UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

A Deep Learning Approach to Biodiversity: Clustering Wildlife Images in Yasuní Using CNNs

Proyecto de Titulación

Jenner Francois Baquero Morales

Felipe Grijalva, Ph.D.

Director de Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito para la obtención del título de Magíster en Ciencia de Datos

Quito, 01 de diciembre de 2024

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

A Deep Learning Approach to Biodiversity: Clustering Wildlife Images in Yasuní Using CNNs

Jenner Baquero

Nombre del Director del Programa:	Felipe Grijalva
Título académico:	Ph.D. en Ingeniería Eléctrica
Director del programa de:	Ciencia de Datos
Nombre del Decano del colegio Académico:	Eduardo Alba
Título académico:	Doctor en Ciencias Matemáticas
Decano del Colegio:	Ciencias e Ingenierías
Nombre del Decano del Colegio de Posgrados:	Dario Niebieskikwiat
Título académico:	Doctor en Física

Quito, diciembre 2024

© DERECHOS DE AUTOR

Nombre del estudiante: Jenner Francois Baquero Morales

Código de estudiante: 00339983

C.I.: 0605169416

Lugar y fecha: Quito, 01 de Diciembre de 2024.

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following graduation project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

DEDICATORIA

A Dios, por su amor, por cuidarme y guiarme en cada paso, por hacer posible lo que pensé imposible y por poner a las personas correctas en mi camino. A mi hermano Daniel, por toda la alegría que trae a mi vida y por acompañarme desde la niñez. A mi madre, Luz Morales, por creer en mí y apoyarme; a mi padre, Jenner Baquero, por incentivarme en la ciencia y la fe. A mis abuelos, Olivia Cargua y Manuel Morales. A mi hermano Esteban, por enseñarme sobre la ciencia y la vida desde niño. A mi tía Rosa Morales, por su cariño. A la Familia Guerrero Chávez por su ayuda incondicional y fe. A mis amigos y amigas: a Pamela Hidalgo, quien ha estado desde el principio de este camino con su apoyo siempre firme; a Nicole Marín, con quien comparto el cosmos; y a Andrés Jiménez, a quien quiero como a un hermano. Dedico este logro a todos ustedes. Disfruto saber que constantemente nos vemos crecer. Sin su apoyo, esto no habría sido posible. Que el tiempo demuestre que somos capaces de cosas grandes. Que el amor de Dios nos acompañe siempre.

AGRADECIMIENTOS

Expreso mi agradecimiento al Centro de Investigación Tiputini por proveer el conjunto de datos necesario para este proyecto. Mi reconocimiento especial a Óscar Cajamarca, quien debe ser considerado coautor por su participación activa y colaborativa durante la primera etapa del desarrollo. Agradezco también a Daniel Riofrío y Felipe Grijalva por su dirección y valiosos consejos en la toma de decisiones a lo largo del trabajo de titulación. A mis padres y hermanos, por su constante apoyo económico y emocional, y a Pamela y Nicole, por su incondicional respaldo en este proceso.

RESUMEN

El Parque Nacional Yasuní, un punto global de biodiversidad en Ecuador, enfrenta desafíos ecológicos significativos debido a las actividades humanas, lo que resalta la urgente necesidad de técnicas efectivas para el monitoreo de la vida silvestre. Este estudio explora la aplicación de Redes Neuronales Convolucionales (CNNs), específicamente ResNet-152, para procesar y agrupar un extenso conjunto de datos de 100,000 imágenes captadas por cámaras trampa recolectadas en Yasuní durante 15 años. Al aprovechar técnicas avanzadas de aprendizaje profundo, incluyendo la extracción de características y la visualización mediante t-Distributed Stochastic Neighbor Embedding (t-SNE), organizamos exitosamente el conjunto de datos en 11 clústeres distintos. Estos clústeres proporcionaron información crítica sobre la biodiversidad del parque, destacando especies de las familias FELIDAE, Tapiridae, Tayassuidae y Cervidae, junto con otras faunas únicas. Aunque muchos clústeres demostraron una alta precisión al agrupar especies similares, algunos revelaron limitaciones, como agrupaciones basadas en factores ambientales como la iluminación o el color de la imagen en lugar de características biológicas. Estos hallazgos subrayan la importancia de integrar métodos automatizados con el conocimiento ecológico experto para refinar los resultados de la clasificación. Además, desafíos como los clústeres superpuestos y la separación de imágenes diurnas y nocturnas de una misma especie resaltan oportunidades para mejoras metodológicas. Esta investigación ofrece un marco escalable para la limpieza, preparación y procesamiento de grandes conjuntos de datos de imágenes, contribuyendo al monitoreo y la preservación de la extraordinaria biodiversidad de Yasuní. También proporciona un enfoque integral de ciencia de datos para la gestión de conjuntos de imágenes extensos. El conjunto de datos curado, los análisis de clústeres y las visualizaciones sirven como herramientas valiosas para que los biólogos estudien la distribución y el comportamiento de las especies, allanando el camino para futuros avances en la tecnología de monitoreo de vida silvestre.

Palabras clave: Biodiversity, Environmental Conservation, Deep Learning, Convolutional Neural Networks, Image Clustering, Camera Traps, Wildlife Monitoring, Ecological Data Processing, Artificial Intelligence, Yasuní National Park.

ABSTRACT

Yasuní National Park, a global biodiver- sity hotspot in Ecuador, faces significant ecological challenges due to human activities, highlighting the urgent need for effective wildlife monitoring techniques. This study explores the application of Convolutional Neural Networks (CNNs), specifically ResNet-152, to process and cluster a vast dataset of 100,000 camera-trap images collected in Yasuní over 15 years. By leveraging advanced deep learning techniques, including feature ex- traction and t-Distributed Stochastic Neighbor Embed- ding (t-SNE) visualization, we successfully organized the dataset into 11 distinct clusters. These clusters provided critical insights into the park's biodiversity, showcasing species from the FELIDAE, Tapiridae, Tayassuidae, and Cervidae families, along with other unique fauna. While many clusters demonstrated high accuracy in grouping similar species, some revealed limitations, such as grouping based on environmental factors like lighting or image color rather than biological traits. These findings underscore the importance of integrating automated methods with expert ecological knowledge to refine classification outcomes. Furthermore, challenges such as overlapping clusters and the separation of day and night images for the same species highlight opportu- nities for methodological enhancements. This research offers a scalable framework for cleaning, preparing, and processing large image datasets, contributing to the monitoring and preservation of Yasuní's extraor- dinary biodiversity. It also provides a comprehensive data science approach for managing extensive image datasets. The curated dataset, clustering analyses, and visualizations serve as valuable tools for biologists to study species distribution and behavior while paving the way for future advancements in wildlife monitoring technology.

Key words: Biodiversity Monitoring, Yasuní National Park, Wildlife Conservation, Convolutional Neural Networks (CNNs), Deep Learning, ResNet-152, t-SNE Visualization, Image Clustering, Camera Trap Analysis, Ecological Data Processing, Big Data in Ecology, Species Classification, Gaussian Mixture Models (GMMs), En- vironmental Machine Learning, Wildlife Image Analysis.

TABLA DE CONTENIDO

ÍNDICE DE FIGURAS

A Deep Learning Approach to Biodiversity: Clustering Wildlife Images in Yasuní Using CNNs

Jenner Baquero Morales, Felipe Grijalva

***Abstract*—Yasuní National Park, a global biodiversity hotspot in Ecuador, faces significant ecological challenges due to human activities, highlighting the urgent need for effective wildlife monitoring techniques. This study explores the application of Convolutional Neural Networks (CNNs), specifically ResNet-152, to process and cluster a vast dataset of 100,000 camera-trap images collected in Yasuní over 15 years. By leveraging advanced deep learning techniques, including feature extraction and t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization, we successfully organized the dataset into 11 distinct clusters. These clusters provided critical insights into the park's biodiversity, showcasing species from the FELIDAE, Tapiridae, Tayassuidae, and Cervidae families, along with other unique fauna. While many clusters demonstrated high accuracy in grouping similar species, some revealed limitations, such as grouping based on environmental factors like lighting or image color rather than biological traits. These findings underscore the importance of integrating automated methods with expert ecological knowledge to refine classification outcomes. Furthermore, challenges such as overlapping clusters and the separation of day and night images for the same species highlight opportunities for methodological enhancements. This research offers a scalable framework for cleaning, preparing, and processing large image datasets, contributing to the monitoring and preservation of Yasuní's extraordinary biodiversity. It also provides a comprehensive data science approach for managing extensive image datasets. The curated dataset, clustering analyses, and visualizations serve as valuable tools for biologists to study species distribution and behavior while paving the way for future advancements in wildlife monitoring technology.**

***Index Terms*—Biodiversity Monitoring, Yasuní National Park, Wildlife Conservation, Convolutional Neural Networks (CNNs), Deep Learning, ResNet-152, t-SNE Visualization, Image Clustering, Camera Trap Analysis, Ecological Data Processing, Big Data in Ecology, Species Classification, Gaussian Mixture Models (GMMs), Environmental Machine Learning, Wildlife Image Analysis.**

I. Prior Works

The increasing volume of data stored in wildlife monitoring datasets has created new opportunities for utilizing advanced analytical methods to process information that would otherwise be unmanageable by a single individual. Recent advancements in wildlife monitoring have been significantly enhanced by the integration of artificial intelligence (AI) and camera trap technologies, which allow for more efficient data collection and analysis.

Henrich et al. [1] introduced a semi-automated camera trap distance sampling method for estimating population densities. Their approach combines AI with traditional wildlife monitoring techniques, addressing challenges such as detection probability and observation distances. This innovative framework provides robust methodologies for population estimation, exemplifying the potential of deep learning in ecological research. Similarly, Leorna and Brinkman [2] investigated the effectiveness of AI in identifying wildlife species from camera trap images. Their findings revealed that AI systems can match, and in some cases surpass, human capabilities under controlled conditions. Moreover, their study highlighted the strengths and limitations of using AI tools compared to manual human review for wildlife detection.

Velasco-Montero et al. [3] emphasized the importance of integrating AI with continual learning to enhance the reliability and accuracy of camera trap systems. Their research demonstrated significant improvements in inference accuracy, showcasing the potential of adaptive AI methodologies for long-term ecological studies.

Herraiz et al. [4] focused on the importance of rigorous experimental designs in studying animal interactions using camera traps. They stressed that camera trap placement is critical for obtaining reliable results, as spatial continuity is often lacking in such setups. Their study revealed that while camera trapping provides valuable insights into animal behavior, factors such as GPS accuracy can limit its effectiveness in detecting direct interactions.

In the realm of Internet of Things (IoT), Mamidi et al. [5] presented a cost-effective and scalable animal detection framework that combines embedded systems and interdisciplinary approaches. Their IoT-based system facilitates real-time wildlife monitoring, demonstrating its potential for widespread ecological applications.

Deep learning applications, particularly Convolutional Neural Networks (CNNs), have also played a pivotal role in advancing the field. Borude and Dand [6] leveraged deep learning architectures for animal recognition, achieving substantial improvements in accuracy. Meanwhile, for marine ecosystems, Yi et al. [7] developed a Coordinate-Aware Mask R-CNN model for underwater animal segmentation. Their work addressed the unique challenges of underwater environments, providing a novel approach to marine biodiversity monitoring.

Collectively, these studies illustrate the transformative

impact of AI, IoT, and deep learning on biodiversity monitoring. By enabling scalable, efficient, and accurate wildlife management, these technologies pave the way for actionable ecological insights and underscore the critical role of artificial intelligence in advancing environmental and wildlife conservation efforts.

II. Introduction

Yasuní National Park represents one of the most biologically diverse regions in Ecuador, characterized by an extraordinary concentration of species. As highlighted by Bass et al. (2010), Yasuní occupies a distinct biogeographical zone where species richness across four major taxonomic groups reaches its peak diversity [8]. However, human activities in this area pose serious threats to its ecological integrity. In efforts to mitigate potential ecological damage, the use of camera traps has proven to be a highly effective and reliable tool for monitoring wildlife in their natural environments [9]. In 1994, the Universidad San Francisco de Quito, in collaboration with Boston University, established the Tiputini Biodiversity Station within the Yasuní Forest. By 2009, a total of 43 cameras were deployed throughout the forest: 11 trail cameras, and 32 plot cameras. The trail cameras were motion-activated, capturing images of wildlife in response to movement. This paper focuses on the analysis and study of the images collected from these cameras.

Over the course of 15 years, 126 107 images were captured. Due to the sheer volume of data, zoologists have faced significant challenges in developing a study capable of efficiently classifying and identifying the animals present in these motion-triggered photographs. Given the ecological significance of the Yasuní forest and the importance of the collected data, this paper aims to process and curate the dataset by identifying images containing wildlife and organizing the new dataset into clusters using a Convolutional Neural Network, specifically ResNet-152, as a feature extractor. This approach will facilitate further analysis by biologists, aiding in the classification and monitoring of wildlife species.

This paper integrates techniques acquired during the Data Science Master's program, showcasing their practical application to real-world challenges in biodiversity conservation. The primary goal is to support nature preservation by leveraging deep learning techniques. Specifically, this study employs convolutional neural networks to cluster image datasets, uncovering patterns and insights that facilitate wildlife tracking and ecological research.

III. Materials and Methods

The main material of study was the image dataset from the camera traps. Due to the great amount of data found in the dataset, a high-speed processor was needed. It was decided to use the A100 NVIDIA GPU.

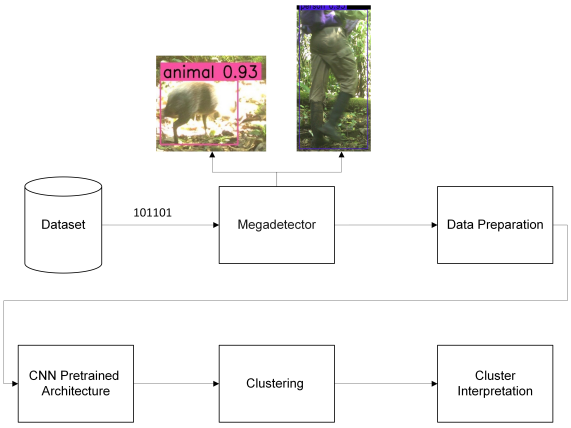

Figure 1. Data Pipeline.

1) Understanding the Dataset: The dataset was stored on a server provided by the Universidad San Francisco de Quito and was organized into folders separated by year, ranging from 2004 to 2019. It comprises 126,107 images in '.jpg' format captured by 43 camera traps across the Tiputini research area. The images include a mix of different types:

- Images containing animals
- Images containing people
- False positives (vegetation triggered by movement but containing neither animals nor people)

These images were captured both during the day and at night. Additionally, some animals were recorded in multiple frames, resulting in multiple images for a single instance of movement.

To better understand the dataset, the metadata for each image was stored in a JSON file. This file saved crucial information such as the time, GPS location, file location, format, SHA identification, and any additional details. Utilizing JSON allowed for efficient data organization and retrieval, which is essential for managing large-scale datasets. As Yusof and Man noted, JSON's lightweight format and ability to handle high-throughput queries without sacrificing scalability make it particularly well-suited for such applications [10]. This process was critical for preparing the dataset for cleaning and ensuring accurate clustering.

2) Cleaning the Dataset: The first step in cleaning the dataset was to separate the images, retaining only those with animals, while eliminating images of people and false positives. A SHA filter was applied to remove any duplicate images found in the dataset.

The PyTorch Wildlife Megadetector, was used for this process, a pre-trained machine learning model developed by Microsoft [11], has become a pivotal tool for processing large datasets, significantly reducing the manual effort required for analyzing camera trap data. As Beery et al. [11] note, "The MegaDetector provides robust and geographically generalizable animal detection in camera

trap data, drastically reducing data processing time and costs, sometimes by up to 90 percent." This capability is especially important given the challenges posed by vast datasets and diverse environmental conditions.

We processed the metadata in the JSON file, leveraging the SHA identification code to locate corresponding images on the server. Each image was analyzed by the neural network, with results stored in a JSONL file, preserving characteristics identified by the model. Using the JSONL format, where each entry is stored on a new line, proved advantageous for appending data without reading the entire file. Only images where animals were detected with a probability of 0.8 or higher were included, with bounding boxes recorded for the identified animals. Celis et al. [12] emphasize that workflows like this, "integrating MegaDetector with site-specific enhancements, can achieve animal detection rates exceeding 90 percent, even in challenging Arctic environments." An algorithm processed the metadata found in the JSON file, using the SHA identification code to locate the corresponding image on the server. Each image in the dataset was analyzed by the neural network, and the results were stored in a JSONL file, which preserved specific characteristics identified by the neural network. The following image illustrates an example of a discarded image where a human was detected. It also highlights the bounding box used to identify the subject in the image.

Figure 2. Discarded image from the dataset.

The JSONL format, which stores each entry on a new line, was beneficial for appending data without the need to read the entire file. Only images where animals were detected with a probability of 0.8 or higher were included in the JSONL file. Additionally, the bounding boxes indicating where animals were located within the images were recorded.

This selective data capture was crucial for reducing noise in the dataset, which could otherwise hinder the clustering process. Ultimately, this procedure enabled us to consolidate the data extracted from the Megadetector with the metadata from the images, resulting in a comprehensive JSONL file. As a result, 97,406 images were retained for processing.

3) Data Preparation: Once the dataset was cleaned, it was necessary to process all the data to feed the Convolutional Neural Network (CNN). This preparation ensured that the input data was compatible with the model's architecture and maximized the quality of the extracted features.
4) Initialization of the Convolutional Neural Network: We employed the ResNet-152 model, a deep convolutional neural network pre-trained on the ImageNet dataset. As noted by Athisayamani et al. [13], ResNet-152 has demonstrated exceptional capability in extracting deep-level visual features, particularly in biomedical imaging tasks where precise segmentation and classification are critical. The model was adapted for our purposes by removing the final classification layer, effectively converting it into a feature extractor. This modification allowed us to leverage the model's strength in capturing intricate visual patterns without engaging in specific classification tasks.

The model was set to evaluation mode, and all parameters were frozen to ensure no updates occurred during the feature extraction process. This step was important for stabilizing the model's parameters and maintaining the integrity of the learned features.
5) Image Transformation and Standardization: Prior to feature extraction, each image underwent a series of transformations, including:

- Resizing the image to a fixed dimension of 224×224 pixels
- Converting the image to a tensor
- Normalizing the image based on the mean and standard deviation values derived from the ImageNet dataset

These steps standardized the input data and adjusted it to the optimal format required by the ResNet-152 model, facilitating consistent feature extraction across all images.

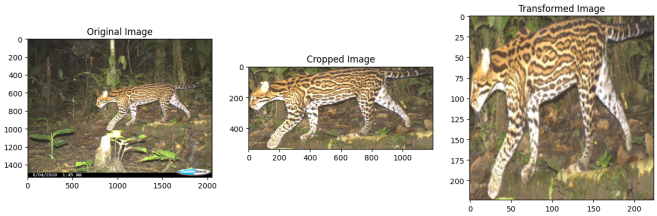

Figure 3. Image Transformation Process.

6) Cropping Mechanism: A critical component of our preprocessing pipeline involved cropping the images based on bounding boxes identified by an external detection algorithm. The cropping function adjusted the bounding boxes to form a square region centered around the detected object, ensuring uniformity in the size and scale of the regions being analyzed. This uniformity was vital for minimizing variations in input data that could potentially influence the performance of the CNN.
7) Feature Extraction Routine: Each preprocessed image was then fed into the ResNet-152 model to extract a dense feature vector. The model, operating in forward-pass

mode, processed the image and outputted a feature map that was subsequently flattened into a vector. This vector encapsulated the essential characteristics of the image, which were crucial for the subsequent clustering analysis.

8) Data Storage and Management: The extracted features, along with relevant metadata, were stored in a JSONL file format. This format facilitated efficient data manipulation and storage, allowing for incremental writes without the need to reload the entire dataset. This capability was particularly useful for handling large volumes of data, significantly reducing computational overhead and enhancing the scalability of the data processing pipeline.

In conclusion, the preparation of the dataset through these systematic and methodical steps was fundamental to ensuring that the CNN could perform optimally. This prepared dataset served as the foundation for the next stages of our analysis, where clustering algorithms classified the images based on the extracted features, grouping together images with similar visual traits.

IV. Results and Discussion

After processing the data through the CNN, we stored the features from each image in a JSONL file. This approach helped organize the information in a manner that is easily understandable for the clustering process. We extracted and stored 2048 features from 97,406 images. Subsequently, we proceeded to conduct silhouette and elbow analyses to determine the optimal number of clusters for this dataset. These methods are commonly used to evaluate clustering performance by providing insights into cluster cohesion and separation [14]. The following images illustrate the results of these analyses.

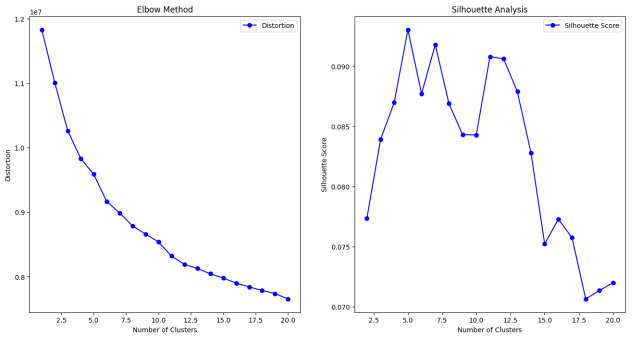

Figure 4. Elbow method, and Silhouette Score.

Upon analyzing the data, it was observed that five clusters might initially seem optimal. However, considering the known diversity of animal species in the area, this number of clusters appeared to be insufficient. Therefore, based on the analysis we opted for a more accurate representation, which supported eleven clusters as the most suitable configuration for our dataset.

After determining an optimal cluster configuration, employing t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization proved to be highly advantageous. By reducing high-dimensional data to a more interpretable two-dimensional space while preserving the local structure of the data, we were able to gain a better understanding of the data distribution and assess the appropriateness of selecting 11 clusters. The t-SNE visualization was a critical step in ensuring that the clustering results were robust, understandable, and actionable.

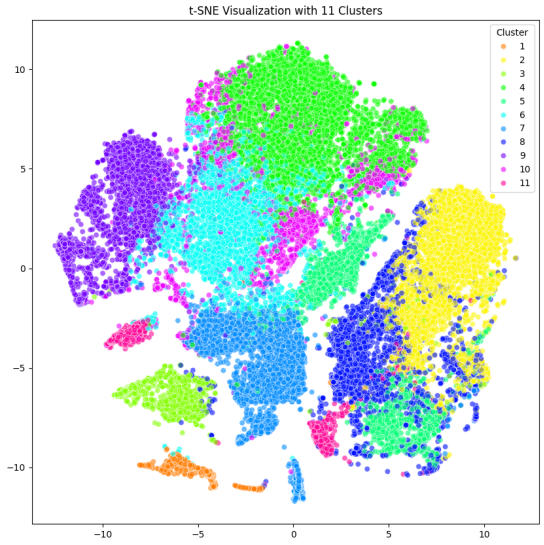

Figure 5. t-SNE Visualization.

As seen in the image, the t-SNE visualization distinctly delineates the 11 clusters, each differentiated by color. Each point represents an element from the dataset, and each color corresponds to a cluster. This visualization supports our decision to select eleven as the optimal number of clusters, as it is evident that the groups are distinguishable and well-distributed across the image. However, there are instances where elements from one cluster appear within another, suggesting interspersion. This mixing of different colors can be attributed to several factors: the inherent complexity and potential overlap of data features, the influence of t-SNE parameters such as perplexity, which balances the focus between local and global aspects of the data, and the effects of noise and outliers.

Following the insights from the t-SNE visualization, we chose to explore the structure of our dataset further using Gaussian Mixture Model (GMM) analysis. The t-SNE visualization had clearly supported our decision to use eleven clusters, showing that each cluster was distinct and well-spread across the image. However, the occasional overlap of elements between clusters, indicated that a more detailed analysis might be beneficial.

Using GMM allowed us to take a deeper look at how data points fit into each cluster, giving us a probabilistic view rather than a simple yes-or-no answer. This method helped us understand the subtle differences and similarities between clusters better, confirming the strength of our initial clustering and providing more detailed insights.

Additionally, by pairing the GMM analysis with visualizations that included representative images for each cluster, we were able to make our findings clearer and more accessible. Combining the visual evidence from both the t-SNE and GMM analyses enriched our interpretation, linking numbers and images together to give a fuller picture of the dataset. This approach enhances our ability to obtain reliable scientific conclusions and supports further research.

1) Cluster Description: The images described in this section can be found in Appendix A.

Cluster 1: This cluster primarily showcases images characterized by the frequent appearance of leopards and similar species, different species of spotted felines that belong to the *FELIDAE* family captured in various poses such as walking, prowling, or captured mid-motion by camera traps.

Cluster 2: This cluster predominantly features a variety of mammals such as *Tapirus terrestris* and many animals that belong to the *Tayassuidae* family, captured incidentally during nighttime. These photographs show these animals with significant amounts of fur. While the images were taken at night, it is important to note that the appearance of these mammals during these hours does not necessarily indicate nocturnal behavior; rather, it highlights the opportunistic nature of wildlife photography, which captures moments of activity regardless of the time of day.

Cluster 3: This cluster showcases a collection of large, ground-dwelling birds, specifically identified as *Psophia crepitans*, or trumpeters. These birds are characterized by their robust build and distinctive plumage, which ranges from solid and sleek to speckled, aiding their camouflage. Notable features captured in these images include their prominent tail feathers and strong, sturdy legs, clearly demonstrating their adaptability to ground movement.

Cluster 4: This cluster captures the same animal species as Cluster 2, but with the distinctive difference that these images were taken during daylight. The vivid daytime photographs enhance the visibility of physical features, including their coarse, bristly fur and robust bodies. Unlike the nocturnal images of Cluster 2, these daylight images also offer clearer views of the animals' coloring and patterns.

Cluster 5: We found that this cluster was characterized by the unique feature of smooth fur across various animal species, distinguishing these individuals from those in previous clusters. While this cluster includes repetitions of animals from the *Tapiridae* family, they do not dominate the grouping. Instead, the shared trait of smooth fur provides a distinct texture and appearance compared to the coarse or bristly fur seen in previous clusters. This cluster captures a variety of species in both color and grayscale images, highlighting the sleek fur that enhances their streamlined body shapes, which is notable in their natural settings.

Cluster 6: Cluster 6 predominantly features images with a notable green hue, a result of either dense vegetation in the background or a green filter effect on the photographs. The collection does not display a uniform feature among the animals depicted, suggesting that the primary criterion for grouping these images together was the color scheme rather than specific animal characteristics. The green tint permeates each image, impacting the visibility and perception of the animals, which range from mammals to birds, caught in various natural settings. Given the emphasis on color over distinguishable animal features, this cluster might be deemed less relevant and potentially excluded for purposes of focused animal studies.

Cluster 7: This cluster primarily features the *Mazama americana*, commonly known as the red brocket deer, a species belonging to the *Cervidae* family. These animals are characterized by their reddish-brown fur, slender build, and relatively small antlers in males, which are visible in several images. The clustering also included some animals with physical similarities to the *Mazama americana*, such as comparable body shapes or postures, leading to slight overlaps in classification. Additionally, the reddish-brown fur of this species resulted in the inclusion of other animals that, due to sunlight reflection or lighting conditions, displayed a similar color tone in the images. Despite these overlaps, this cluster provides a detailed collection showcasing the distinct physical traits of *Mazama americana*.

Cluster 8: This cluster presents a diverse mix of animals, with no clear dominance of a specific species. The images feature a variety of body parts, such as close-up views of fur, limbs, or tails, as well as full-body shots. Some images include blurred or unclear features, likely due to motion or low visibility. The clustering appears to be influenced more by random variations in photographic features, such as image texture or brightness. This cluster lacks a definitive unifying trait across its elements, which may reduce its relevance for targeted species identification or behavioral studies.

Cluster 9: This group features the same animal species as Cluster 3, mainly *Psophia crepitans*. However, unlike Cluster 3, which consists of nighttime images, this cluster captures these birds during daylight. The improved lighting conditions helped to provide greater clarity of their physical characteristics, such as their plumage details and body structure, offering a complementary perspective to the nocturnal images in Cluster 3.

Cluster 10: This cluster primarily showcases a variety of rodents, characterized by their small to medium-sized bodies, short legs, and prominent incisors. The images highlight species that look alike, such as agoutis, recognizable by their smooth fur and rounded body shapes. Additionally, this cluster includes a few non-rodent species, such as small mammals with similar ecological niches or physical appearances, including some resembling armadillos or small carnivores. The inclusion of these non-rodent animals may result from overlapping features such as size. This cluster offers a diverse representation of ground-dwelling mammals, emphasizing their shared traits.

Cluster 11: This cluster primarily features armadillos, with most individuals likely belonging to the species *Dasypus novemcinctus* (nine-banded armadillo) and *Priodontes max-*

imus (giant armadillo). These animals are easily recognized by their armored bodies, consisting of bony plates that form a protective shell, and their long, tapered tails. The images highlight their distinctive body structure, including their elongated snouts and strong, clawed limbs adapted for digging. The size variation among individuals supports the inclusion of both species, with *Priodontes maximus* being significantly larger. This cluster provides a clear representation of these unique mammals and is also one of the most accurate clusters for grouping animal features.

V. Conclusion

This article demonstrates the successful application of Convolutional Neural Networks (CNNs) to cluster wildlife images from Yasuní National Park, one of the most biologically diverse regions in Ecuador. It also highlights the process of cleaning and preparing a large dataset of images before applying deep learning techniques. By utilizing the ResNet-152 model for feature extraction, combined with methods such as t-SNE visualizations and Gaussian Mixture Model analyses, we successfully identified and grouped over 97,000 images into 11 distinct clusters. These clusters provided valuable insights into the biodiversity of the region, capturing a range of species from the FELIDAE family and ground-dwelling birds to various rodents, armadillos, and mammals from the Tapiridae and Cervidae families.

The results highlight the potential of automated image processing to address challenges associated with large datasets, particularly in scenarios where monitoring biodiversity is critical for conservation efforts. While certain clusters, such as those featuring spotted felines or armadillos, demonstrated a high degree of accuracy and ecological relevance, other clusters grouped images based on factors like color or unclear animal characteristics rather than similar species. This revealed limitations in clustering precision due to environmental factors such as lighting conditions and motion blur. These findings underscore the importance of combining automated methods with ecological expertise to refine data processing pipelines and improve classification outcomes. It is highly recommended that future studies include the active participation of a biologist to better identify the specific animals represented in each cluster. Additionally, some clusters were found to represent the same species but were separated based on whether the images were taken during the day or at night.

This study not only provides a scalable and efficient approach for processing image datasets but also emphasizes the importance of integrating advanced technologies with field research. The curated dataset and detailed cluster analyses serve as a foundation for future biodiversity monitoring, enabling scientists to identify patterns, track species distributions, and evaluate the impact of human activities in Yasuní. Moving forward, integrating additional contextual data and refining clustering algorithms will enhance the accuracy and applicability of such methodologies, contributing to the global effort to protect and preserve biodiversity.

Appendix A
Representative Images per Cluster

The most representative image for each cluster is presented, as explained in the cluster description.

Figure 6. Representative images for each cluster.

Acknowledgment

The authors would like to express their gratitude to Andrés Cajamarca for his active participation during the initial stages of this project. We also extend our thanks to Daniel Riofrío, and Felipe Grijalva for providing valuable advice and insights throughout the investigation.

References

[1] M. Henrich, M. Burgueño, J. Hoyer, T. Haucke, V. Steinhage, H. S. Kühl, and M. Heurich, "A semi-automated camera trap distance sampling approach for population density estimation," *Remote Sensing in Ecology and Conservation*, vol. 10, no. 2, pp. 156–171, 2024.

[2] S. Leorna and T. Brinkman, "Human vs. machine: Detecting wildlife in camera trap images," *Ecological Informatics*, vol. 72, p. 101876, 2022.

[3] D. Velasco-Montero, J. Fernández-Berni, R. Carmona-Galán, A. Sanglas, and F. Palomares, "Reliable and efficient integration of ai into camera traps for smart wildlife monitoring based on continual learning," *Ecological Informatics*, vol. 83, p. 102815, 2024.

[4] C. Herraiz, D. Ferrer Ferrando, J. Vicente, and P. Acevedo, "Camera trapping and telemetry for detecting and quantifying animal interactions: Not anything goes," *Ecological Indicators*, vol. 160, p. 111877, 2024.

[5] K. K. Mamidi, S. N. Valiveti, G. C. Vutukuri, A. K. Dhuda, H. Alabdeli, R. Chandrashekar, S. Lakhanpal, and P. Praveen, "An iot-based animal detection system using an interdisciplinary approach," in *E3S Web of Conferences*, vol. 507. EDP Sciences, 2024, p. 01041.

[6] V. Borude and H. Dand, "Animal recognition using deep learning architecture," *Asian Journal of Biological Sciences*, vol. 6, pp. 916–921, 2024.

[7] D. Yi, H. B. Ahmedov, S. Jiang, Y. Li, S. J. Flinn, and P. G. Fernandes, "Coordinate-aware mask r-cnn with group normalization: A underwater marine animal instance segmentation framework," *Neurocomputing*, vol. 583, p. 127488, 2024.

[8] M. S. Bass, M. Finer, C. N. Jenkins, H. Kreft, D. F. Cisneros-Heredia *et al.*, "Global conservation significance of ecuador's yasuní national park," *PLOS ONE*, vol. 5, no. 1, p. e8767, 2010. [Online]. Available: https://doi.org/10.1371/journal.pone.0008767

[9] H. Nguyen, S. J. Maclagan, T. D. Nguyen, T. Nguyen, P. Flemons, K. Andrews, E. G. Ritchie, and D. Phung, "Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring," in *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2017, pp. 40–49.

[10] M. K. Yusof and M. Man, "Efficiency of json for data retrieval in big data," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 7, no. 1, pp. 250–262, 2017.

[11] S. Beery, "The megadetector: Large-scale deployment of computer vision for conservation and biodiversity monitoring," in *Proceedings of the Conference on Biodiversity Monitoring*. Microsoft AI for Earth, 2024. [Online]. Available: https://github.com/microsoft/CameraTraps

[12] G. Celis, P. Ungar, A. Sokolov, N. Sokolova, H. Böhner, D. Liu, O. Gilg, I. Fufachev, O. Pokrovskaya, R. A. Ims, W. Zhou, D. Morris, and D. Ehrich, "A versatile, semi-automated image analysis workflow for time-lapse camera trap image classification," *Ecological Informatics*, vol. 81, p. 102578, 2024. [Online]. Available: https://doi.org/10.1016/j.ecoinf.2024.102578

[13] S. Athisayamani, R. S. Antonyswamy, V. Sarveshwaran, M. Almeshari, Y. Alzamil, and V. Ravi, "Feature extraction using a residual deep convolutional neural network (resnet-152) and optimized feature dimension reduction for mri brain tumor classification," *Diagnostics*, vol. 13, no. 4, p. 668, 2023.

[14] D. M. Saputra, D. Saputra, and L. D. Oswari, "Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method," in *Sriwijaya international conference on information technology and its applications (SICONIAN 2019)*. Atlantis Press, 2020, pp. 341–346.