

Entrega 1- Modelo de Clasificación de imágenes de Cámara Trampa

Julian David Santamaría Pabón
Universidad de Antioquia

Alejandro Ruiz
Universidad de Antioquia

1 Contexto de aplicación

La observación de la fauna mediante cámaras fototrampa se ha consolidado como una herramienta valiosa en estudios de ecología, permitiendo la recopilación de datos de manera no invasiva y automatizada. En proyectos de conservación, estas cámaras se utilizan para monitorear la vida silvestre sin perturbar a los animales, lo que resulta crucial para obtener información precisa sobre sus comportamientos y hábitats. Sin embargo, el uso de cámaras fototrampa genera grandes volúmenes de datos no curados, lo que hace que la anotación manual de las imágenes sea un proceso extremadamente laborioso y que consume mucho tiempo. Esta necesidad de etiquetar datos de manera eficiente ha impulsado la búsqueda de técnicas que automatizen este trabajo, facilitando así el análisis de la información recolectada [1, 2].

2 Objetivo de Machine Learning

El objetivo de este trabajo es predecir la clase de los animales presentes en las imágenes capturadas por cámaras trampa en el proyecto Snapshot Serengeti. Las imágenes fueron preprocesadas utilizando el modelo MegaDetector [3], un modelo de detección de objetos preentrenado para identificar animales. Gracias a este preprocesamiento, se garantiza la presencia de un animal en cada imagen, las cuales tienen una resolución de 380×380 píxeles. El conjunto de datos incluye un total de 46 clases de animales [4].

Las variables de entrada de nuestro modelo son las imágenes preprocesadas descritas anteriormente. Esta clasificación no solo permitirá una mejor organización y análisis de los datos recolectados, sino que también facilitará el monitoreo y la conservación de la fauna en su hábitat natural.

3 Dataset

3.1 Tipo de datos

El dataset de Snapshot Serengeti [4] comprende 46 clases. Esta versión del dataset incluye recortes de imágenes de 380×380 píxeles generados por el modelo MegaDetector [3] del proyecto Snapshot Serengeti, aplicando un umbral de confianza superior a 0.7. Solo se seleccionaron imágenes de cámaras trampa que contenían un solo animal para esta versión.

3.2 Tamaño

El dataset incluye un total de 340,972 imágenes, divididas en 230,971 para entrenamiento, 24,059 para validación y 85,942 para prueba. En total el dataset ocupa un tamaño en disco de 22.3 GB.

3.3 Distribución de las clases

La distribución de clases en el dataset se muestra en la siguiente tabla:

Especie	Especie	Especie
Aardvark (0.17%)	Baboon (0.96%)	Cheetah (1.32%)
Aardwolf (0.09%)	Bat Eared Fox (0.02%)	Civet (0.02%)
Buffalo (5.31%)	Bushbuck (0.13%)	Dik Dik (0.96%)
Caracal (0.04%)	Eland (0.02%)	Elephant (4.64%)
Gazelle Grants (1.32%)	Gazelle Thomsons (4.16%)	Genet (0.14%)
Giraffe (1.04%)	Guinea Fowl (0.02%)	Hare (0.04%)
Hartebeest (1.66%)	Hippopotamus (0.38%)	Honey Badger (0.03%)
Hyena Spotted (0.29%)	Hyena Striped (0.27%)	Impala (0.68%)
Jackal (0.11%)	Kori Bustard (0.02%)	Leopard (0.11%)
Lion Female (5.48%)	Lion Male (5.81%)	Mongoose (0.02%)
Ostrich (0.02%)	Porcupine (0.04%)	Reedbuck (0.30%)
Reptiles (0.38%)	Rhinoceros (0.24%)	Rodents (0.11%)
Secretary Bird (0.02%)	Serval (0.14%)	Topi (0.47%)
Vervet Monkey (0.92%)	Warthog (0.92%)	Waterbuck (0.68%)
Wildcat (1.04%)	Wildebeest (17.88%)	Zebra (17.44%)
Zorilla (0.01%)		

Table 1: Distribución de clases en el dataset

4 Métricas de desempeño

En esta sección, se describen las métricas de desempeño que se utilizarán para evaluar el modelo de *Machine Learning*.

- **Métricas de *Machine Learning*:**

- **Precisión (Accuracy):** Mide la proporción de predicciones correctas sobre el total de predicciones realizadas. Es fundamental para evaluar la efectividad del modelo en la clasificación de las especies de animales.

- **F1-Score:** Es la media armónica entre la precisión y el recall. Esta métrica es especialmente útil en situaciones donde hay un desbalance en las clases, ya que proporciona una medida más equilibrada de la precisión del modelo.
- **Precisión (Precision):** Indica la proporción de verdaderos positivos sobre el total de predicciones positivas. Es importante en contextos donde las falsas alarmas pueden tener un costo alto.
- **Recall:** También conocido como sensibilidad, mide la proporción de verdaderos positivos sobre el total de casos positivos reales. Es crucial para identificar correctamente las especies presentes en las imágenes.

5 Resultados previos

En esta sección presentamos algunos resultados del estado del arte utilizando el conjunto de datos Snapshot Serengeti.

Model	Test Acc (%)
WildCLIP [5]	61.778
WildCLIP-LwF [5]	64.390

Table 2: Comparación de desempeño en Snapshot Serengeti.

References

- [1] S. Beery, G. Van Horn, and P. Perona, “Recognition in terra incognita,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 456–473.
- [2] S. Beery, G. Wu, V. Rathod, R. Votel, and J. Huang, “Context r-cnn: Long term temporal context for per-camera object detection,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 13 075–13 085.
- [3] A. Hernandez, Z. Miao, L. Vargas, R. Dodhia, and J. Lavista, “Pytorch-Wildlife: A collaborative deep learning framework for conservation,” 2024.
- [4] A. Swanson, M. Kosmala, C. Lintott, R. Simpson, A. Smith, and C. Packer, “Data from: Snapshot serengeti, high-frequency annotated camera trap images of 40 mammalian species in an african savanna,” 2015.

- [5] V. Gabeff, M. Rußwurm, D. Tuia, and A. Mathis, “WildCLIP: Scene and animal attribute retrieval from camera trap data with domain-adapted vision-language models,” *International Journal of Computer Vision*, 2024.