Wildlife Image Processing & Semantic Search System V2

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

Abstract

Version 1 of this project introduced a system that combined computer vision, metadata extraction, and semantic embeddings to organize and retrieve wildlife imagery. Version 2 builds on that foundation by expanding the platform into a more comprehensive AI-driven framework capable of deeper ecological reasoning and more accurate species identification. This iteration incorporates transfer learning, multimodal vision–language embeddings, supervised species classification, and object detection to improve robustness in complex wildlife imagery [1], [2]. By integrating these models with geospatial analysis and ecoregion-aware filtering, the system grounds predictions in biologically plausible context. The addition of agentic orchestration enables dynamic re-ranking, validation, and context-sensitive inference across multiple signals. Together, these enhancements unify visual, semantic, and spatial information within a modular architecture that improves classification reliability, supports natural language exploration, and enables more meaningful ecological interpretation of wildlife image datasets.

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 language-enabled vision models, the system transforms wildlife photographs into searchable, analyzable data that supports semantic search, behavior analysis, and ecological interpretation across personal and professional image collections. When augmented with geospatial tools, these approaches allow users to uncover patterns related to migration, seasonal behavior, and habitat use, contributing to ecological research and conservation efforts.

Prior work has demonstrated that wildlife imagery presents unique challenges for automated analysis due to environmental variability, background complexity, and fine-grained interspecies similarity, limiting the effectiveness of general-purpose computer vision systems [3]. At the same time, advances in deep learning for ecological monitoring have shown that automated species identification at scale is both feasible and scientifically valuable when models are grounded in ecological context [4]. These findings motivate the continued development of integrated systems that unify computer vision, semantic embeddings, and spatial reasoning, as explored in this second iteration of the platform.

II. Problem Statement

Traditional photo management tools such as Apple Photos or Google Photos lack the domain-specific functionality required to analyze and derive insight from wildlife imagery. These platforms are not designed to interpret complex ecological context, including animal behavior, environmental conditions, or temporal patterns across geographic locations. As a result, they are poorly suited for researchers and conservationists who require systems capable of organizing wildlife images while supporting advanced species classification, semantic search, and predictive analysis. Prior ecological machine learning research emphasizes that meaningful interpretation of wildlife observations depends on integrating visual data with spatial and environmental context rather than treating images as isolated artifacts [5].

Version 2 of this platform addresses this gap by focusing on improved species identification accuracy while strengthening the underlying architecture to make the system more accessible, extensible, and secure. It emphasizes protecting intellectual property through architectural separation of concerns and providing modular tools that others can build upon. These enhancements improve the overall reliability of the system and allow analytical focus to extend beyond identification toward the generation of meaningful ecological insight.

III. Methodology/Approach

This project builds on an existing system developed in Python and PostgreSQL. The current platform provides automated wildlife image ingestion, metadata extraction, and initial species identification.

The new enhancements extend this architecture (see Figure 1), which rests on two core pillars: a Python-based processing pipeline and a PostgreSQL database extended with spatial (PostGIS) and vector (PGVector) search capabilities. Python serves as the engine for ingestion, pre-processing, and analysis, utilizing libraries such as Streamlit for the interactive user interface, OpenCV for image processing and feature extraction, and BeautifulSoup for scraping reference images. Multimodal image and text embeddings are generated using CLIP-based vision–language models, which provide transferable semantic representations across visual domains [1].

Object detection and automated cropping are performed using a YOLO-based architecture, enabling accurate localization of wildlife subjects within complex scenes [6]. Isolating focal subjects prior to embedding generation improves downstream classification and similarity search performance. At the core of the system, the World Wildlife Fund's WildFinder [7] database provides authoritative species-by-region information for grounding predictions in ecological context [5].

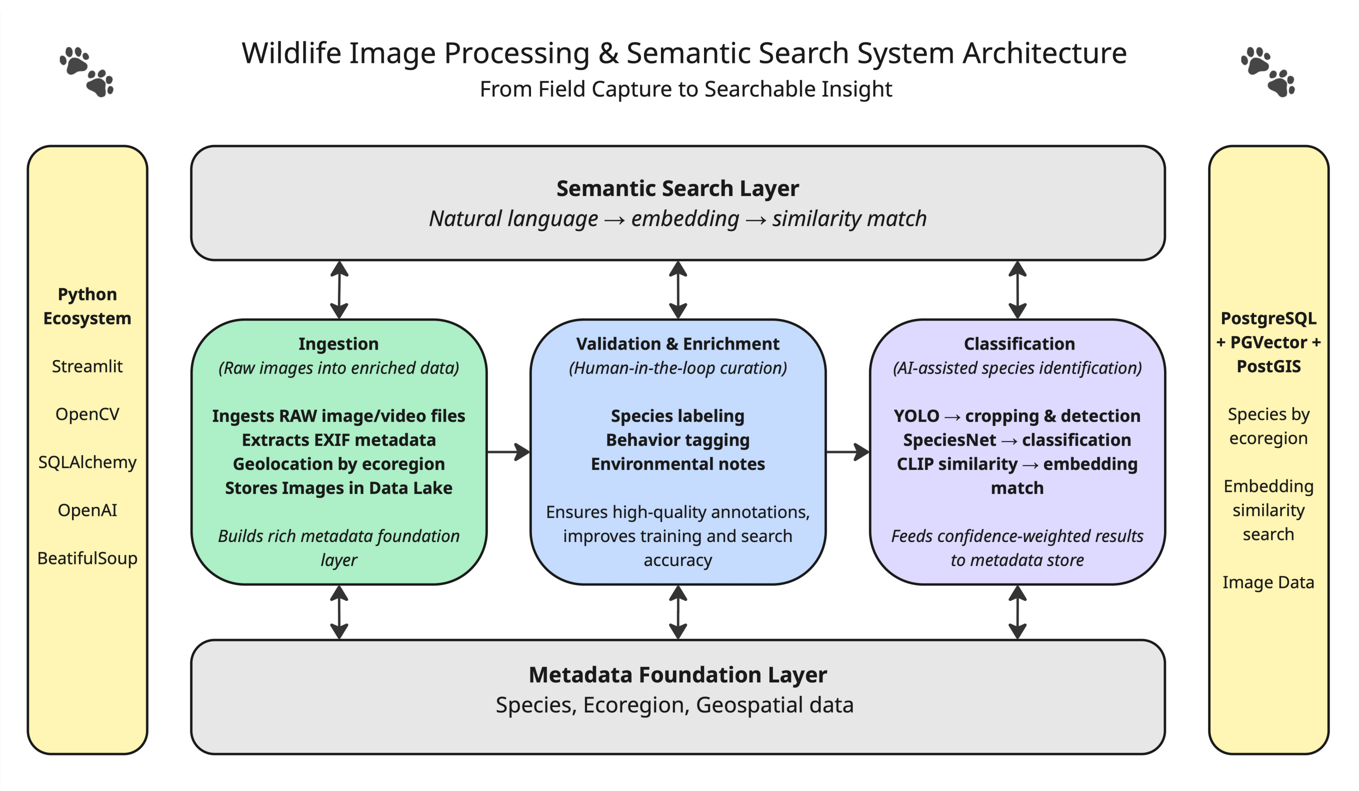

Fig. 1. Wildlife Image Processing and Semantic Search System Architecture

To enrich the feature space, OpenCV is used to generate histograms of dominant colors and to construct a custom color palette that can be queried for similarity analysis. Image metadata, particularly focal length and sensor size, is leveraged to estimate species size as an additional discriminating feature. These attributes supplement embeddings, enabling multi-dimensional comparison beyond purely visual similarity. The combination of embeddings, color, and size properties produces a richer representation of each image and provides more robust inputs for downstream reasoning.

A key methodological advancement in Version 2 is the integration of a custom-trained SpeciesNet model through transfer learning [2]. By fine-tuning a pretrained architecture on a curated wildlife dataset, the model improves classification accuracy for species most relevant to the system. Within the workflow, SpeciesNet operates alongside embeddings and ecological filters, acting either as a primary classifier when its label set matches the query species or as an additional validation signal to confirm or re-rank predictions. This dual role increases flexibility and ensures that model specialization strengthens the broader multi-model pipeline.

Another advancement is the introduction of agentic workflows using LangGraph. Species identification is no longer a single-model decision but a dynamic process in which embeddings, SpeciesNet predictions, color and size features, and ecological presence data are jointly evaluated. The agentic workflow enables conditional logic such as candidate re-ranking, fallback to supervised classification, or adjustment of weighting when text similarity provides stronger signals. This approach produces a transparent, stepwise inference process in which intermediate results and reasoning can be inspected.

The system has also been containerized as a desktop application, providing photographers with local control over sensitive images and models. In parallel, cloud-hosted services expose species datasets, ecoregion layers, and search functionality through an API and Model Context Protocol (MCP) server. This client–server design follows a separation of concerns in which local execution protects intellectual property and sensitive data, while cloud endpoints deliver shared ecological knowledge in a controlled manner. Together, these design choices improve reproducibility, support responsible data use, and enable collaboration without compromising ownership or wildlife protection.

Building on the foundation of Version 1, the project integrates advanced feature extraction, multi-model prediction, supervised transfer learning, and agentic orchestration to support more robust validation and species identification while maintaining clear safeguards. These enhancements collectively strengthen the reliability and extensibility of the platform, ensuring it can evolve into a practical tool for ecological research and wildlife monitoring, as illustrated in Figure 2.

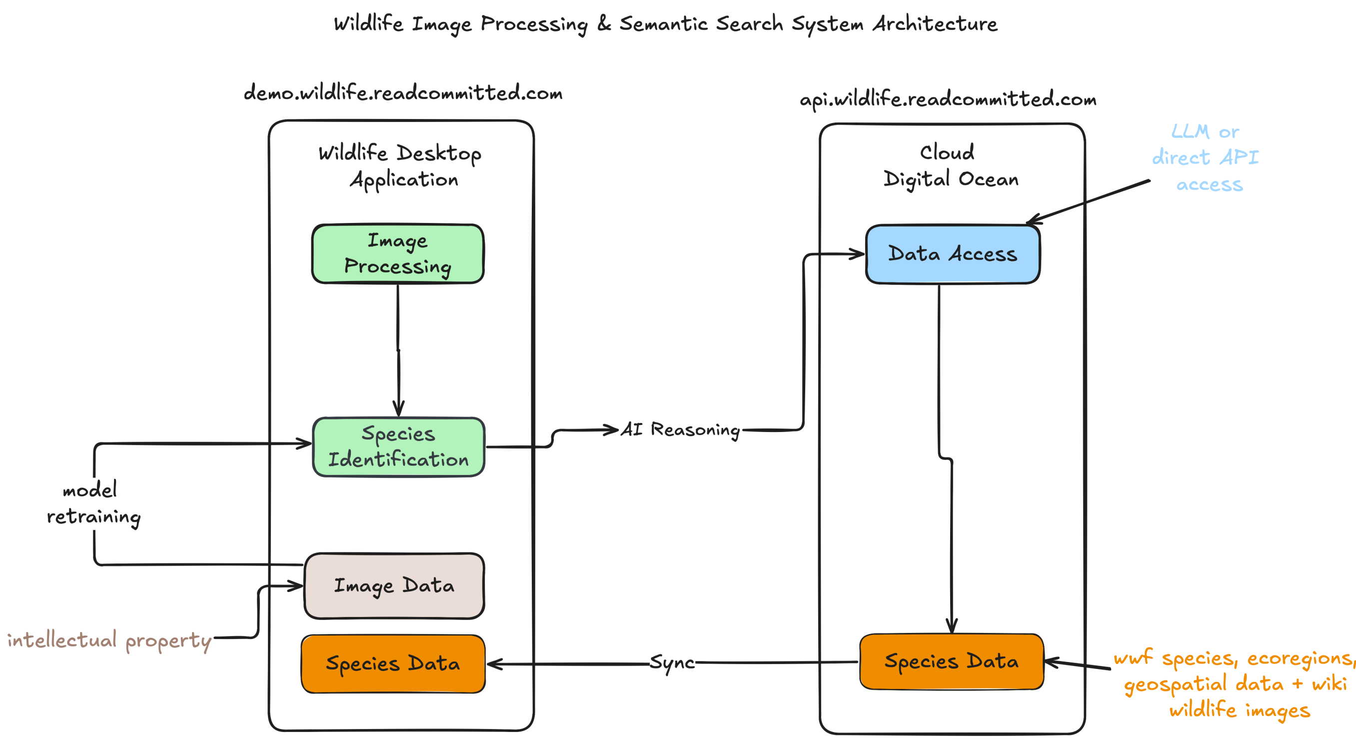

Fig. 2. Wildlife Image Processing and Semantic Search System Architecture V2

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 locations are inadvertently exposed. Public-facing implementations would therefore require location masking or spatial generalization to protect vulnerable wildlife from human disturbance.

These concerns are addressed not only through policy considerations but through the architecture of the system itself. The design follows a deliberate separation of concerns in which sensitive image data and custom-trained models remain under the photographer's control within the desktop client, while cloud-based services provide curated species datasets, ecoregion boundaries, and geospatial layers. This client–server separation preserves intellectual property, reduces the risk of raw imagery misuse, and protects sensitive wildlife locations, while still enabling shared ecological services to support broader research efforts. As illustrated in Figure 2, this architectural balance allows local processing and secure cloud services to coexist in a manner that supports both conservation priorities and collaborative ecological research.

V. Expected Outcomes

The expected outcome of this project is improved species prediction accuracy and enhanced ecological insight through the introduction of agentic pipelines and AI-driven reasoning. Building on the existing platform, Version 2 introduces advanced feature extraction, supervised transfer learning for species classification, and multi-model reasoning workflows designed to support more robust validation and contextual inference.

Transfer learning with a custom-trained SpeciesNet model strengthens species classification by adapting pretrained architectures to a curated wildlife dataset, leveraging feature representations learned from large-scale visual data [8]. Prior work has shown that such fine-tuning approaches improve performance in ecological imagery when compared to generic classifiers, particularly when domain-specific visual variation is present [2].

Together, these components create a unified platform that not only organizes and classifies wildlife imagery, but also enables robust search, advanced validation, and insight generation grounded in ecological context. The resulting system provides a flexible and extensible foundation for future research in wildlife monitoring and applied ecological machine learning.

VI. Timeline

Week	**Focus**	**Deliverables**
Week 1	Project Setup	Setup up the graph database, modeling the schema to support the WWF [7] taxonomy, locations, and relationships needed for species detection and observation. Import the initial data, focusing on getting core nodes and properties in place.
Week 2	Graph Database	Continue working on the graph data model, ensuring that taxonomy, location, and observation nodes are fully connected and queryable. Expand the model to include user, geospatial, images, and embeddings.
Week 3	OpenCV Feature Extraction	Integrate OpenCV to extract additional features from images, such as color histograms and texture metrics, which will be used to enhance embeddings. Begin refactoring the codebase to route all new data and embeddings into the graph database, and update embeddings.
Week 4	Agentic Development	Start agentic development by scaffolding the LangChain backend, including basic MCP prompt templates. Implement the first agent pipeline, where an uploaded image is routed through a species prediction workflow, and the results are pushed to the graph database.
Week 5	Agentic Development Continued	Enhance the LangChain integration, expanding the agent workflows to include prompt-based branching and additional validation logic. Ensure agent outputs are transparently stored, with clear links between predictions, inputs, and derived properties for each observation.
Week 6	Custom Model	Shift focus to model development, introducing a custom-trained SpeciesNet [2] model using transfer learning. The goal is to adapt the pretrained architecture to the project's curated wildlife dataset, improving accuracy for target species and establishing a stronger foundation for multi-model prediction. By fine-tuning on local images, SpeciesNet [2] becomes specialized for the species most relevant to the system, creating a dedicated classification component that complements embedding-based methods and supports more reliable ecological analysis.
Week 7	Final Presentation & Demo	Prepare for the final presentation by polishing documentation, capturing demo videos, and rehearsing the end-to-end workflow. Assemble supporting visuals and summaries that highlight system capabilities, technical choices, and the overall impact of combining graph databases, agentic AI, and enhanced computer vision in wildlife research.

TABLE I
PROJECT TIMELINE

VII. Results

The evaluation of Version 2 highlights the complementary strengths and limitations of embedding-based similarity search and supervised species classification within a multi-model pipeline. When relying primarily on semantic embeddings combined with ecological filtering and re-ranking, the system achieved an overall species prediction accuracy of 77.9 percent. Performance varied by species, with higher accuracy observed for visually distinctive or well-represented species, and lower accuracy for species with subtle visual differences or limited training examples, as illustrated in Figure 3.

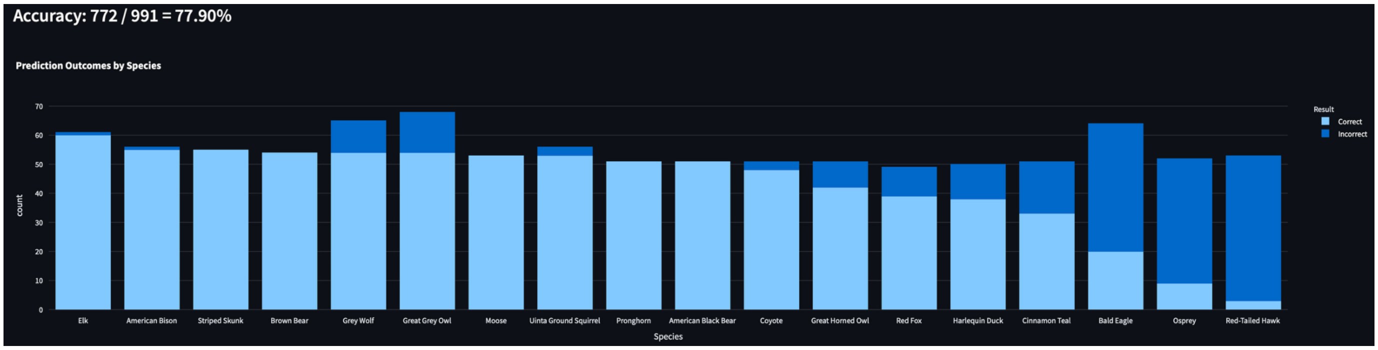

Fig. 3. Multi-Model Species Prediction

Initial experiments using transfer learning with a custom-trained SpeciesNet model produced promising results, with apparent accuracy approaching 97 percent. However, further analysis revealed overlap between training and evaluation images caused by non-independent random splits. Because many images were captured during the same days and sessions, the model exhibited signs of memorization rather than true generalization. To address this issue, grouped train–test splits were introduced based on capture date, ensuring that all images from a given day were assigned exclusively to either the training or evaluation set.

After applying grouped splits, classification performance stabilized, with accuracy reaching approximately 89 percent for species with sufficient diversity across lighting conditions, poses, and capture sessions. Species with limited or homogeneous training data continued to exhibit weaker performance, underscoring the importance of dataset size and variability for effective transfer learning. With an average of approximately 40 training images per species, the model remains underpowered for robust generalization.

Despite these limitations, the results demonstrate the effectiveness of a multi-model approach that combines embeddings, supervised classification, and ecological constraints. The flexibility of agentic re-ranking and validation enables the system to balance confidence, and accuracy, providing a strong foundation for future improvements as additional data becomes available.

VIII. Ethical Considerations

The application of AI to wildlife imagery raises important ethical considerations related to species protection, data stewardship, and responsible use of location-based information. Wildlife photographs often capture sensitive ecological contexts, such as nesting sites, den locations, or breeding behavior, where unintended disclosure can increase the risk of disturbance, harassment, or habitat degradation. As a result, systems designed to analyze and organize wildlife imagery must prioritize ecological responsibility alongside technical accuracy.

This project adopts a "tag responsibly" approach, emphasizing that species identification and metadata enrichment should not automatically imply public disclosure of precise locations or sensitive behavioral information. While geospatial data is essential for ecological reasoning and species validation, public-facing outputs must avoid exposing fine-grained coordinates for vulnerable species. The system is therefore designed to support location generalization, masking, or region-level abstraction, particularly for nesting, denning, or endangered species, ensuring that analytical capability does not translate into ecological harm.

Ethical considerations are further addressed through architectural design choices that preserve photographer ownership and control. Sensitive images and custom-trained models remain local to the desktop client, while shared cloud services provide curated species distributions and ecological context without requiring access to raw imagery. This separation of concerns protects intellectual property, reduces the risk of misuse, and aligns with ethical principles of consent and data minimization.

Finally, the system acknowledges the limitations of AI-driven species identification and avoids presenting predictions as authoritative labels. Human-in-the-loop validation, confidence scoring, and transparent workflows are used to prevent overconfidence in automated outputs. By treating AI as an assistive tool rather than a definitive arbiter, the platform promotes responsible interpretation and use of wildlife data. Collectively, these design choices reflect a commitment to ethical AI practices that respect wildlife, photographers, and the broader goals of conservation-focused research.

IX. Conclusion

This project advances the application of data science to ecological analysis by transforming unstructured wildlife imagery into structured, actionable insight through a modular, AI-driven system. Building on the foundations established in Version 1, Version 2 introduces multi-model species identification, transfer learning, and agentic reasoning to improve prediction accuracy, transparency, and ecological grounding. By integrating semantic embeddings, supervised classification, and spatial constraints, the system addresses key limitations of general-purpose image management tools and aligns automated predictions with biological reality.

Beyond its technical contributions, this work demonstrates the importance of designing AI systems that respect ecological context, intellectual property, and responsible data practices. The platform's separation of local and cloud components enables collaboration and extensibility while preserving photographer ownership and protecting sensitive wildlife data. As a result, the system provides a reproducible and ethically grounded framework for applying machine learning to wildlife research and monitoring.

As the culmination of my data science studies, this project reflects a transition from exploratory experimentation to the design of robust, extensible systems. It unites machine learning, transfer learning, geospatial analysis, data engineering, and responsible AI practices into a cohesive platform with real-world applicability. Version 2 demonstrates readiness for professional practice and the ability to develop applied data science solutions that extend beyond academic exercises to support ecological research, conservation, and wildlife monitoring.

References

[1] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021.

[2] 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.

[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] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Nature Communications*, vol. 9, no. 1, 2018.

[5] 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.

[6] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

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

[8] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.