Wildlife Image Processing & Semantic Search System

Matt Scardino
Anderson College of Business and Computing
Regis University
Regis University, Denver, CO, USA
scard659@regis.edu

Abstract

This project presents the development of a modular AI-enhanced system to process, classify, and retrieve wildlife images and videos. The system integrates traditional computer vision techniques with advanced semantic understanding powered by multimodal AI models, including vision–language embeddings that enable transferable visual representations across domains [1]. The platform supports both manual and AI-assisted annotation, preserves visual metadata and learned embeddings, and enables intuitive natural language queries for content discovery. By incorporating semantic similarity search alongside ecological context informed by prior wildlife machine learning research [2], the system transforms unstructured wildlife media into structured, searchable information and enables more meaningful exploration and utilization of wildlife imagery.

I. Introduction/Background

The proliferation of wildlife imagery from researchers, conservationists, and photographers has created a need for intelligent systems that extend beyond basic storage and retrieval. This project envisions an AI-enhanced platform explicitly tailored for wildlife media, offering deeper analytical capability and more domain-specific functionality than general-purpose photo applications. By combining image processing techniques with vision–language models, the system supports semantic search, behavior analysis, and the extraction of ecological information from both personal and professional wildlife image collections. When augmented with geospatial tools, such systems enable users to uncover patterns related to migration, seasonal behavior, and habitat use, contributing to ecological research and conservation efforts.

Prior research has demonstrated that wildlife imagery presents unique technical challenges compared to general-purpose visual datasets, particularly due to environmental variability, background complexity, and fine-grained interspecies similarity [3]. At the same time, advances in deep learning for ecological monitoring have shown that automated species identification at scale is feasible and scientifically valuable [2]. These findings motivate the development of integrated systems that unify computer vision, semantic embeddings, and geospatial reasoning to better align automated analysis with ecological realities.

II. Problem Statement

Traditional photo management tools such as Apple Photos or Google Photos are designed to organize personal memories rather than to support ecological analysis. These systems lack domain-specific capability to interpret complex wildlife imagery and cannot reliably analyze species identity, behavior, habitat conditions, or patterns across space and time. As a result, they are poorly suited for ecological workflows that require contextual reasoning and longitudinal analysis.

For researchers, conservationists, and wildlife photographers, this limitation presents a significant gap. Large volumes of wildlife imagery remain disconnected from structured ecological knowledge that could inform species management, habitat protection, or human–wildlife conflict mitigation. Prior ecological research emphasizes that meaningful interpretation of wildlife observations depends on integrating visual evidence with spatial and temporal context, rather than treating images as isolated artifacts [4].

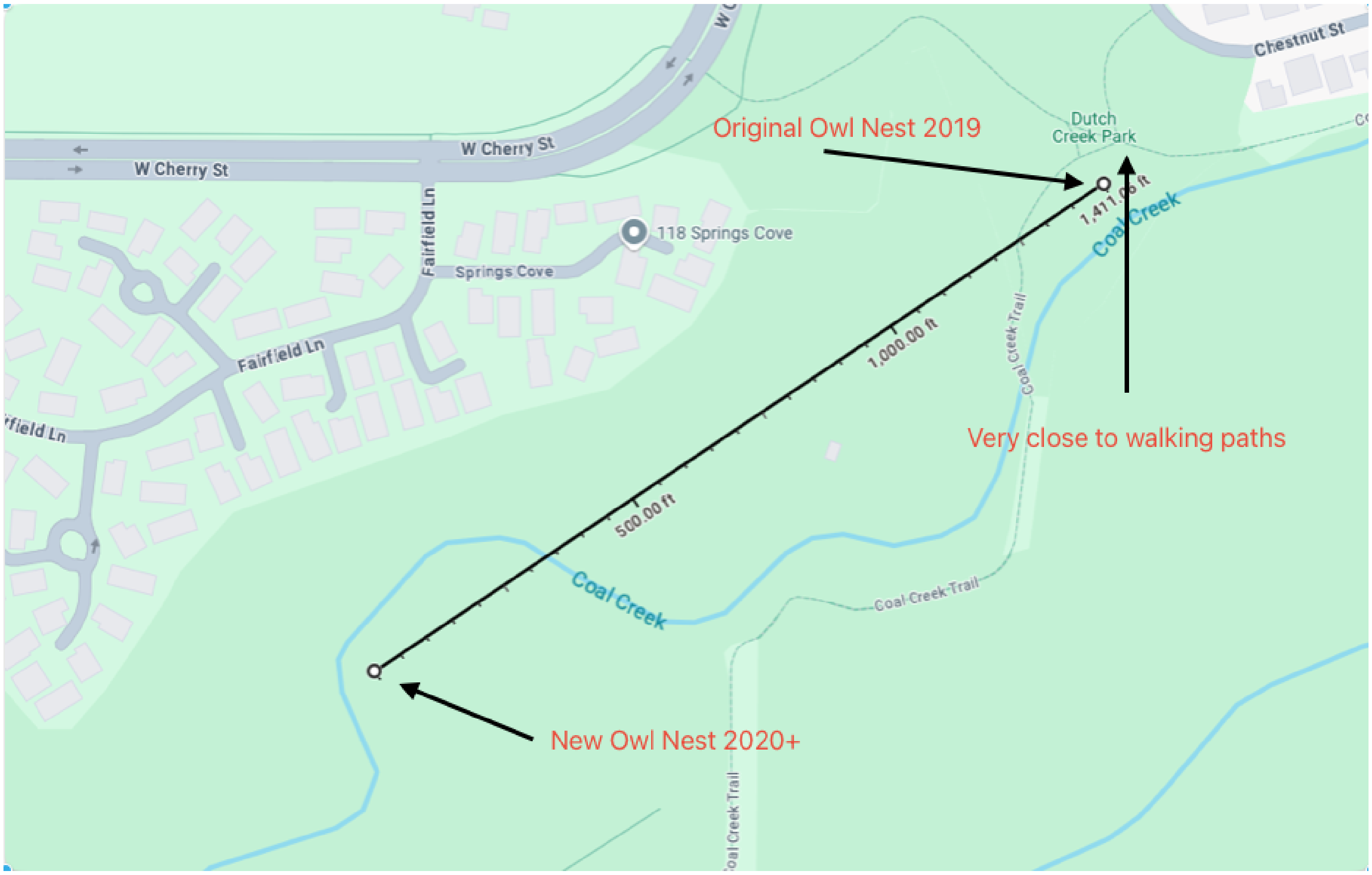

Fig. 1. Great Horned Owl Nesting Patterns

This project addresses that gap by developing a specialized platform that transforms wildlife images into structured ecological data, supporting species classification, behavior tagging, semantic search, and spatial analysis. Understanding wildlife responses to human activity, for example, requires both spatial awareness and temporal tracking—capabilities that general-purpose tools cannot provide. Figure 1 illustrates this need through a case study involving a pair of Great Horned Owls that relocated their nesting site following repeated human disturbance at Coal Creek Trail in Louisville, Colorado.

In 2019, the owls nested near Dutch Creek Park, directly adjacent to walking paths. After repeated human presence, they moved approximately 1,400 feet downstream for subsequent breeding seasons. Detecting, tracking, and analyzing changes of this nature requires a purpose-built system capable of integrating imagery, geospatial metadata, and AI-driven analysis. This project responds to that requirement by enabling unstructured wildlife media to be analyzed as actionable ecological insight.

III. Methodology/Approach

The architecture for this project is built on two core pillars: a Python processing pipeline and a PostgreSQL database with spatial and vector search capabilities. Python provides the foundation for ingesting, processing, and analyzing wildlife images, leveraging libraries such as Streamlit for user interaction, OpenCV for image handling, BeautifulSoup for scraping reference images, and OpenAI's CLIP model for generating joint image and text embeddings [1]. Object detection and automated cropping are performed using a YOLO-based architecture, which provides efficient and accurate localization of visual subjects in complex scenes [5]. Isolating wildlife subjects prior to embedding generation improves downstream classification and similarity search quality. PostgreSQL, extended with PostGIS for spatial queries and PGVector for similarity search, manages structured ecological data, species embeddings, and metadata.

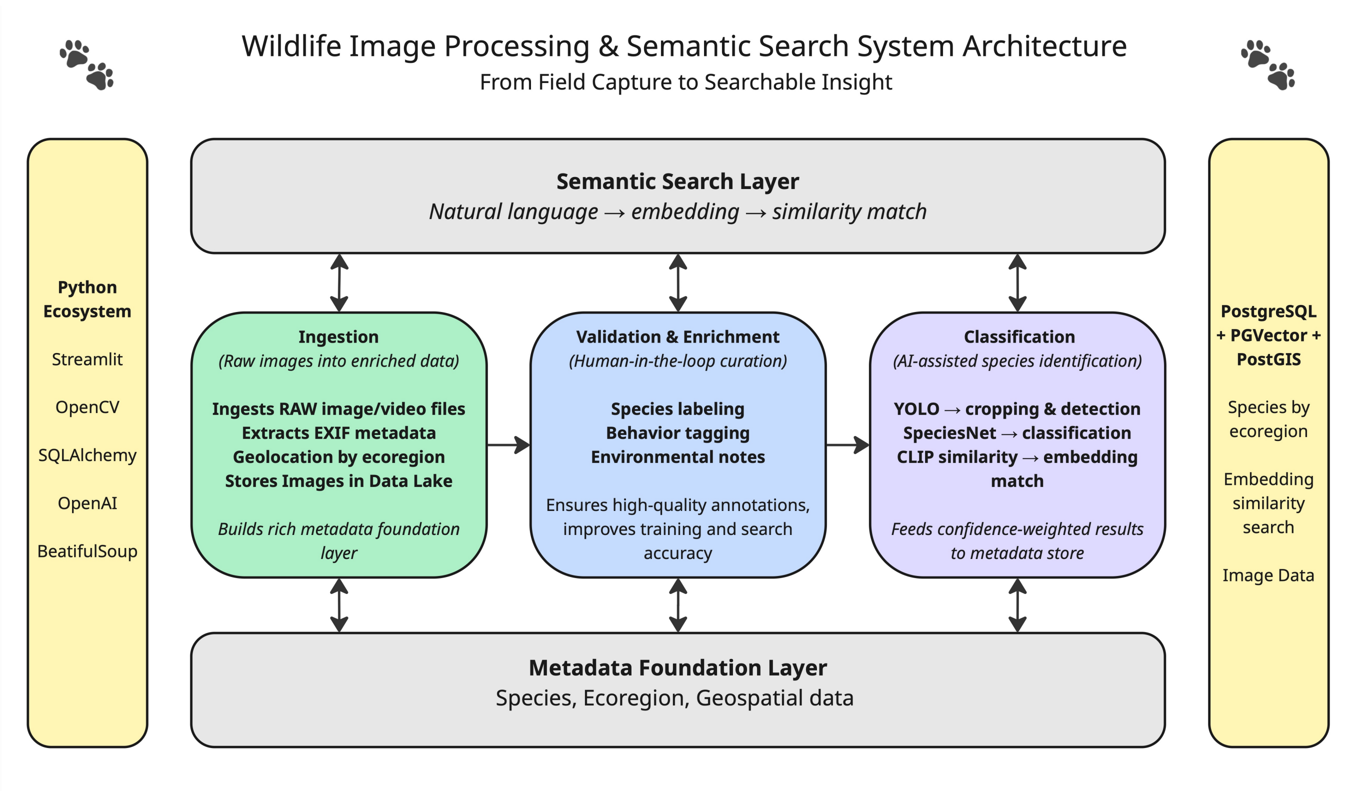

Fig. 2. Wildlife Image Processing and Semantic Search System Architecture

This architecture, shown in Figure 2, illustrates the modular workflow from image ingestion to searchable ecological insight. The system integrates species detection, embedding-based similarity search, human-in-the-loop validation, and ecological filtering based on ecoregions.

The foundation of the system is the Metadata Layer, which integrates species information, ecoregion boundaries, and geospatial data. This layer is critical for grounding AI predictions within ecological reality and directly influences the accuracy of species identification across the pipeline.

The dataset consists of a curated subset of personal wildlife photography captured over several years across diverse geographic locations. These high-resolution RAW images document species and behaviors in natural settings and include embedded EXIF metadata such as capture date, time, GPS coordinates (when available), and camera settings [6]. To expand taxonomic coverage, the system scrapes publicly available species images from Wikipedia using BeautifulSoup, focusing on mammals and birds. These reference images are processed to generate a baseline set of canonical embeddings, which serve as a comparison model for similarity search and classification.

A core strength of the system is its integration of ecological context directly into species prediction by geocoding image locations and constraining candidate species to those known to occur within the corresponding ecoregion. Species range data and ecoregion boundaries from the World Wildlife Fund's WildFinder database [7] provide the basis for this process. Incorporating spatial constraints improves prediction reliability by eliminating biologically implausible results and prioritizing species based on realistic geographic distributions, consistent with established ecological research [4].

The system combines this spatial awareness with CLIP-based semantic embeddings [1], enabling flexible similarity search and clustering across both images and species concepts. While vision–language embeddings demonstrate strong transfer learning capabilities, prior research has shown that fine-grained visual categories often cluster densely in embedding space, particularly in wildlife imagery characterized by subtle interspecies variation [3]. As a result, embedding similarity alone may be insufficient for precise species identification.

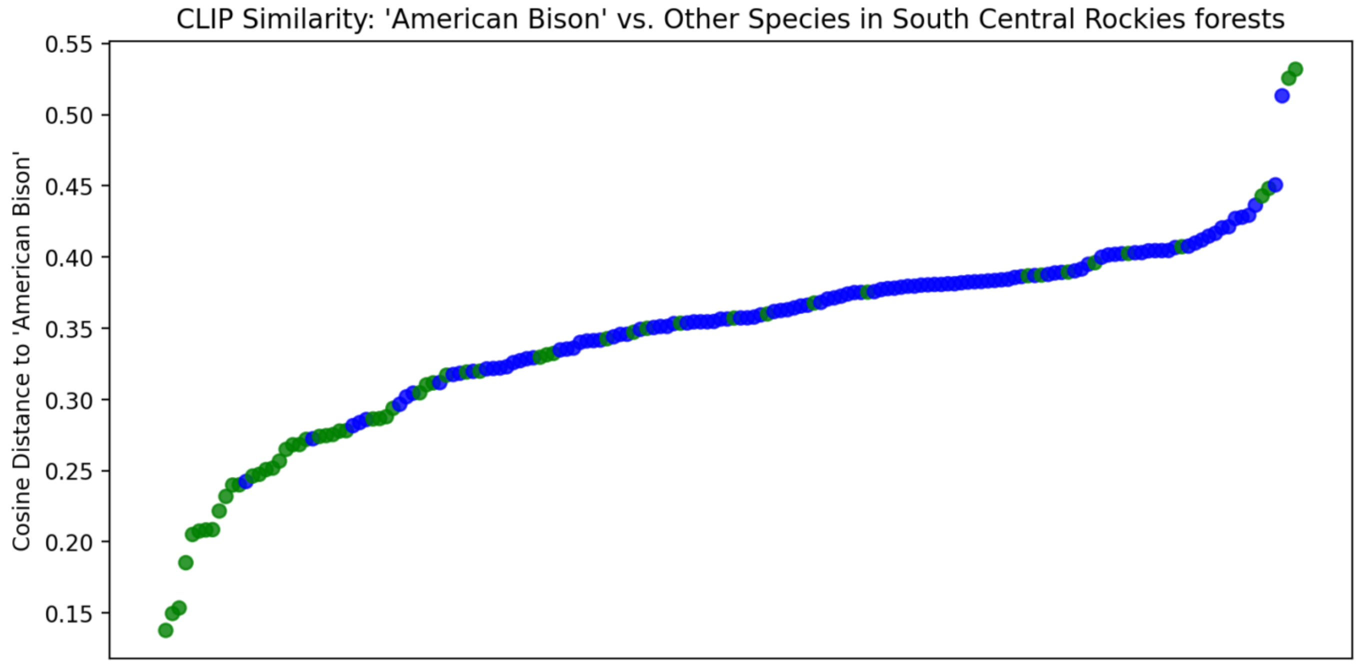

Fig. 3. Cosine distance between the CLIP [1] embedding for *American Bison* and other species within the South Central Rockies forests ecoregion. Dense clustering and compressed distance ranges highlight the challenge of interpreting raw embedding similarity scores for fine-grained species identification.

This limitation is illustrated by comparing the CLIP-generated embedding for *American Bison* against other species within the South Central Rockies forests ecoregion. As shown in Figure 3, cosine distance values occupy a narrow range, with many species overlapping in vector space. Although lower distances generally indicate greater similarity, the compressed scale complicates confident species differentiation. These observations reinforce the need to integrate ecological constraints and supervised classification rather than relying solely on embedding proximity.

Classification performance is further improved through a supervised wildlife model, SpeciesNet, which is trained specifically for species-level recognition in ecological imagery [8]. Prior work has demonstrated that subject-focused inputs and domain-specific supervision significantly improve classification accuracy in camera trap and wildlife datasets. Together, these components form a modular and scalable pipeline that converts raw imagery into structured ecological insight, supported by validation workflows, human-in-the-loop enrichment, and semantic search tools.

IV. Data Description

This project is designed primarily for personal use; however, broader application of the system raises important considerations related to data governance, intellectual property, and responsible research practices. Wildlife photographers, in particular, have valid concerns regarding image ownership and intended use. The system preserves original file structures, metadata, and attribution, and any expansion beyond personal datasets would require explicit consent policies, access controls, and safeguards to respect photographer rights. Additionally, the system processes location data for species detection, which introduces potential risk if sensitive sites such as nesting or denning locations are inadvertently exposed. Public-facing implementations would therefore require location masking or spatial generalization to protect vulnerable wildlife from human disturbance.

The system architecture follows best practices for data management and reproducibility, with structured storage of images, metadata, and AI-derived embeddings in a PostgreSQL database, supported by version control for schema evolution and model updates. AI-driven species identification also introduces risks of algorithmic bias, particularly when training data overrepresent common species, specific habitats, or visual conditions. Prior research in wildlife machine learning highlights that rare species, seasonal variation, and environmental heterogeneity can significantly degrade model generalization and reliability [3]. To mitigate these limitations, the system combines supervised classification with human-in-the-loop validation and communicates uncertainty through similarity-based ranking rather than deterministic predictions. Scaling the system to broader datasets or user populations would require additional safeguards, including transparency around AI limitations, secure handling of sensitive data, and ongoing monitoring to prevent misuse or unintended ecological impact.

V. Expected Outcomes

This project will produce a fully functional prototype system that transforms unstructured wildlife imagery into structured ecological insight through AI-driven classification, semantic search, and spatial filtering. The key technical deliverables include an integrated species detection pipeline, a canonical embedding model for known species, and interactive tools for geospatial analysis, validation, and semantic similarity search. Together, these components form a unified platform that organizes wildlife images, improves species identification accuracy, and enables pattern discovery grounded in ecological context.

The system demonstrates how combining vision–language embeddings generated by CLIP [1], supervised species classification, and ecoregion-based filtering improves the reliability of species predictions, particularly in cases where visually similar species overlap. Prior ecological machine learning research has shown that constraining predictions using spatial and environmental context significantly enhances classification robustness in wildlife datasets [2]. Applied examples, such as tracking owl nest relocations or comparing species similarity within an ecoregion, illustrate the practical value of the system for researchers, conservationists, and wildlife photographers working with large, unstructured image collections.

While the system addresses core challenges outlined in the problem statement, it represents an early-stage prototype with known limitations. AI models may struggle with rare species, atypical image conditions, or incomplete metadata. Future work is required to expand training data, incorporate additional ecological layers, and refine outputs for large-scale or public deployment. Nonetheless, this project establishes a foundation for advancing applied machine learning in ecological monitoring and demonstrates a reproducible framework for integrating AI, spatial data, and domain expertise in support of wildlife research and conservation.

VI. Timeline

Week	Focus	Deliverables
Week 1	Project Setup & Planning	Define scope, set up Git repository, document architecture, configure PostgreSQL with pgvector extension, and install required Python libraries (OpenCV, Streamlit, OpenAI, etc.)
Week 2	Image/Video Ingestion Pipeline	Build Python pipeline to ingest RAW image and video files, extract EXIF [6] metadata, convert RAW to JPG, and extract frames using OpenCV; organize media by timestamp in a structured directory tree
Week 3	Streamlit UI for Upload & Metadata	Develop Streamlit interface for uploading images, viewing thumbnails, manually entering species/location/behavior metadata, and staging files for import
Week 4	Embedding Generation & Storage	Integrate OpenAI Embedding API (‘text-embedding-ada-002‘) to generate text embeddings from image captions; store embeddings and related metadata in PostgreSQL with pgvector for semantic indexing
Week 5	Natural Language Search	Create a natural language query interface using Streamlit; convert queries to embeddings and perform similarity search against stored vectors to retrieve and display relevant images
Week 6	Annotation Tools & Data QA	Build tools for editing and validating annotations, display metadata coverage stats, handle missing EXIF [6] data, and support batch annotation
Week 7	Evaluation & Visualization	Use dimensionality reduction techniques (e.g., t-SNE, UMAP) to visualize the semantic structure of the embedding space; analyze tag consistency and data quality
Week 8	Final Presentation & Demo	Prepare slide deck, demo the full ingestion-to-search pipeline, showcase example queries and visualizations, and summarize key technical and research findings

TABLE I
PROJECT TIMELINE FOR WEEKS 1–8

VII. Conclusion

This project applies advanced data science techniques to a real-world ecological challenge: transforming unstructured wildlife imagery into structured, actionable insight. Existing tools fail to bridge the gap between raw images, species identification, and ecological understanding, particularly for researchers, conservationists, and photographers working with large, unorganized image collections. The system presented here addresses that gap by combining AI-driven classification, semantic embeddings, and spatial filtering informed by known species distributions and ecoregions.

The project integrates machine learning models, geospatial data, and human-in-the-loop validation within a scalable and modular pipeline, providing a reproducible framework for extracting ecological knowledge from imagery. Beyond its technical contributions, this work demonstrates how AI systems can be grounded in ecological context to produce reliable and interpretable outputs, a challenge that lies at the core of applied data science in environmental domains.

As the culmination of the data science curriculum, this project reflects both technical proficiency and domain-specific understanding. It brings together machine learning, database design, geospatial analysis, and responsible data management into a coherent, purpose-built system. The educational value extends beyond the application of individual techniques, demonstrating the ability to design and implement an end-to-end solution with practical relevance for ecological research and wildlife monitoring, and illustrating readiness for professional practice in applied data science.

References

[1] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal *et al.*, “Learning transferable visual models from natural language supervision,” in *Proceedings of the 38th International Conference on Machine Learning (ICML).* PMLR, 2021. [Online]. Available: https://openai.com/research/clip

[2] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, “Automatically identifying animals in camera trap images using deep learning,” *Nature Communications*, vol. 9, no. 1, 2018.

[3] S. Beery, G. Van Horn, and P. Perona, “Recognition in terra incognita,” in *Proceedings of the European Conference on Computer Vision (ECCV).* Springer, 2018.

[4] W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers, “The global diversity of birds in space and time,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 28, pp. 11 061–11 066, 2012.

[5] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.

[6] P. Harvey, “Exiftool,” 2024, online; accessed June 25, 2025. [Online]. Available: https://exiftool.org

[7] World Wildlife Fund, “Wildfinder database,” online; accessed June 25, 2025. [Online]. Available: https://www.worldwildlife.org/publications/wildfinder-database

[8] T. Gadot, S. Istrate, H. Kim, D. Morris, S. Beery, T. Birch, and J. Ahumada, “To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images,” *IET Computer Vision*, 2024.