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. Integrates traditional computer vision techniques with advanced semantic understanding powered by AI models. The platform supports manual and AI-assisted annotation, stores visual metadata and embeddings, and enables intuitive natural language queries to discover relevant visual content. By enabling contextual insights and advanced search capabilities, the system transforms how wildlife media can be explored and utilized.

I. Introduction/Background

The proliferation of wildlife imagery from researchers, conservationists, and photographers has created a need for intelligent systems beyond basic storage and retrieval. This project envisions an AI-enhanced platform explicitly tailored for wildlife media, offering a deeper understanding and more sophisticated tools than general-purpose photo applications. Combining image processing with powerful language models enables semantic search, behavior analysis, and extraction of ecological information from personal and professional wildlife image collections. Through geospatial tools, users can uncover patterns such as migration, seasonal behaviors, and habitat changes, ultimately contributing to ecological research and conservation.

II. Problem Statement

Traditional photo management tools such as Apple Photos or Google Photos are designed for personal memories, not ecological insight. These systems lack domain-specific capability to interpret complex wildlife imagery. They cannot analyze species, behavior, habitat conditions, or track patterns in space and time.

For researchers, conservationists, and wildlife photographers, this presents a significant gap. Vast amounts of wildlife imagery remain disconnected from structured ecological knowledge that could inform species management, habitat protection, or human-wildlife conflict mitigation.

This project bridges the gap by building a specialized platform that transforms wildlife images into structured ecological data. Supporting species classification, behavior tagging, semantic search, and spatial analysis.

For example, understanding wildlife responses to human activity requires spatial awareness and temporal tracking. General-purpose tools cannot support this. Figure 1 shows how a pair of Great Horned Owls relocated their nest after a human disturbance at Coal Creek Trail, 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 like this requires a purpose-built system capable of integrating imagery, location data, and AI analysis. This project addresses this need, turning unstructured wildlife media into actionable ecological insights.

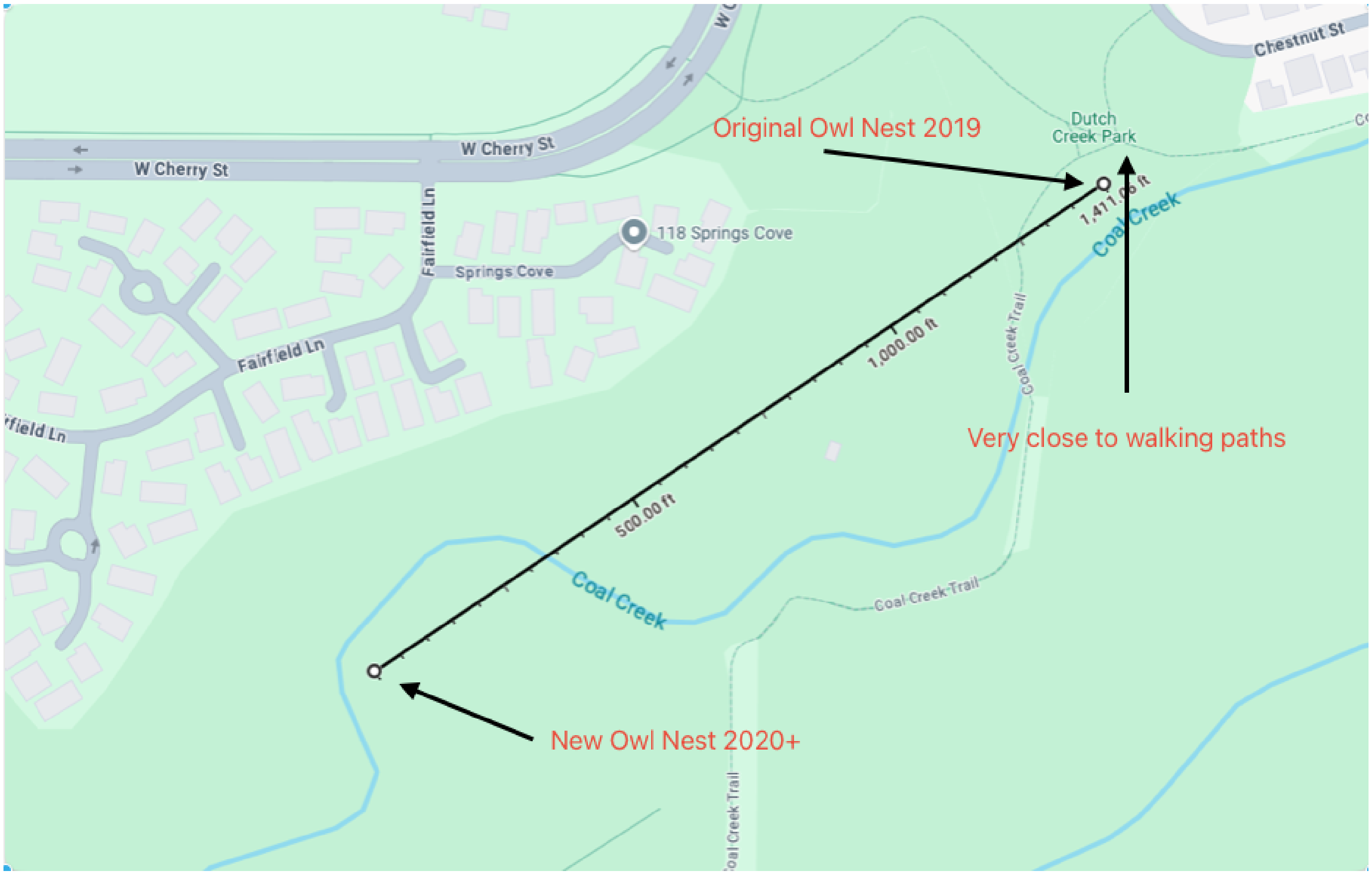

Fig. 1. Great Horned Owl Nesting Patterns

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 [1] model for generating image and text embeddings. The system uses YOLOv8 [2] for object detection and automated cropping, isolating subjects within images to improve classification and embedding quality. PostgreSQL, extended with PostGIS for spatial queries and PGVector for similarity search, manages structured ecological data, species embeddings, and metadata.

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 drives the accuracy of species identification across the pipeline.

The dataset consists of a curated subset of my personal wildlife photography collection, captured over several years across diverse locations. These high-resolution RAW images document species and behaviors in natural settings and include embedded EXIF [3] metadata such as capture date, time, GPS coordinates (when available) and camera settings. To expand taxonomic coverage, the system scrapes species images from Wikipedia using BeautifulSoup, focusing on mammals and birds. These public images are processed to generate a baseline set of canonical embeddings, which serve as a reference model for similarity search and classification.

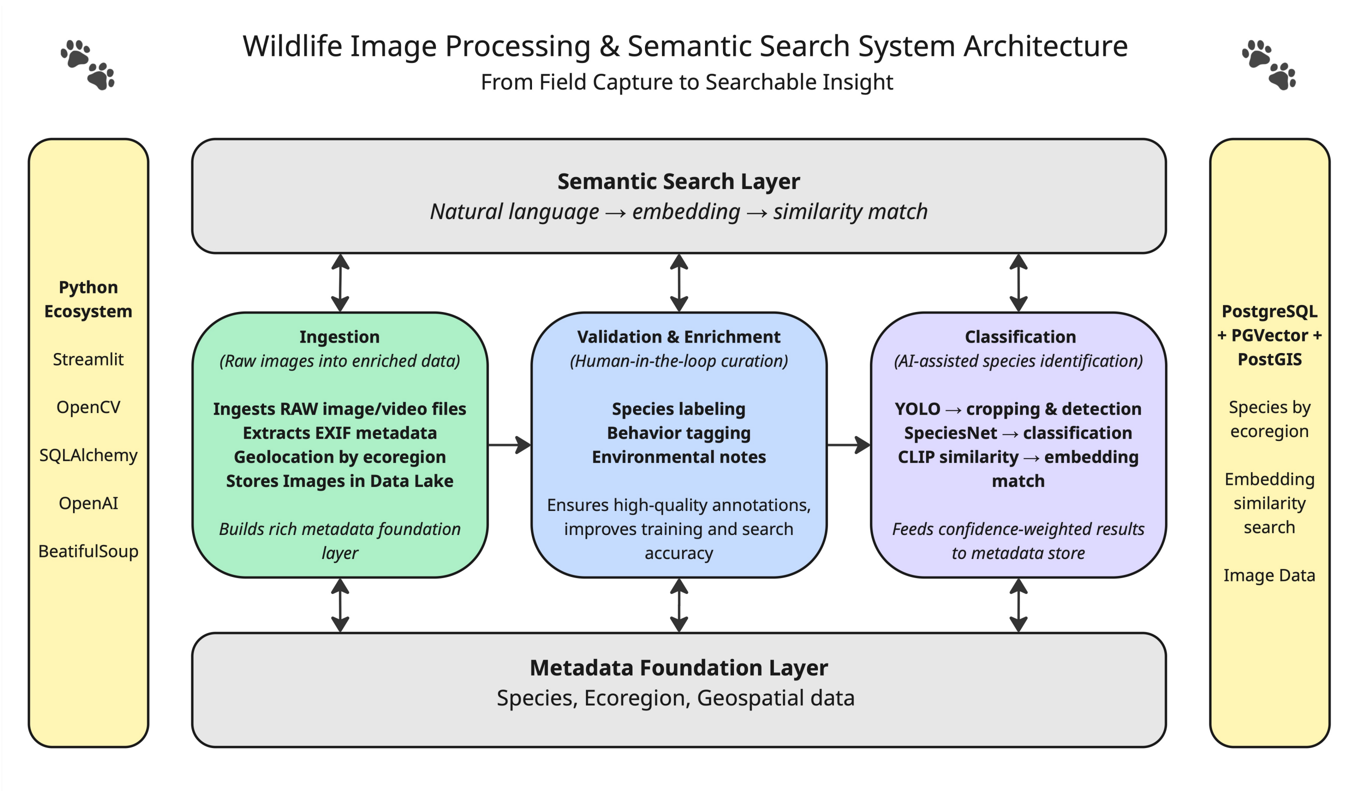

Fig. 2. Wildlife Image Processing and Semantic Search System Architecture

A core strength of the system is its ability to integrate ecological context directly into species prediction by geocoding image locations and constraining comparisons to species known to occur within the corresponding ecoregion. Species range data and ecoregion boundaries from the World Wildlife Fund's [4] normalized species database and spatial shapefiles form the basis for this process. This spatial filtering significantly improves prediction accuracy by eliminating biologically impossible results and prioritizing species based on realistic geographic constraints.

The system combines this spatial awareness with OpenAI's CLIP [1] model, which generates semantic embeddings for images and species concepts. CLIP [1] enables flexible similarity search and clustering, but presents challenges for precise identification. Closely related species often cluster tightly in vector space, limiting separation. By incorporating ecoregion boundaries into the workflow, the system boosts the ranking of likely species and prevents common misclassifications that occur when relying on visual features alone.

A key methodological challenge with embedding-based species identification is the lack of calibrated confidence scores and the dense clustering of similar species within vector space. To illustrate this, the system compares the CLIP [1]-generated embedding for American Bison against other species present in the South Central Rockies ecoregion. As shown in Figure 3, cosine distance values range narrowly, with many species occupying overlapping regions of the embedding space. Although lower distances generally correspond to greater similarity, the compressed scale makes precise species differentiation difficult. This reinforces the need to integrate ecoregion filtering and supervised classification to improve prediction accuracy, rather than relying on embeddings alone.

Classification results are further improved through a supervised model, SpeciesNet [5], trained for wildlife detection and species labeling. Together, these components form a modular, scalable pipeline that converts raw images into structured ecological insight, supported by validation workflows, human-in-the-loop enrichment, and semantic search tools.

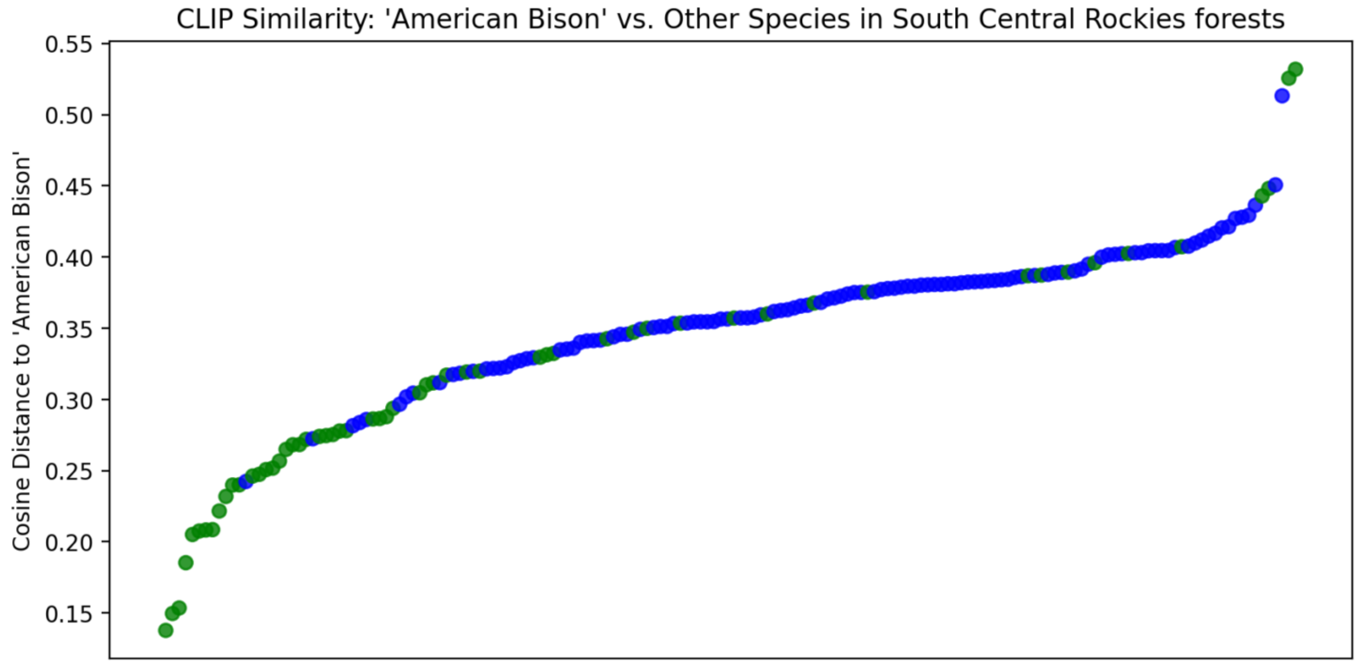

Fig. 3. Cosine distance between the CLIP [1] embedding for *American Bison* and other species within the South Central Rockies forests ecoregion. Dense clustering and compressed distance ranges highlight the challenge of interpreting raw embedding similarity scores for fine-grained species identification.

IV. Data Description

This project is designed primarily for personal use; however, the broader application of this system raises important considerations around data governance, intellectual property, and responsible research practices. Wildlife photographers, in particular, have valid concerns about protecting the ownership and intended use of their images. The system preserves original file structures, metadata, and attribution, and any expansion beyond personal datasets would require clear consent policies, access controls, and safeguards to respect photographer rights. Additionally, the system processes location data for species detection, which presents risks if sensitive sites such as owl nests or den locations are inadvertently exposed. Future public-facing implementations would require location masking or 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 model checkpoints and schema evolution. AI-driven species identification introduces risks of algorithmic bias, especially where training data overrepresents common species or specific visual characteristics. Rare species, seasonal variations, and incomplete datasets may reduce model reliability. To mitigate this, the system combines supervised classification with human-in-the-loop validation and explicitly communicates confidence levels in its outputs. While current use is controlled, scaling to broader datasets or user groups would require additional safeguards, including transparency around AI limitations, secure handling of sensitive data, and ongoing monitoring to prevent misuse or unintended ecological impacts.

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 create a unified platform that organizes wildlife images, improves species identification accuracy, and enables pattern discovery grounded in ecological context.

The system demonstrates how combining OpenAI's CLIP [1] embeddings, supervised species classification, and ecoregion-based filtering improves the reliability of species predictions, especially in complex cases where visually similar species overlap. Applied examples, such as tracking owl nest relocations or comparing species similarity within an ecoregion, showcase the practical value for researchers, conservationists, and photographers seeking to extract meaningful information from large, unstructured photo 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 needed to expand training data, incorporate additional ecological layers, and refine outputs for large-scale or public deployment. Nonetheless, this project provides a foundation for advancing applied machine learning in ecological monitoring and demonstrates a reproducible framework for integrating AI, spatial data, and domain expertise to support 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 [3] 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 [3] 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
EXPANDED 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 photo collections. This system addresses that gap by combining AI-driven classification, semantic embeddings, and spatial filtering based on known species distributions and ecoregions.

The project integrates machine learning models, geospatial data, and human-in-the-loop validation within a scalable, modular pipeline, providing a reproducible framework for extracting ecological knowledge from imagery. Beyond the technical contributions, this work demonstrates how AI can be grounded in ecological reality to produce reliable, interpretable outputs—a challenge central to applied data science.

As the culmination of my data science studies, this project reflects both technical expertise and domain-specific understanding. It integrates machine learning, database design, geospatial analysis, and responsible data management into a coherent, purpose-built system. The educational value lies not only in applying these methods, but in building a tool with practical impact for ecological research and wildlife monitoring. This project demonstrates both readiness for professional practice and the ability to develop applied data science solutions that extend beyond academic exercises into meaningful, real-world applications.

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] G. Jocher, A. Chaurasia, and J. Stoken, "Yolo by ultralytics," 2023, online; accessed June 25, 2025. [Online]. Available: https://github.com/ultralytics/yolov5

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

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

[5] 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.