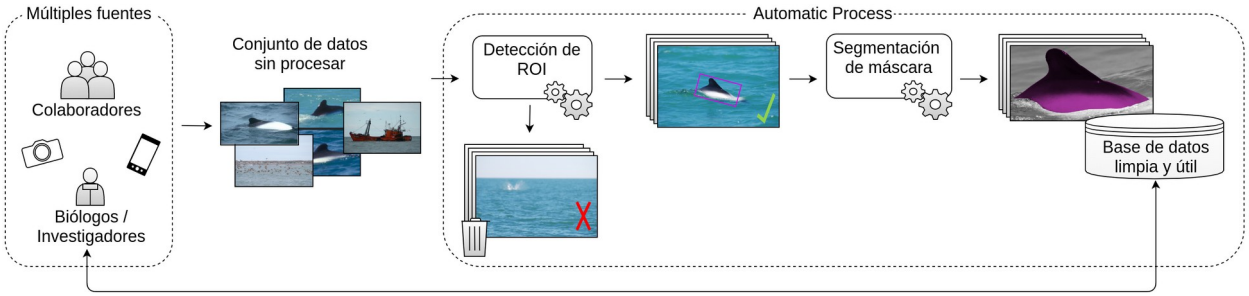

Figura 1: Representación de las primeras etapas en el proceso de foto-identificación

Desarrollo

El modelo de red Mask R-CNN pre-entrenado con el dataset MS COCO (Lin et al., 2014) fue utilizado en un re-entrenamiento con transferencia de conocimiento, alimentado con las anotaciones manuales de las máscaras (ground truth masks) del conjunto de fotografías de un tipo de delfín conocido como tonina overa (*Cephalorhynchus commersonii*). A pesar de que COCO no contiene imágenes de la clase delfín, posee imágenes de otras clases diferentes de animales y objetos que se encuentran en ambientes no controlados (~120K), lo cual hace que la red se adapte bastante bien para reconocer la mayoría de las características comunes al dominio del problema.

Las anotaciones manuales de las máscaras se realizaron con la herramienta Sloth en archivos de formato .json. Esta tarea requirió un tiempo y esfuerzo considerable, por lo cual se realizó en dos etapas. Consecuentemente hubieron dos entrenamientos. El primero (*I*), con un conjunto total de 150 imágenes y el segundo (*II*) con 499. En ambos casos las imágenes utilizadas, sólo tenían presencia de dos clases: background y delfín. En cada caso, el conjunto de imágenes fue dividido en dos: 80% para el entrenamiento y 20% para validación. Para obtener resultados comparativos entre ambos entrenamientos, el conjunto de validación del entrenamiento *I* constituyó un subconjunto del dataset de validación del entrenamiento *II*. Por otro lado, se realizaron pruebas de validación con otro conjunto de imágenes de campañas fotográficas (dataset de campañas), con múltiples clases de objetos que no son de interés (embarcaciones, aves, personas, etc.) entre las cuales puede o no haber presencia de delfines y además algunas imágenes que no tienen ninguna clase de objeto.

El modelo Mask R-CNN utilizado fue el basado en ResNet101 (He, Zhang, Ren, & Sun, 2016) como backbone para la extracción de características. La última capa (conocida como head) de predicción de clase, se modificó para reconocer solo una clase (delfín: tonina overa) en lugar de las 80 clases de MS COCO. La tasa de aprendizaje (learning rate) se estableció en 0.001. Las relaciones de aspecto de anclaje y las escalas para la RPN se mantuvieron en [0.5, 1, 2] y [32, 64, 128, 256, 512] respectivamente. El tamaño de la imagen con relleno cero (zero padding) se estableció en 1024x1024px. En el entrenamiento I se realizaron 10 epochs, mientras que en el *II*, fueron 100. En la Figura 1 puede observarse el diagrama de bloques simplificado de la arquitectura de la red con las adaptaciones realizadas. Para poder realizar una comparación, los resultados de ambos entrenamientos se validaron con el conjunto de validación *I* (30 imágenes con un total de 32 instancias de delfines) y se calcularon las matrices de confusión de las detecciones, junto con otras métricas como tasa de verdaderos positivos, precisión, exactitud y medida F1.

Adicionalmente, sobre la segmentación obtenida, se desarrolló un proceso posterior para optimizar el recorte de ROI ajustado a la máscara. El mismo fue realizado calculando el centro de masa de la figura mediante análisis de componentes principales (PCA) con la posterior rotación de la máscara sobre los nuevos ejes definidos.

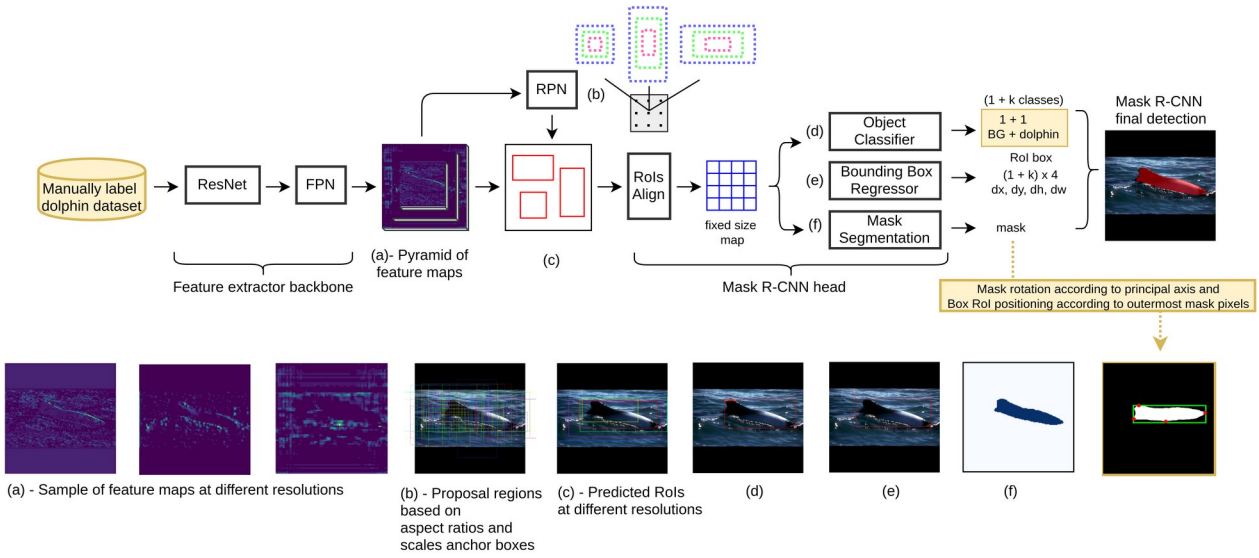

Figura 2: Arquitectura Mask R-CNN personalizada

Conclusiones

Entrenamiento *I*	Presencia	Ausencia
Detección Positiva	31	1
Detección Negativa	1	0

Entrenamiento *II*	Presencia	Ausencia
Detección Positiva	31	0
Detección Negativa	1	0

	E *I*	E *II*
Sensitivity: TPR = TP / (TP + FN)	0.9688	0.9689
Precision: PPV = TP / (TP + FP)	0.9688	1.0000
Accuracy: ACC = (TP + TN) / (P + N)	0.9394	0.9688
F1 Score: F1 = 2TP / (2TP + FP + FN)	0.9688	0.9841

Tabla 1: Métricas sobre el conjunto de validación

Figura 3: Ejemplos de máscaras de detecciones positivas de ROI en los conjuntos de validación.

Si bien la tasa de verdaderos positivos es igual en ambas pruebas, la variación de casos de falsos positivos y negativos, arroja una mejoría en las métricas calculadas del segundo entrenamiento, en relación al primero (observar Tabla 1). Respecto de la segmentación de máscaras, en ambos casos, no se logra un ajuste suficientemente preciso y exacto, es decir no se ajustan perfectamente a los bordes de la figura del delfín (ver Figura 3). No obstante, en los resultados del segundo entrenamiento pueden verse algunas mejoras sobre la definición de las máscaras. Respecto de la validación de resultados sobre el dataset de campañas, se obtuvo una mejora considerable del segundo entrenamiento respecto del primero, dado que son descartados una gran cantidad de objetos considerados fuera de interés (aves, personas, delfines con oclusiones, etc.). Todas las comparativas, demuestran que la red esta siendo más precisa a la hora de clasificar las figuras detectadas. Estos alentadores resultados marcan un camino a seguir, indicando que podría mejorarse la precisión y exactitud en la detección, clasificación y segmentación, aumentando las anotaciones manuales (ground truth) y realizando los ajustes de configuración pertinentes sobre la red pre-entrenada.

Aún a pesar de las limitaciones observadas, este trabajo proporcionó un prominente avance en la automatización del procesamiento de imágenes sobre un área multidisciplinar prácticamente vacante, entre informática y biología. En particular, se avanzó en la fase de detección de ROI, del proceso de foto-identificación sobre una especie endémica, cuya distribución está limitada a un ámbito geográfico reducido y no se encuentra de forma natural en ninguna otra parte del mundo. De ahí su importancia en los estudios de investigación y conservación.

Referencias

Berger-Wolf, T. Y., Rubenstein, D. I., Stewart, C. V, Holmberg, J. A., Parham, J., Menon, S., … Joppa, L. (2017). Wildbook: Crowdsourcing, computer vision, and data science for conservation. *CoRR*, *abs/1710.0*. Recuperado de http://arxiv.org/abs/1710.08880

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, *2017-Octob*, 2980-2988. https://doi.org/10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770-778. https://doi.org/10.1109/CVPR.2016.90

Klaich, M. J., Kinas, P. G., Pedraza, S. N., Coscarella, M. A., & Crespo, E. A. (2011). Estimating dyad association probability under imperfect and heterogeneous detection. *Ecological Modelling*, *222*(15), 2642-2650. https://doi.org/10.1016/j.ecolmodel.2011.03.027

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *14*(10), 551-560. https://doi.org/10.1002/fee.1436

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8693 LNCS*(PART 5), 740-755. https://doi.org/10.1007/978-3-319-10602-1_48

Parham, J., Crall, J., Stewart, C., Berger-Wolf, T., & Rubenstein, D. (2017). Animal Population Censusing at Scale with Citizen Science and Photographic Identification. *Association for the Advancement of Artificial Intelligence*, 37-44. https://doi.org/10.1016/j.wasman.2010.12.019