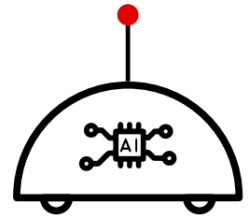




universidad
de león



MÁSTER EN ROBÓTICA E INTELIGENCIA ARTIFICIAL

RECONOCIMIENTO DE OBJETOS

Entrega 1: Dataset analysis

Autor:
Edgar Alcoba Casado

(marzo, 2024)

INDICE

1	Introducción	3
2	Dataset	4
3	Análisis del dataset	5
4	Creación del subset	10
5	Análisis del subset	11
6	Conclusiones	15

1 Introducción

El objetivo de esta práctica consiste en describir el dataset AnimalTrack utilizando diferentes tipos de gráficas para facilitar su comprensión. También se creará un dataset nuevo que será un subconjunto del original, compuesto por tres categorías diferentes y cinco vídeos por categoría, con el objetivo de reducir los recursos necesarios para computar los modelos que utilizarán este dataset en próximas prácticas.

2 Dataset

El dataset utilizado se llama AnimalTrack. Este dataset está diseñado para la detección y seguimiento de animales. Cuenta con un total de 58 vídeos de distintas resoluciones y duración. Los vídeos capturan momentos sobre animales en entornos silvestres y están divididos en 10 categorías: chicken, deer, dolphin, duck, goose, horse, penguin, pig, rabbit y zebra. Cada categoría cuenta con entre 5 y 7 vídeos distintos. El número promedio de objetos en seguimiento en cada secuencia es de más de 30.

Además, el dataset incluye anotaciones detalladas hechas a mano, con información sobre la posición de cada animal en cada fotograma en formato de bounding boxes. El formato que han seguido para realizar las anotaciones es el siguiente: [frame_id, target_id, top_left_x, top_left_y, width, height, confidence, class, visibility].

Para facilitar su uso, el dataset proporciona un script en Python que permite extraer automáticamente los fotogramas de los vídeos y los guarda como imágenes individuales separándolas en carpetas dependiendo del vídeo del que proceden. Las imágenes se guardan en formato JPG que es un formato que intenta equilibrar la calidad con un tamaño de archivo reducido. Después de ejecutar el script se puede ver que el dataset cuenta con un total de más de 25000 imágenes. Los vídeos cuentan con una mediana de 316 fotogramas, siendo de 196 el vídeo más corto y de 2269 el de mayor duración.

Asimismo, el dataset cuenta con archivos específicos para la división de los conjuntos de entrenamiento y prueba, asegurando que esta separación se realice a nivel de vídeo y no de imágenes individuales. Esto significa que el entrenamiento se realiza con imágenes de un conjunto de vídeos, mientras que la evaluación se hace con imágenes de vídeos completamente distintos. Esta estrategia busca simular una situación más realista, en la que un modelo de seguimiento recibe un vídeo nuevo en lugar de imágenes aleatorias de diferentes vídeos. Además, las anotaciones también están organizadas de la misma forma y especifican las detecciones que corresponden a cada vídeo.

Este dataset se puede encontrar en el siguiente enlace:

<https://hengfan2010.github.io/projects/AnimalTrack/index.html>

Existen varias opciones diferentes para descargar el dataset. Entre estas opciones está la posibilidad de descargarlo utilizando la plataforma Google Drive o Baidu Pan. También se puede elegir entre descargar el dataset completo o descargar únicamente algunas categorías. Se proporciona también un enlace diferente para descargar las divisiones de los conjuntos de entrenamiento y test.

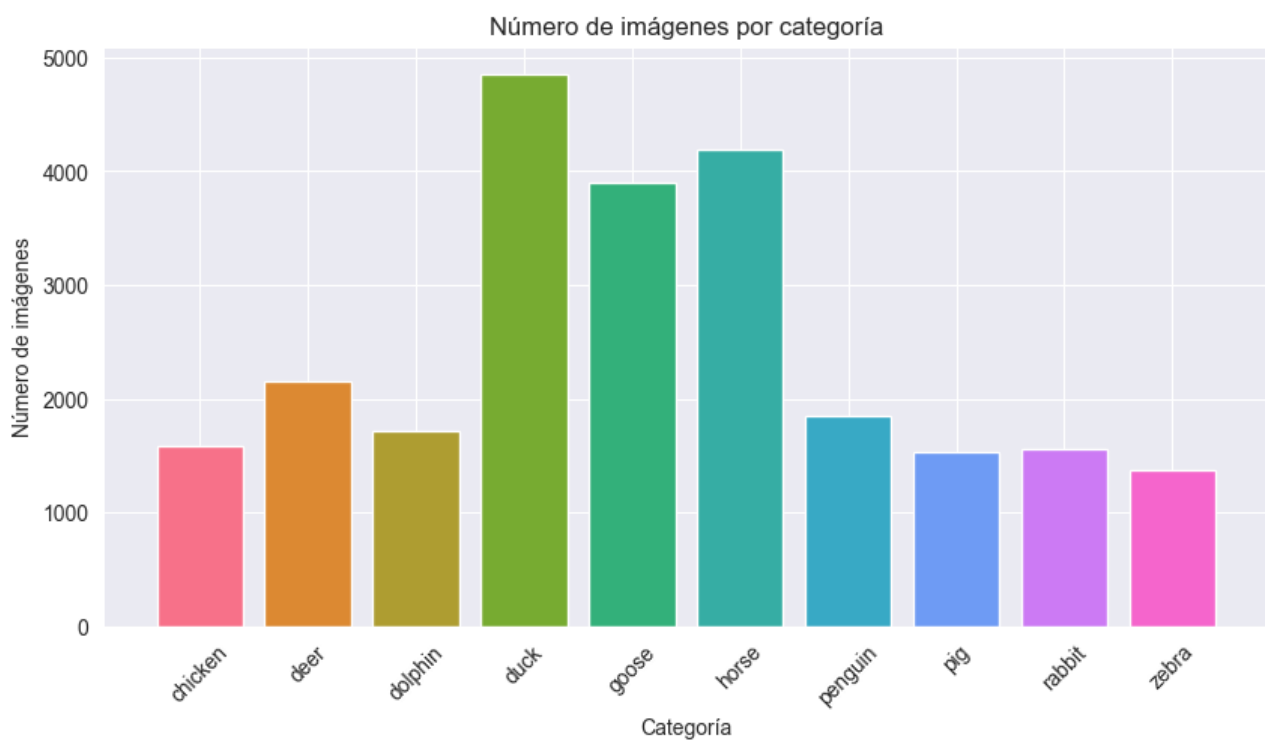
Los autores del dataset son:

Zhang, L., Gao, J., Xiao, Z., & Fan, H. (2023). [AnimalTrack: A Benchmark for Multi-Animal Tracking in the Wild](#). International Journal of Computer Vision (IJCV), 131, 496-513.

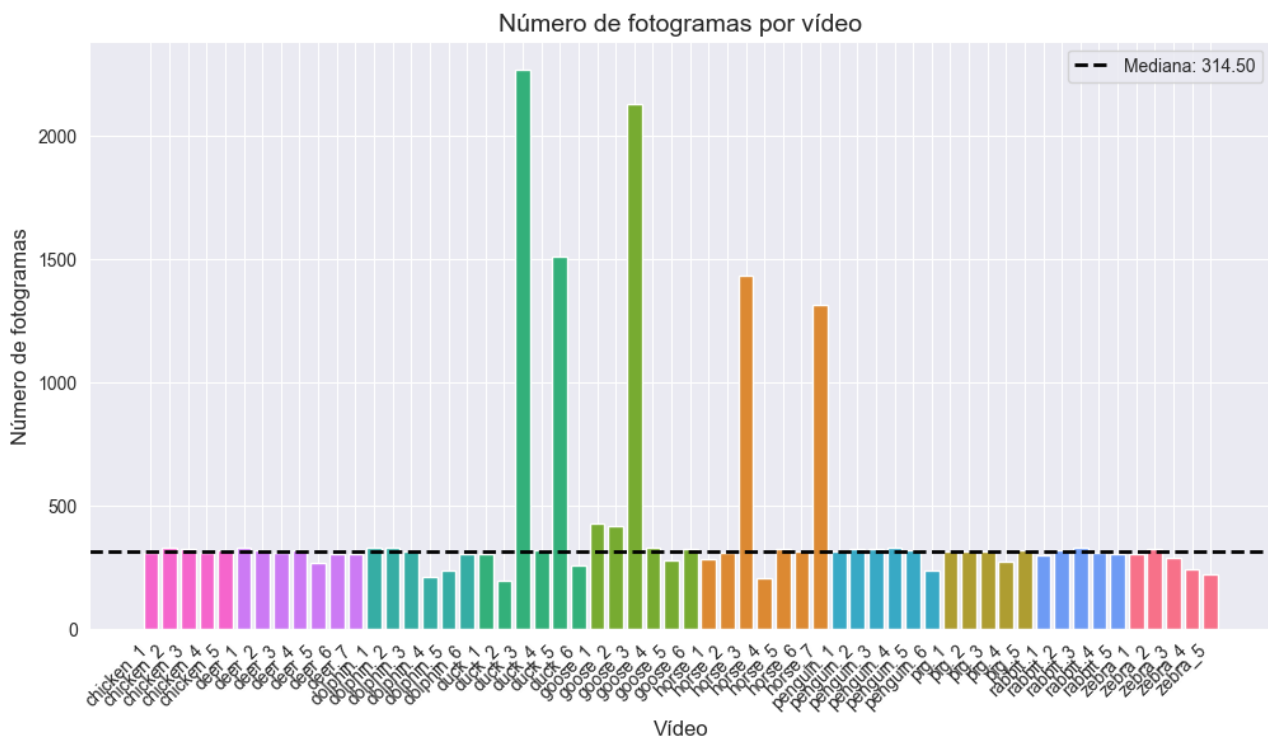
3 Análisis del dataset

Para comprender mejor la distribución y características del dataset, se han generado varias gráficas que ilustran sus relaciones.

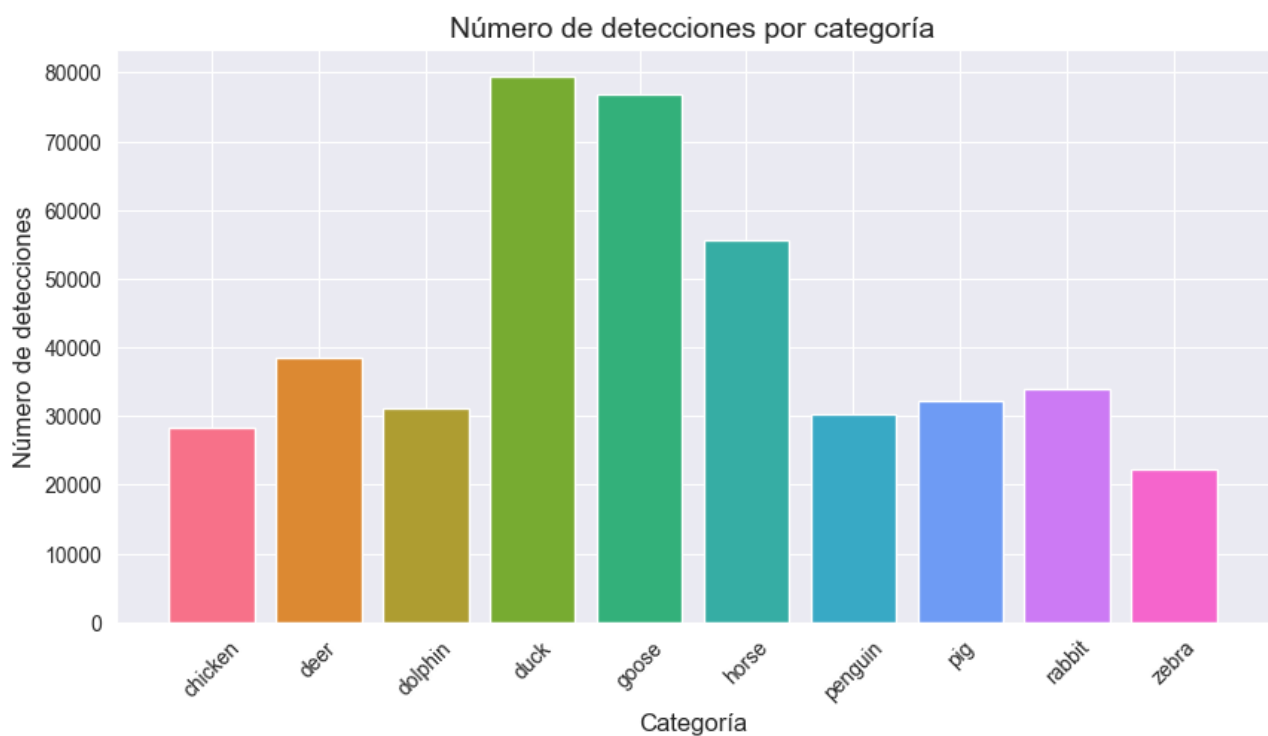
En la primera gráfica ilustra el número de imágenes que tiene cada categoría del dataset después de extraer todos los fotogramas de los vídeos. Se puede ver que las clases no están balanceadas en cuanto al número de fotogramas, lo que significa que algunas categorías tienen un mayor grado de representación que otras. Este desequilibrio puede afectar el rendimiento de los modelos que utilicen este dataset, por lo que será necesario utilizar técnicas de balanceo de datos o ponderación de clases durante el entrenamiento.



En la segunda gráfica se representa el número de fotogramas por vídeo junto con la mediana de todas las categorías. La mediana ayuda a comprender mejor la distribución del número de fotogramas por vídeo, reduciendo el impacto los valores de los extremos que afectan a métricas como la media.

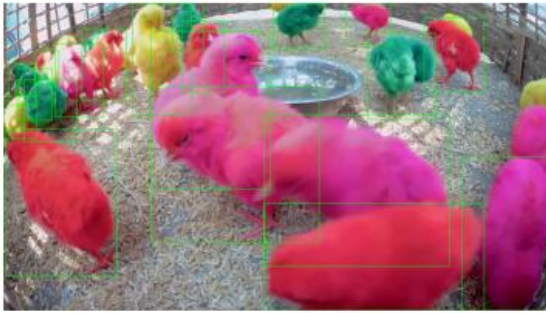


En la tercera gráfica se muestra el número de detecciones por categoría. La similitud entre esta gráfica y la primera indica que el número de detecciones está directamente relacionado con el número de fotogramas por categoría. Esto confirma el desequilibrio de clases mencionado previamente.



A continuación, se presentan varios ejemplos de imágenes del dataset. Cada una pertenece a una categoría distinta y se han aplicado las bounding boxes que le corresponden. Estas imágenes permiten visualizar cómo están anotados los objetos dentro del dataset.

chicken



deer



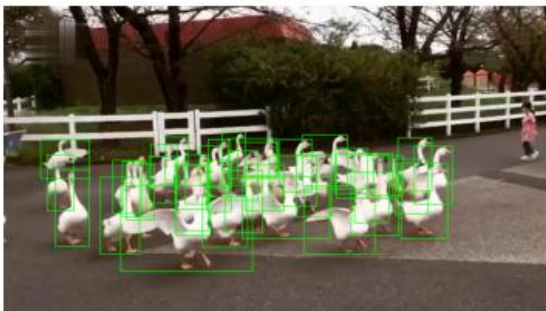
dolphin



duck



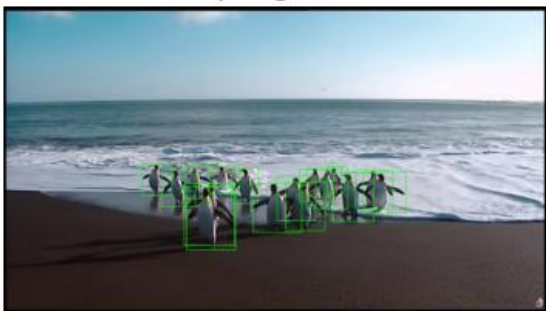
goose



horse



penguin



pig



rabbit

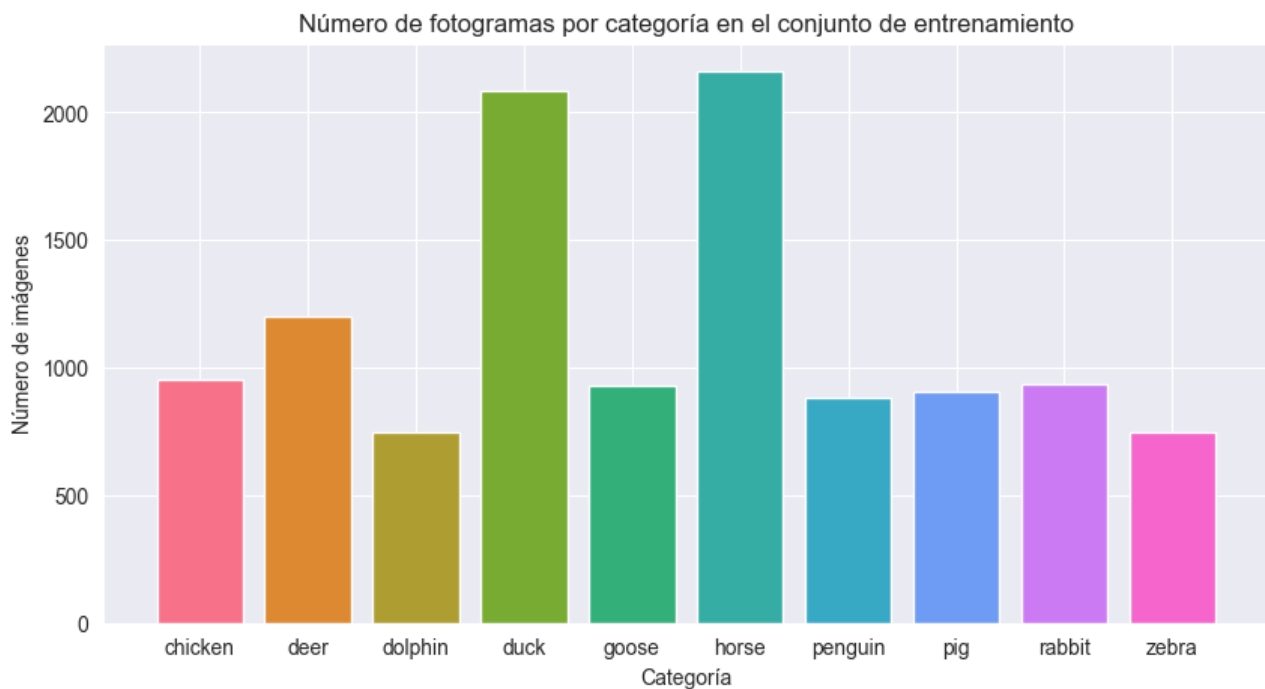


zebra

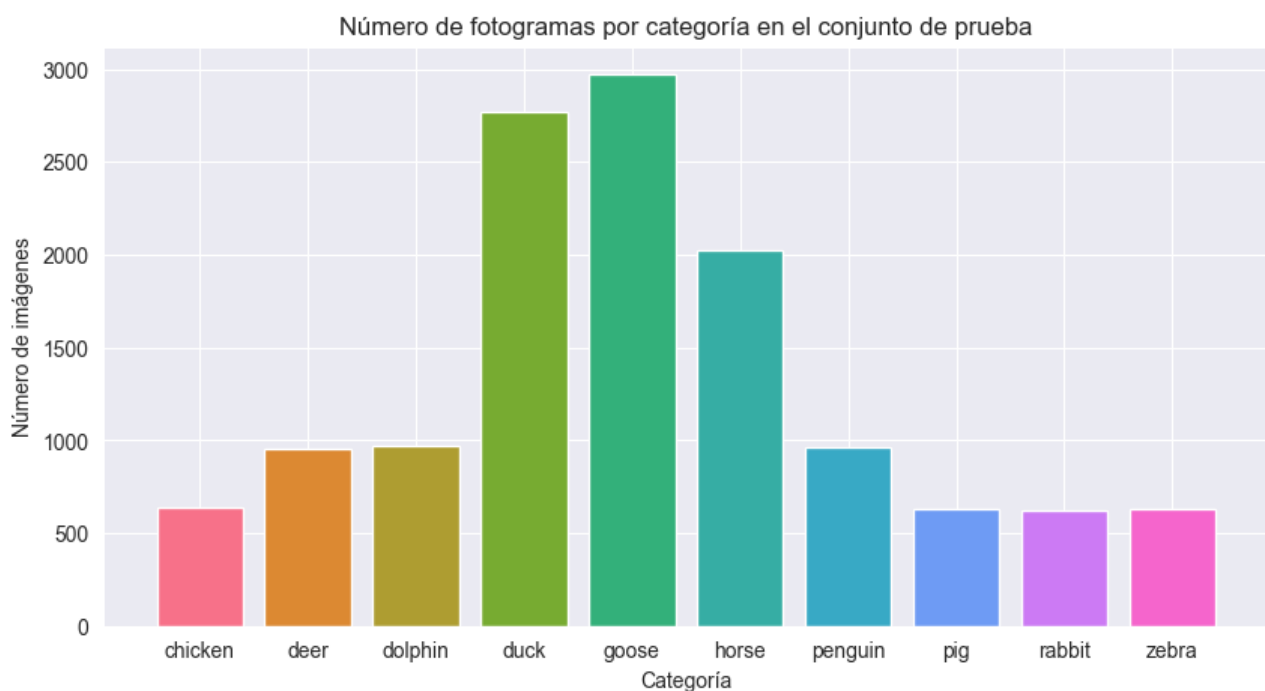


Se ha realizado la división de los conjuntos de entrenamiento y test como se especifica en los archivos del dataset destinados a este propósito, asegurando que la separación se haga a nivel de vídeo y no de imagen.

En la cuarta gráfica se puede ver el número de fotogramas por categoría en el conjunto de entrenamiento.



En la quinta gráfica se muestra el número de fotogramas por categoría en el conjunto de prueba.



Se observa que, debido a la variabilidad en la duración de los vídeos, en algunos casos el conjunto de prueba puede contener más fotogramas que el conjunto de entrenamiento. Sin embargo, dado que el conjunto de entrenamiento contiene un mayor número de vídeos con situaciones diferentes, esto puede contribuir a una mejor capacidad de generalización del modelo.

El código utilizado para crear todas las gráficas que se muestran en este documento se puede encontrar en el siguiente enlace:

https://github.com/ULE-Informatica-2024-2025/reob25-EdgarAlcoba/blob/main/Entrega_1/D1_Dataset.ipynb

o en este otro:

https://github.com/EdgarAlcoba/reob25-EdgarAlcoba/blob/main/Entrega_1/D1_Dataset.ipynb

4 Creación del subset

Se ha creado un subconjunto del dataset AnimalTrack con el objetivo de reducir el consumo de memoria y la capacidad de cómputo necesaria para entrenar y evaluar modelos que usen este subset.

Se han seleccionado las categorías pig, chicken y rabbit porque todas tienen un número similar de fotogramas (consultar gráfica 1), permitiendo que haya un mejor equilibrio entre clases. Además, cada una de estas categorías cuenta con cinco vídeos, lo que permite cumplir con el requisito que se solicitaba sin necesidad de realizar selecciones manuales adicionales para balancear el número de imágenes entre clases.

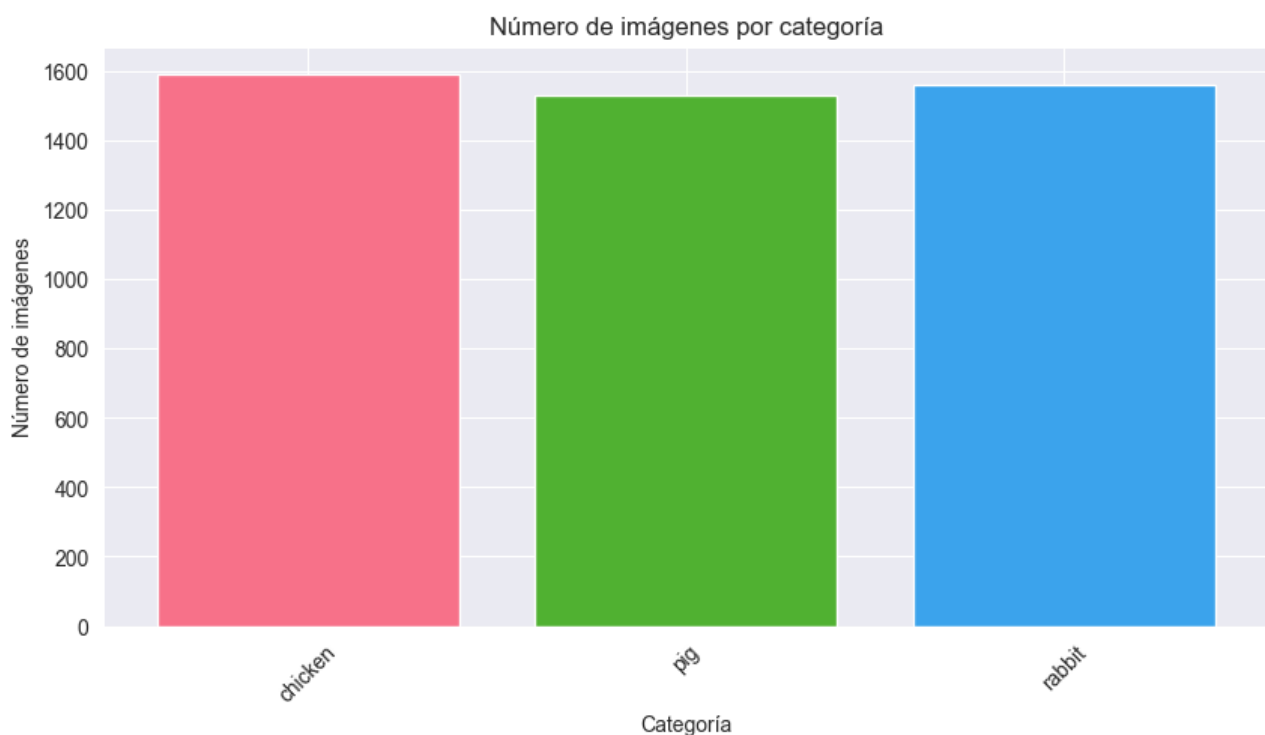
Este subconjunto seguirá utilizando el mismo formato de anotaciones que en el dataset original, utilizando la estructura: [frame_id, target_id, top_left_x, top_left_y, width, height, confidence, class, visibility]. También se conservará la misma división de los conjuntos de entrenamiento y test.

Para facilitar su uso, el subset mantiene la misma estructura de carpetas y nombres de archivos que el dataset original. Gracias a esto, es posible trabajar con el dataset completo o con el subset sin necesidad de realizar modificaciones adicionales en el código de preprocesamiento o entrenamiento. Además, se puede descargar únicamente las categorías seleccionadas utilizando los enlaces de descarga originales, facilitando su distribución al no ser necesario crear nuevos enlaces de descarga.

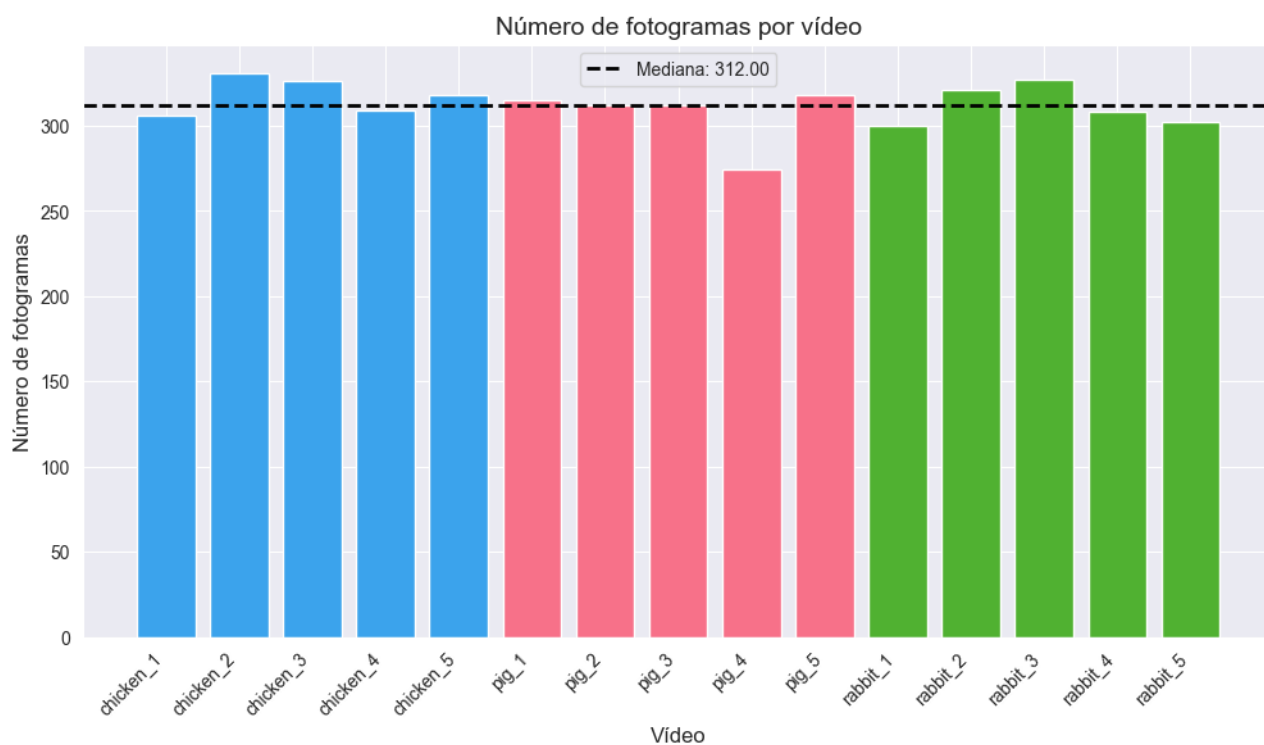
5 Análisis del subset

Para evaluar la composición del dataset reducido, se han generado diversas gráficas que permiten analizar su distribución y características principales.

En la primera gráfica, se muestra el número total de imágenes por categoría tras extraer los fotogramas de los vídeos seleccionados. Se observa que las clases están más balanceadas en comparación con el dataset original, con un rango de 1,500 a 1,600 imágenes por categoría. Aunque no es un equilibrio perfecto, la diferencia es mínima, lo que reduce el riesgo de sesgo en el entrenamiento de un modelo.



En la segunda gráfica, se representa la cantidad de fotogramas por vídeo y su mediana. En este caso, la mediana es de 312 fotogramas por vídeo, y no se observan valores extremos demasiado alejados de ella, lo que indica que los vídeos seleccionados tienen una duración uniforme.

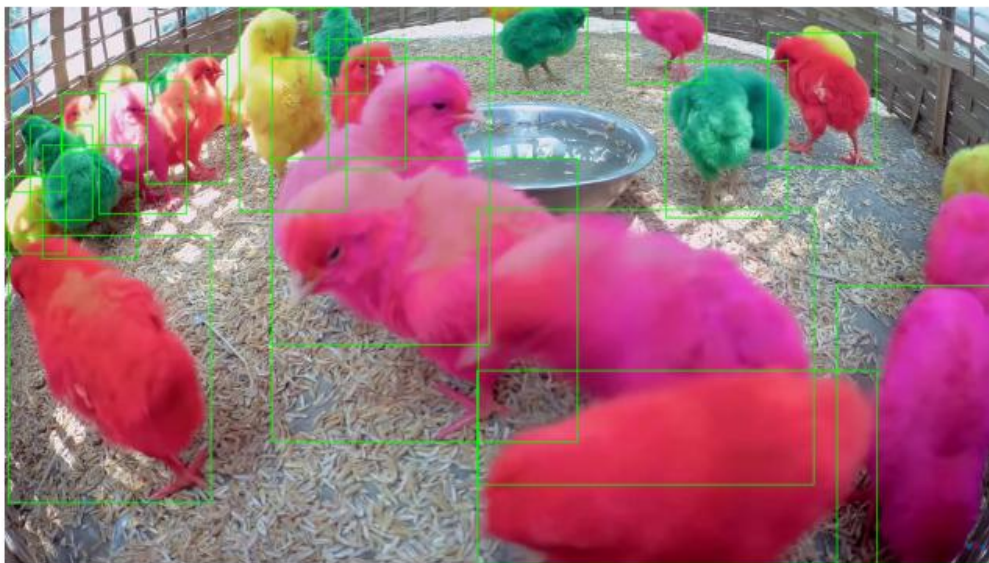


En la tercera gráfica, se analiza el número de detecciones por categoría. Aunque hay algunas diferencias entre clases (con alrededor de 28000 detecciones para chicken, 32000 para pig y 34000 para rabbit), el subset sigue siendo significativamente más equilibrado que el dataset completo.



A continuación, se presentan ejemplos de imágenes de cada categoría del subset con sus correspondientes bounding boxes aplicadas, lo que permite visualizar cómo son las anotaciones de los animales.

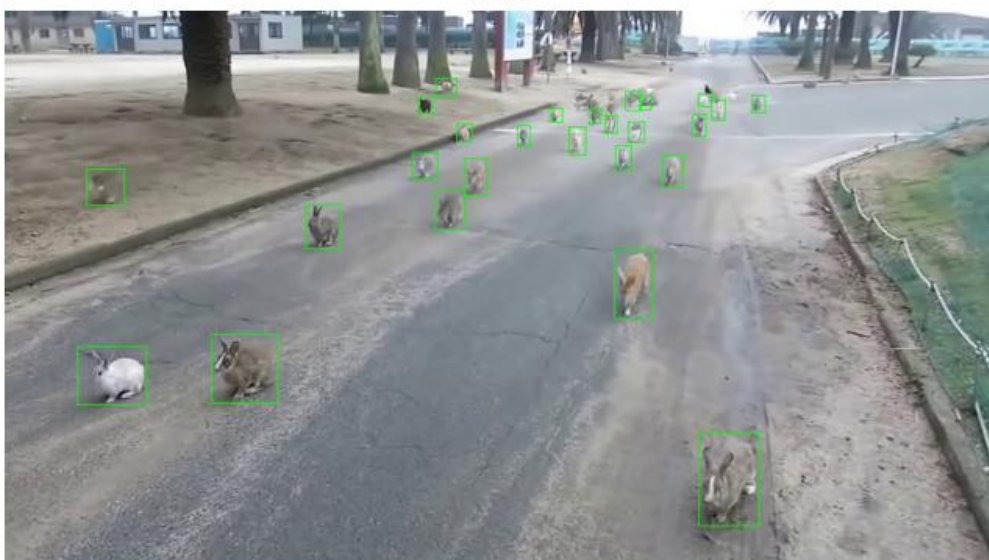
chicken



pig



rabbit

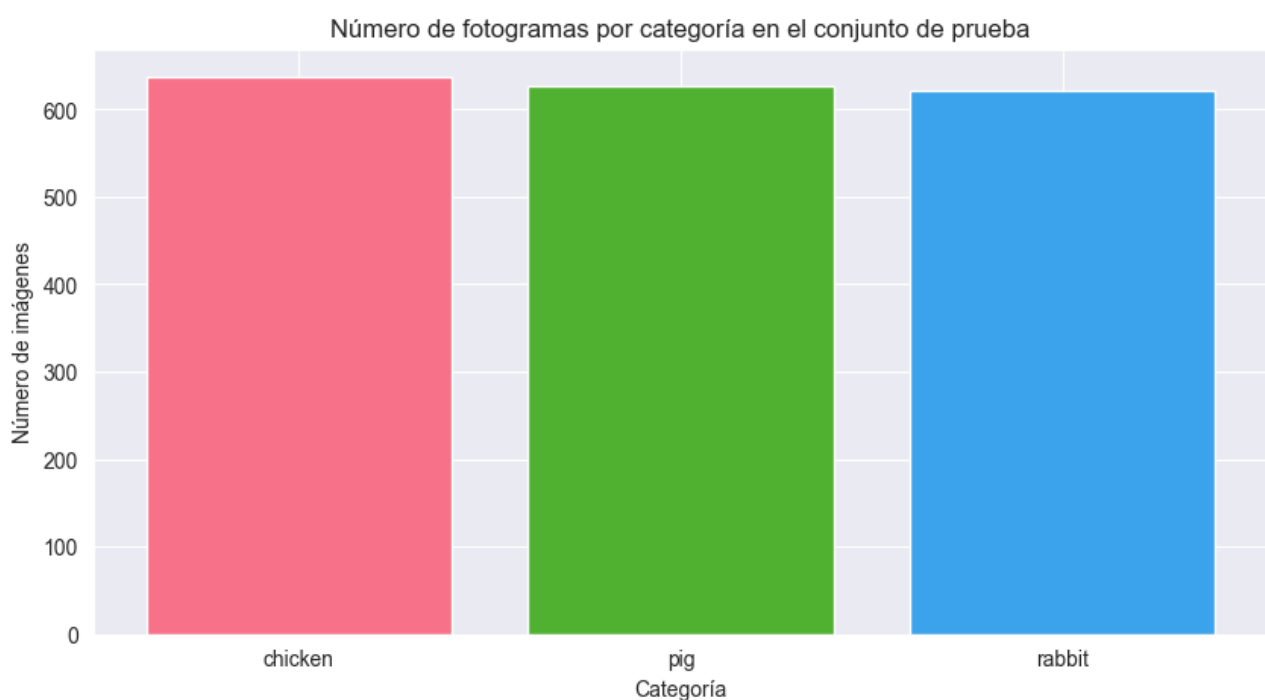
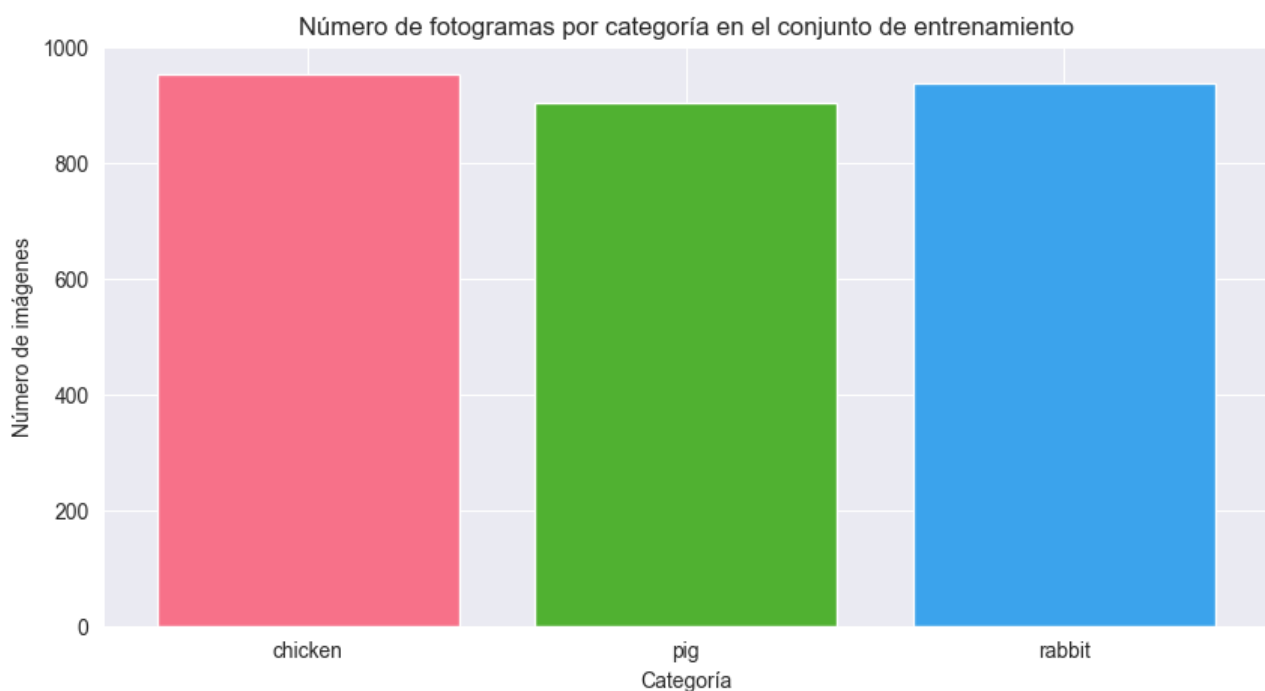


Se ha realizado la división de los conjuntos de entrenamiento y prueba manteniendo la especificación del dataset original y se ha conseguido un balance adecuado entre clases.

El conjunto de entrenamiento contiene entre 850 y 900 imágenes por categoría.

El conjunto de prueba es más reducido, con aproximadamente 650 imágenes por categoría.

En la cuarta y quinta gráfica, se muestra esta distribución, reflejando una proporción cercana al 60% de muestras para realizar el entrenamiento y un 40% para prueba. Esta división permite que el modelo aprenda con una cantidad suficiente de datos sin que la evaluación se vea afectada por un conjunto de prueba demasiado pequeño.



6 Conclusiones

En este informe se ha analizado el dataset AnimalTrack y se ha creado un subset más pequeño para reducir los recursos computacionales necesarios para procesarlo y almacenarlo.

El análisis del dataset original mostró un desequilibrio entre categorías, tanto en número de fotogramas como en el de detecciones, además de grandes variaciones en la duración de los vídeos, lo que afecta a la división entre los conjuntos de entrenamiento y prueba. La división de los conjuntos de entrenamiento y prueba también era algo particular, ya que en algunos casos había una mayor cantidad de datos en el conjunto de prueba que en el de entrenamiento.

Para crear el subset se seleccionaron las categorías chicken, pig y rabbit, consiguiendo un conjunto más equilibrado en cuanto a la cantidad de imágenes y la duración de los vídeos. Aunque hay pequeñas diferencias en el número de detecciones por categoría (28,000-34,000), su distribución es más uniforme que la del dataset original. Además, la división en entrenamiento y prueba (60%-40%) permite que el modelo se entrene con más datos sin comprometer la calidad de la evaluación.

En conclusión, el subset mantiene la misma estructura y formato que el dataset original, pero es más equilibrado y requiere una menor cantidad de recursos computacionales, lo que facilitará su uso en la creación y evaluación de los modelos que se crearán en prácticas posteriores.