Inteligencia artificial como herramienta automatizada para monitoreo de fauna silvestre mediante el modelo YOLOv5

Ing. Rubén Darío Calderón Cepeda[1], Dr. Edgar Gustavo López Saut[1], Ing. Ramiro Robles Villanueva[1], Dra. Merit Cisneros González[1]

[1]Tecnológico Nacional de México / Instituto Tecnológico del Valle del Guadiana, División de Estudios de Posgrado e Investigación, Maestría en Ingeniería.

Km. 22.5 Carretera Durango-México, Villa Montemorelos, Dgo. México. C.P. 34371

Resumen

En el presente artículo, se mostrarán los resultados obtenidos al desarrollar una red neuronal de tipo convolucional (CNN) a manera de prototipo mediante el modelo de YOLOv5 (You Only Look Once). Esto se hizo utilizando los servicios en línea de Google colab, Roboflow y MakeSense.ai para la detección de fauna silvestre en imágenes de estudio en áreas naturales protegidas. Esto con el fin de encontrar una alternativa a las técnicas de monitoreo actuales, buscando ser más eficiente y con buen desempeño para el monitoreo de animales en áreas de interés utilizando modelos de detección con inteligencia artificial (I.A.).

Es importante mencionar que el modelo aún se encuentra en fase de prototipo, por lo cual, no se cuenta con pruebas en campo del mismo. Para el desarrollo del modelo se utilizaron fotografías obtenidas por cámaras trampa marca "Browning" modelo BTC-1XR que se colocaron en el bioparque Sahuatoba en la ciudad de Durango, Dgo. De igual forma, una parte de ellas fueron brindadas por investigadores del ITVG, las cuales se obtuvieron con cámaras trampa dentro de su hábitat natural. En este artículo, se evaluó el desempeño de 3 modelos diferentes siguiendo la arquitectura YOLOv5x generando una comparativa de los resultados obtenidos por medio de las métricas de precisión, mAP o "mean average precisión" y la función de pérdida, con el fin de seleccionar el modelo con el mejor desempeño en clasificación.

Palabras clave--Red Neuronal Convolucional, fauna silvestre, YOLOv5.

Abstract

In this article, it will be shown the results obtained during the develop of a convolutional type neuronal network (CNN) as a prototype with YOLOv5 model (You Only Look Once). This was made by giving use to the online services Google colab, Roboflow and MakeSense.ai for the detection of wildlife in study images in protected natural areas. This having the purpose of finding an alternative for the actual monitoring technics trying to be more efficient and having a good performance for the monitoring of animals species in areas of interest using detection models with artificial intelligence (A.I.).

It is important to mention that the model is still in a prototype phase, which means it do not have been tested in field. For the development of the model, they were used photographs obtained with "Browning" brand camera traps model BTC-1XR that were placed in the bio park Sahuatoba in the city of Durango, Dgo. In addition, part of them were provided by researchers of the ITVG, which were obtained with camera traps inside their natural habitat. In this article, it was evaluated the performance of 3 different models following YOLOv5x architecture by making a comparative of the results obtained with the metrics precision, mAP or mean average precision and the loss function in order to choose the model with the best classification performance.

Keywords—*Convolutional Neuronal Networks, wildlife, YOLOv5.*

1. INTRODUCCIÓN

Las áreas naturales han permitido a la humanidad lograr un alto nivel de conservación de la biodiversidad al ser puntos centrales que han sido regulados y administrados minuciosamente para lograr este fin [1]. En México, al ser uno de los países megadiversos, estas áreas han cobrado una gran importancia para la conservación de flora y fauna. Debido a esto, se han desarrollado diferentes técnicas de monitoreo que permiten tener un mayor control de las especies que habitan en estas áreas [2].

En el caso de la fauna, la técnica más utilizada es mediante imágenes de estudio, las cuales, son obtenidas mayormente por medio de cámaras trampa debido a su gran funcionalidad tanto en el día como en la noche permitiendo fotografiar todo tipo de especies, tanto diurnas como nocturnas [3]. Sin embargo, realizar monitoreos de la fauna suele ser una tarea muy complicada y lenta al tener que estudiar por separado cada una de las imágenes obtenidas.

En esta investigación se busca innovar la manera en que se realiza este tipo de monitoreos por medio de la integración de inteligencia artificial dando uso de las redes neuronales convolucionales (CNN) con el modelo de código abierto YOLOv5 (You Only Look Once) [4].

2. CONTENIDO

2.1 Metodología

Para el desarrollo del sistema de reconocimiento por medio de redes neuronales convolucionales se requirió obtener un conjunto de fotografías divididas en 8 categorías diferentes incluyendo 6 especies de animales siendo estos venados, mapaches, zorros, coyotes, pecarís y linces. Así mismo, se agregaron 2 categorías más a identificar integrando fotografías de personas y objetos que se categorizaron como otros y estando conformados por automóviles, bicicletas, gatos y perros [5]. La figura 1 muestra un ejemplo de las fotografías obtenidas, en este caso se muestra un venado cola blanca.

Una parte de estas fotografías fueron obtenidas mediante cámaras trampa marca “Browning” modelo BTC-1XR que se colocaron en el bioparque Sahuatoba en la ciudad de Durango, Dgo. El resto de las imágenes utilizadas fueron brindadas por investigadores del Instituto Tecnológico del Valle del Guadiana, las cuales obtuvieron con cámaras trampa dentro de su hábitat natural, es decir, dentro del área del parque ecológico “El Tecuán”, Dgo, lugar donde se busca implementar el modelo en un futuro. En total se obtuvieron 3,000 fotografías, las cuales fueron etiquetadas en un sitio online llamado “makesense.ai” [6].

Figura 1 Foto venado cola blanca utilizada para entrenamiento

Fuente: elaboración propia obtenida con cámara trampa en el parque ecológico “El Tecuán”, Dgo

Las fotografías obtenidas junto con sus respectivas etiquetas conformaron el “dataset”, el cual se almaceno en la nube dentro de un servicio online llamado “Roboflow” donde se hizo un reescalado de las fotografías a 640x640 píxeles permitiéndonos tener cada una de ellas del mismo tamaño con la finalidad de no saturar los recursos que nos brinda el servicio online donde se realizó el entrenamiento de la red neuronal [7]. Así mismo, en “Roboflow” se dividió el “dataset” en 3 partes, la primera usada para el entrenamiento siendo el 70% del “dataset” con 2,101 fotografías, la segunda parte usada para la validación es el 20% con 599 fotografías y la tercera parte usada para prueba ocupa el 10% restante con 300 fotografías [8]. Cada parte mantiene sus etiquetas correspondientes, es decir, en cada una de las partes cada fotografía tiene su respectiva etiqueta [9].

Para el entrenamiento se utilizó un servicio de “Google” llamado “Google colab” dando uso del modelo YOLOv5 para entrenamientos personalizados [10]. Primeramente, se instalaron los requerimientos necesarios para trabajar con este modelo, las cuales fueron obtenidas desde “GitHub”. De igual forma, se incluyeron las librerías de “Roboflow”, “Torch” y “os”. El código fue trabajado en lenguaje Python versión 3.

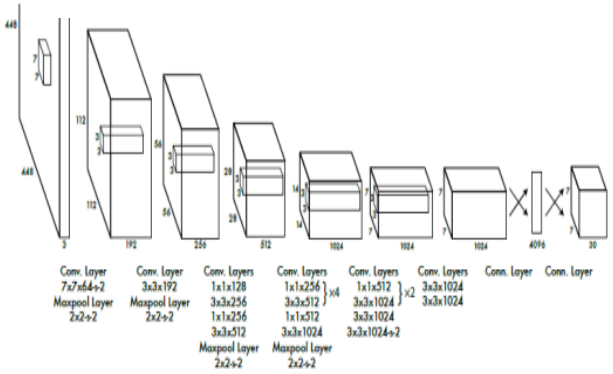
Lo siguiente fue leer el dataset, para lo cual “Roboflow” generó un código para exportarlo a “Google colab” y poder ser usado en el entrenamiento.

Finalmente, para realizar el entrenamiento se le indico al sistema los parámetros necesarios para entrenar la red neuronal, siendo estos el tamaño de las imágenes, el número de lotes, el número de épocas de aprendizaje, la localización del “dataset”, el cual fue previamente exportado, y la arquitectura a utilizar [11].

Para encontrar el modelo más adecuado se realizaron 3 entrenamientos diferentes donde se modificaron el número de épocas de entrenamiento correspondiendo a 10, 15 y 20 épocas. Sin embargo, los demás parámetros se mantuvieron iguales, las imágenes de 640x640 píxeles, 16 lotes y la arquitectura YOLOv5, mostrada en la figura 2, donde se utilizó un peso preentrenado siendo YOLOv5x [12].

Esta arquitectura está compuesta por 24 capas de convoluciones y 2 capas completamente conectadas. Así mismo, cuenta con 4 capas de “maxpooling” lo que reduce el tamaño de las imágenes y comprime sus características de forma que dichas características serán cada vez más concentradas. Se van alternando capas de convolución de 1x1 y 3x3 píxeles, donde las capas de 1x1 píxeles permiten reducir la profundidad de los mapas de características. La última capa de convolución genera un tensor de 7x7x1024 píxeles, el cual finalmente será reducido al aplicar 2 capas completamente conectadas dando como salida final un tensor de 7x7x30 píxeles [13].

Figura 2 Arquitectura YOLOv5



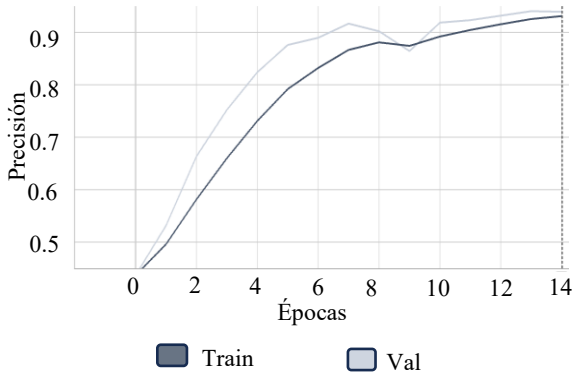
Fuente: Ramos B. (2022) Arquitectura CNN de YOLO [Figura]

La arquitectura YOLOv5 utiliza “LeakyReLu” como función de activación a excepción en la capa lineal en donde utiliza una capa de activación lineal. La función “LeakyReLu” transforma los valores introducidos multiplicando los negativos por un coeficiente para rectificarlos y en el caso de los coeficientes positivos no son rectificados, se dejan sin cambios [14].

2.2 Resultados y discusión

Para seleccionar el modelo más adecuado de los 3 obtenidos, se evaluaron 3 parámetros diferentes, siendo precisión, mAP o “mean average precision” [15] y el error, el cual se divide en 3, teniendo el error de la caja, error de la clase y el error del objeto [16].

Figura 3 Precisión del modelo con 15 épocas de aprendizaje.

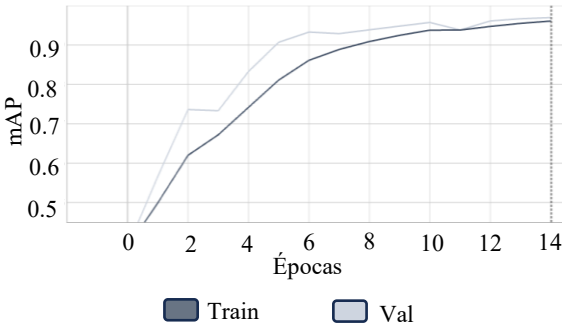


Fuente: elaboración propia a partir de resultados obtenidos en TensorBoard

La precisión, la cual es mostrada en la figura 3, nos indica que tan certera es una predicción al momento de brindarle a la red neuronal una imagen de estudio, en este caso para el modelo de 15 épocas se obtuvo una precisión del 93.3%. Lo que demuestra que este modelo presenta un alto nivel de precisión para la clasificación de especies.

Así mismo, mediante esta grafica se verifica que el modelo no llegue a un nivel de sobreentrenamiento. Para esto, se verifica que nuestro modelo no permanezca en su tope de aprendizaje por mucho tiempo, es decir, que nuestro modelo no este estabilizado o sin incremento durante muchas épocas debido a que este ya ha aprendido lo más que puede y todo aquello que intente aprender a partir de ese punto puede causar errores de clasificación. En el caso de este modelo su tope llego en la época número 15, deteniendo su entrenamiento justo al llegar al momento máximo de aprendizaje y de esta forma evitando llegar a un sobreentrenamiento.

Figura 4 mAP del modelo con 15 épocas de aprendizaje.

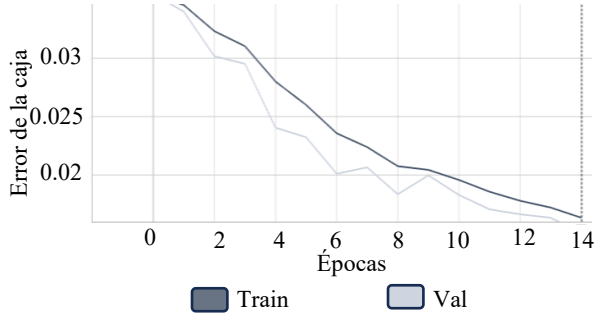


Fuente: elaboración propia a partir de resultados obtenidos en TensorBoard

En la figura 4 se muestra el mAP o “mean Average Precision”, la cual es una métrica utilizada para medir el desempeño general de las redes neuronales. Esta métrica considera 3 aspectos diferentes de manera conjunta, buscando evaluar el desempeño del modelo intentando lograr un balance entre estos 3 aspectos, sin embargo, también pueden ser evaluados de manera individual. Los aspectos que considera esta métrica son precisión, IoU o “Intersection over Union” y “recall”.

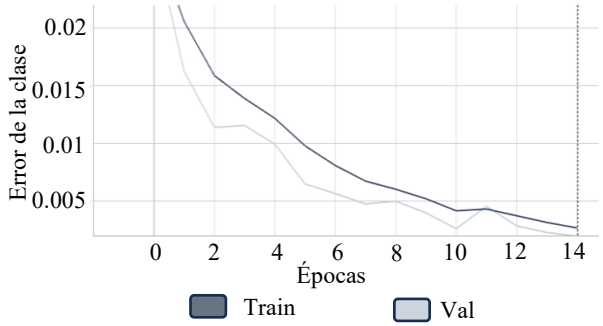
Este modelo muestra un mAP del 97% indicando un gran desempeño y balance de estas 3 métricas al realizar una clasificación, es decir, los niveles de precisión, “IoU” y “recall” son altos y mantienen un balance entre ellos.

Figura 5 Error de la caja del modelo con 15 épocas de aprendizaje.



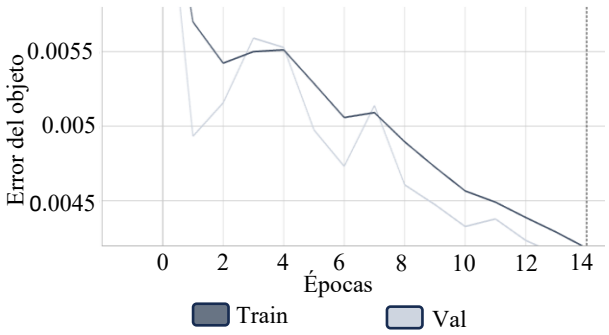
Fuente: elaboración propia a partir de resultados obtenidos en TensorBoard

Figura 6 Error de la clase del modelo con 15 épocas de aprendizaje.



Fuente: elaboración propia a partir de resultados obtenidos en TensorBoard

Figura 7 Error del objeto del modelo con 15 épocas de aprendizaje



Fuente: elaboración propia a partir de resultados obtenidos en TensorBoard

El error es una métrica que indica el grado de realizar una clasificación incorrecta, es decir, esto permite conocer la facilidad que tiene la red neuronal de cometer una equivocación al realizar clasificaciones. En la figura 5 se encuentra el error de la caja, el cual está relacionado con el "IoU". Este error menor del 2% muestra que el modelo tiene buena exactitud al colocar la caja de predicción sobre la caja "Ground truth".

El error de la clase, mostrado en la figura 6, está relacionado con la precisión. Este error menor del 0.5% muestra que el modelo tiene un alto nivel de precisión al clasificar la especie animal dentro de la clase a la que pertenece. En la figura 7 encontramos el error del objeto, el cual está relacionado con el "recall" o exhaustividad. Este error menor del 0.45% muestra que el modelo comete pocos errores al clasificar especies dentro de cada una de las 8 categorías de clasificación consideradas.

Así mismo, se comparó el modelo con el método de monitoreo que busca mejorar, es decir, el método de monitoreo por cámaras trampa. Este método suele ser una tarea muy compleja debido a que requiere de un gran control, concentración y organización por parte del investigador.

En un principio, en ambas alternativas el proceso comienza de la misma manera al obtener las imágenes de estudio de las cámaras trampa. Sin embargo, el revisar la información obtenida suele ser mas complejo en el método común, ya que el investigador debe revisar cada fotografía detenidamente para registrar las observaciones de especies. Además, los resultados también dependerán de la experiencia de los investigadores encargados del censo, pero errores muy comunes como la mala organización de las imágenes de estudio y la mala revisión de las mismas son un factor importante que puede provocar errores en el conteo y control de las especies. En cambio, al utilizar una inteligencia artificial para analizar la información obtenida de las cámaras se pueden evitar estos problemas. Al poder leer lotes de imágenes o incluso todas las imágenes en una sola iteración, dependerá del nivel de control que los investigadores deseen tener, permitiendo tener una mejor organización de las mismas, evitando que no falte la revisión de ninguna imagen o al evitar la revisión doble de alguna fotografía, logrando detectar las especies encontradas con un gran nivel de precisión, en este caso del 93.3%.

Así mismo, es importante mencionar que al implementar el modelo no se requieren realizar demasiados ajustes al solo requerir la instalación de las librerías y programas necesarios para su funcionamiento. De igual forma, no es necesario invertir para implementarlo, ya que no se requiere pagar para utilizar dichas librerías o programas al ser de uso libre, ni tampoco se requiere comprar equipo extra o nuevo para usarse, teniendo la posibilidad de usar las fotografías de las cámaras trampa que se tengan y el

equipo de cómputo con el que se cuenta. El mantener los costos iguales mejorando el rendimiento y evitando los problemas mencionados, se demuestra que este tipo de sistemas con inteligencia artificial pueden ser una gran alternativa para el monitoreo de especies.

3. CONCLUSIONES Y RECOMENDACIONES

Para seleccionar el modelo más adecuado se compararon los desempeños de los 3 modelos según las métricas seleccionadas mostradas en la tabla 1.

Tabla 1 Comparación de desempeño de modelos.

MODELO	PRECISIÓN	RECALL	MAP
10	82.5 %	79.6 %	83.8 %
15	93.3%	94.7%	97%
20	88.4 %	86.2 %	87.6 %

Fuente: elaboración propia

Al evaluar el desempeño de cada uno de los modelos obtenidos se puede decir que el modelo más adecuado para realizar las clasificaciones es el modelo con 15 épocas de aprendizaje, considerando que los otros 2 modelos presentaron resultados menos eficientes que los obtenidos por este modelo. En el caso del modelo de 10 épocas, su bajo desempeño se deriva de problemas por subentrenamiento, considerando que los modelos llegaban a su máximo nivel de aprendizaje en la época número 15, este modelo no logro aprender lo suficiente para mostrar un buen desempeño. Caso contrario con el modelo de 20 épocas, el cual mostro un bajo desempeño derivado de sobreentrenamiento puesto que este modelo continuó con más épocas de aprendizaje después de llegar a la época donde llego a su nivel de aprendizaje máximo.

Considerando el nivel de precisión en la clasificación al que pueden llegar este tipo de modelos y la eliminación de errores en el conteo del censo, se considera que puede ser una gran herramienta para el monitoreo de fauna silvestre, ayudando a mejorar la velocidad en la que se analizan las imágenes de estudio obtenidas en las áreas naturales protegidas.

Para estudios próximos, se pretende variar los pesos del modelo donde se podría cambiar a alguna de las variantes de la familia de YOLOv5 como lo son YOLOv5s, YOLOv5n o YOLOv5m lo que ayudaría a variar la velocidad y precisión de la clasificación. Así mismo, se recomendaría realizar pruebas con diferente número de imágenes y modelos de redes neuronales diferentes.

3.1 Observaciones generales

Los autores agradecen al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) y al Tecnológico Nacional de México (TecNM) por el financiamiento.

4.REFERENCIAS

[1] H. González Ocampo, P. Cortés-Calva, L. I. Íñiguez Dávalos y A. Ortega-Rubio, «Las áreas naturales protegidas de México,» *Investigación y Ciencia,* vol. 22, nº 60, pp. 7-15, 2014.

[2] A. P. Orellana Baque, B. J. Bernabé Mendoza y P. Piedrahita, «Sistema de monitoreo de fauna silvestre del Bosque Protector la Prosperina mediante el uso de cámaras trampa,» Escuela Superior Politécnica del Litoral, 2022.

[3] K. Sánchez Pinzón, E. L. Hernández Pérez, J. F. Moreira Ramírez, N. Mayer y R. Á. Reyna Hurtado, «Fototrampeo: Descubriendo lo que no podemos ver,» Ecofronteras, vol. 21, nº 61, 2017.

[4] B. Huda Husain y T. Osawa, «Advancing Fauna Conservation through Machine Learning-Based Spectrogram Recognition: A Study on Object Detection using YOLOv5,» Jurnal Sumberdaya Alam Dan Lingkungan, vol. 10, nº 2, pp. 58-68, 2023.

[5] M. Choiński, M. Rogowski, P. Tynecki, D. Kuijper, M. Churski y J. Bubnicki , «A First Step Towards Automated Species Recognition from Camera Trap Images of Mammals Using AI in a European Temperate Forest,» Computer Information Systems and Industrial Management, vol. 12883, pp. 299-310, 2021.

[6] S. Hegde y G. Shetty, «Hybrid Approach for apple fruit disease detection, yield estimation and,» International Research Journal of Engineering and Technology (IRJET), vol. 09, 2022.

[7] S. Tariq, A. Hakim, A. A. Siddiqi y M. Owais, «An image dataset of fruitfly species (Bactrocera Zonata and Bactrocera Dorsalis) and automated species classification through object detection,» Data in brief, vol. 43, 2022.

[8] H. Wang, S. Shang, D. Wang, X. He, K. Feng y H. Zhu, «Plant Disease Detection and Classification Method Based on the Optimized Lightweight YOLOv5 Model,» Agriculture, vol. 12, 2022.

[9] T. Hu, R. Yan, C. Jiang, N. V. Chand, T. Bai, L. Guo y J. Qi, «Grazing Sheep Behaviour Recognition Based on Improved YOLOV5,» Sensors, vol. 23, 2023.

[10] F. Jubayer, J. A. Soeb, A. N. Mojumder, M. K. Paul, P. Barua, S. Kayshar, S. S. Akter, M. Rahman y A. Islam, «Detection of mold on the food surface using YOLOv5,» Current Research in Food Science, vol. 4, pp. 724-728, 2021.

[11] Y. Xie, J. Jiang, H. Bao, P. Zhai, Y. Zhao, X. Zhou y G. Jiang, «Recognition of big mammal species in airborne thermal imaging based on YOLO V5 algorithm,» Integrative Zoology, vol. 18, pp. 333-352, 2023.

[12] Z. Li, A. Namiki, S. Suzuki, Q. Wang, T. Zhang y W. Wang, «Application of Low-Altitude UAV Remote Sensing Image Object Detection Based on Improved YOLOv5,» Applied Sciences, vol. 12, nº 16, 2022.

[13] S. Rozada Raneros, «Estudio de la arquitectura YOLO para la detección de objetos mediante deep learning,» Universidad de Valladolid, 2021.

[14] B. D. Ramos Caicedo, «Sistema de reconocimiento y conteo de productos de panadería,» Universidad de Antioquia, 2022.

[15] S. Wu, J. Wang, L. Liu, D. Chen, H. Lu, C. Xu, R. Hao, Z. Li y Q. Wang, «Enhanced YOLOv5 Object Detection Algorithm for Accurate Detection of Adult Rhynchophorus ferrugineus,» Insects, vol. 14, 2023.

[16] Y. Guo, P. Regmi, Y. Ding, R. B. Bist y L. Chai, «Automatic detection of brown hens in cage-free houses with deep learning methods,» Poultry Science, vol. 102, 2023.