

Title: LizardLens: A Two-Stage Deep Learning Pipeline for Detecting and Classifying Similar Species in Visually Complex Environments

Authors: Wen Han Chia¹, Ilia Jahanshahi², Le Yang Loh¹, Anqi Zheng¹, Nirai Verma¹, Stephen Mussmann¹, Breanna Shi¹, James T. Stroud³

¹College of Computing, Georgia Institute of Technology, Atlanta GA, United States

²Department of Computer Science, Iowa State University, Ames IA, United States

³School of Biological Sciences, Georgia Institute of Technology, Atlanta GA, United States

Correspondence: stroud@gatech.edu

Abstract

Community science platforms like iNaturalist generate unprecedented volumes of biodiversity data, but their scientific utility depends critically on accurate species identification—a persistent challenge when contributors often lack taxonomic expertise. We developed “*LizardLens*”, a two-stage machine learning pipeline that decouples object detection from species classification to enable fine-grained identification of morphologically similar organisms in visually complex field photographs. Using 10,000 verified iNaturalist images of five *Anolis* lizard species in Florida, we trained specialized YOLO-based detection and Swin Transformer classification models and compared performance against state-of-the-art single-stage architectures. Our two-stage pipeline achieved 83.0% Top-1 accuracy and a macro-averaged F1-score of 89.0%, indicating strong precision-recall performance across species and outperforming single-stage YOLOv8 and YOLOv12 models across all evaluation metrics for all species, with relative improvements ranging from 10.5% to 13.2%. Gradient-weighted Class Activation Mapping (Grad-CAM) indicated that the model’s predictions were consistently associated with regions corresponding to diagnostic morphological (e.g., head shape, feet, and limb lengths) and pattern features (e.g., ocular rings and body patterning), providing evidence that *LizardLens* leverages biologically relevant visual cues consistent with those used by expert taxonomists. Error analysis identified partial occlusion and multiple proximate individuals as primary sources of missed detections, while spurious detections of lizard-like environmental features (e.g., sticks, bark) represented the dominant false positive error mode. We deployed *LizardLens* as an accessible web application featuring interactive bounding box correction, ranked species predictions with confidence scores, directly supporting the “*Lizards on the Loose*” middle school community science initiative. By combining technical advances in fine-grained visual classification with user-centered design, *LizardLens* demonstrates how machine learning can simultaneously enhance data quality for biodiversity monitoring and provide authentic scientific experiences for student participants. Our approach is generalizable to other small-bodied organisms in complex habitats and provides a framework for translating computer vision advances into practical tools for community science and conservation.

Key words: community science, deep learning, automated species identification, object detection, *Anolis* lizards, image classification

1. Introduction

Digital imagery has revolutionized biology, enabling unprecedented scales of biodiversity data collection through community science initiatives (Chandler *et al.*, 2017; Di Cecco *et al.*, 2021). Platforms such as iNaturalist harness the collective power of thousands of contributors, generating organismal observation records that far exceed the capacity of professional scientists alone (Mason *et al.*, 2025). These crowdsourced datasets have become invaluable resources for tracking species distributions, monitoring invasive species, and documenting phenological shifts (Encarnação, Teodósio and Morais, 2021; Jaskuła *et al.*, 2021; Binley *et al.*, 2023). However, the scientific utility of community science data depends critically on accurate species identification—a challenge that remains substantial when contributors lack taxonomic expertise (Koo *et al.*, 2022).

Species misidentification represents a persistent bottleneck in community science platforms such as iNaturalist. While iNaturalist employs a verification system where community members review and correct uploaded identifications, this process is labor-intensive and prone to error propagation (White *et al.*, 2023; Mesaglio *et al.*, 2025). The problem is particularly acute for taxonomic groups with subtle morphological differences or when photograph quality is suboptimal (Jäckel *et al.*, 2023; Coca-de-la-Iglesia *et al.*, 2024). As community science initiatives continue to expand, developing automated tools to improve identification accuracy has become increasingly important for maintaining data quality and scientific rigor (Van Horn *et al.*, 2018).

Recent advances in deep learning offer promising solutions for automated species identification from digital images. Object detection models, particularly the YOLO (“You Only Look Once”) architecture (Redmon *et al.*, 2016), have demonstrated remarkable performance in rapidly localizing objects within images (Murat and Kiran, 2025). Similarly, classification algorithms employing approaches such as convolutional neural networks have achieved high accuracy in distinguishing between visually similar categories (Krizhevsky, Sutskever and Hinton, 2012). However, applying these methods to wildlife imagery presents unique challenges. Many organisms appear small relative to overall image dimensions, and natural habitats introduce substantial visual complexity through varied colors, textures, and structures (Mulero-Pázmány *et al.*, 2025a). Furthermore, closely related species often exhibit superficial similarity that confounds even state-of-the-art classification models (Shen *et al.*, 2025). While end-to-end models can perform classification from the full image, they often struggle with fine-grained visual discrimination when targets are small and backgrounds are complex (Mulero-Pázmány *et al.*, 2025b). Decoupling these tasks into a two-stage pipeline—where specialized detection models first localize organisms and dedicated classification models subsequently identify species from cropped images—allows each component to be independently optimized, potentially improving overall performance (Ren *et al.*, 2016; Mulero-Pázmány *et al.*, 2025b).

Here, we developed a machine learning pipeline specifically designed to address these dual challenges of fine-grained object detection in visually complex environments and accurate classification of morphologically similar species. Our system integrates object detection and species classification to

process community science photographs of *Anolis* lizards in Florida. This study system is particularly well-suited for method development, as thousands of verified iNaturalist records exist for five common species that vary in their degree of visual similarity (Van Horn *et al.*, 2018). Anoles exemplify the detection and classification difficulties outlined above: they are typically small within photographs, inhabit structurally complex vegetation, and several species share overlapping morphological features (Losos, 2009; Stroud and Losos, 2020; Stroud *et al.*, 2023; Price *et al.*, 2025).

Our pipeline directly supports “*Lizards on the Loose*,” a community science initiative engaging middle school students in biodiversity monitoring. While this program provides valuable scientific training and generates occurrence data for both native and invasive species, student participants often lack the taxonomic expertise for confident species-level identification (Hooykaas *et al.*, 2019). By providing immediate, probability-based species classifications, our LizardLens pipeline dramatically improves data quality while maintaining student engagement. This approach addresses two critical conservation objectives: first, it creates authentic research experiences that foster scientific literacy in young students (Queiruga-Dios *et al.*, 2020); second, it generates high-quality verified records for four invasive anole species, supporting management efforts and invasion biology research (Encarnação, Teodósio and Morais, 2021). This integration of technological innovation with community science demonstrates how machine learning can enhance both the scientific value and educational impact of public participation in biodiversity research.

Our study had four primary objectives: (1) develop and validate a two-stage machine learning pipeline optimized for fine-grained detection and species-level classification of small animals (here, *Anolis* lizards) in visually complex field photographs; (2) compare our pipeline's performance against state-of-the-art single-stage detection models to assess the advantages of task decoupling; (3) deploy an accessible web-based application (LizardLens) for community science use; and (4) interpret the classification model's decision-making process through gradient-weighted class activation mapping (Grad-CAM) (Selvaraju *et al.*, 2017) to understand which morphological features drive species predictions. In the following four sections, we describe our contributions towards each of these four objectives.

2. Methods

Our classification pipeline consisted of two sequential stages: (1) object detection to localize lizards within digital images, and (2) species-level classification of detected individuals. Below, we describe dataset assembly, model architecture, training procedures, and performance evaluation for each stage.

2.1. Training dataset assembly

We assembled a training dataset from iNaturalist photographic records of the five most common *Anolis* lizard (anole) species in Florida: the American green anole (*Anolis carolinensis*), Puerto Rican crested anole (*A. cristatellus*), Hispaniolan bark anole (*A. distichus*), Cuban knight anole (*A. equestris*), and Cuban

brown anole (*A. sagrei*). For each species, we filtered for “Research Grade” images, photographs that have received community verification for correct species identification, and downloaded the images from the iNaturalist records using the API link (*Export Observations, iNaturalist*, 2026). As of 17th Jan 2025, the image counts of brown anole, green anole, crested anole, bark anole and knight anole are 44756, 44242, 6075, 2671 and 2301 respectively. Next, most of the anoles’ bounding boxes were automatically annotated using GroundingDINO (Liu *et al.*, 2024), a zero-shot Vision-Language model on Roboflow (Dwyer, Nelson and Hansen, 2026). The remaining unannotated anoles images were manually annotated by the authors (W.H.C., A.Z., N.V.). Next, the authors then retained only images that display distinct species features as stated in the Florida Anole field identification guide on iNaturalist (Thawley and Stroud, 2026). To mitigate the issue of class imbalance, we included only 2000 images of each species, yielding a total dataset count of 10,000 images.

Images were randomly partitioned into training (n = 8,000; 80% of photos), validation (n = 1,000; 10%), and test (n = 1,000; 10%) subsets, stratified by species to maintain equal class representation (1,600 training, 200 validation, and 200 test images per species). Given that some images contain more than one anole lizard detection, the per-species lizard detection counts slightly exceed 2000 (Table 1). However, the resulting class imbalance is negligible (maximum deviation <1% from the mean). Note that all of the images with multiple anole lizard detections contain only one species; no images contain more than one anole species.

2.2. Stage 1: Detection model training

The primary objective of the object detection stage was to localize lizards within input images and generate bounding boxes around detected individuals. We evaluated three state-of-the-art object detection architectures: YOLOv8x with 68.2 million parameters (Varghese and M., 2024), YOLOv11x with 56.9 million parameters (Khanam and Hussain, 2024), and Real-Time Detection Transformer RT-DETR(L) with 42 million parameters (Zhao *et al.*, 2024). The best-performing model was selected for integration into the classification pipeline. For object detection training, we created a modified version of the dataset by replacing species-level labels with a single "lizard" class label, enabling the model to detect *Anolis* lizards regardless of species identity. Each model was fine-tuned with the training dataset for 100 epochs using a batch size of 128 and Stochastic Gradient Descent (SGD) as the optimizer. Input images were resized to 640x640 using letterbox padding to preserve aspect ratio. Model training was conducted on two A100 GPUs with 80GB VRAM using Ultralytics package (Jocher, Qiu and Chaurasia, 2023).

2.3. Detection model performance evaluation

We evaluated detection performance using precision, recall, F1-score at an evaluation IoU threshold of 0.5, and mean average precision (mAP) calculated across intersection over union (IoU) thresholds ranging from 0.50 to 0.95 (mAP@50-95) (Padilla *et al.*, 2021) (Equations 1 - 4). During inference, non-maximum suppression (NMS) was performed at an IoU threshold of 0.7 to remove redundant overlapping detections prior to evaluation. Precision quantifies the proportion of correct detections

among all predicted bounding boxes, while recall measures the proportion of actual lizards successfully detected, revealing whether the model systematically misses certain species (Sokolova and Lapalme, 2009). F1-score provides a harmonic mean of precision and recall, and mAP@50-95 integrates precision across multiple IoU thresholds to assess both localization accuracy and detection sensitivity (Powers, 2020; Padilla *et al.*, 2021).

Equation 1: Precision = TP / (TP + FP)

where TP is true positives (correct predictions for the detections) and FP is false positives (incorrect predictions assigned to the detections). Recall measures the proportion of actual lizards of each species that were correctly identified, revealing whether the model systematically misses certain species:

Equation 2: Recall = TP / (TP + FN)

where FN is false negatives (lizard detections incorrectly assigned to other classes in test images). F1-score provides a harmonic mean of precision and recall, offering a balanced assessment of model performance that accounts for both prediction accuracy and detection sensitivity:

Equation 3: F1-score = $2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

Mean average precision (mAP) is computed by first calculating the average precision (AP) for each class as the area under its precision–recall curve, then averaging AP values across all classes. mAP@50–95 averages these values across ten evaluation IoU thresholds from 0.50 to 0.95 (in 0.05 steps), providing a stricter measure of localization quality than single-threshold mAP (Padilla *et al.*, 2021):

Equation 4: Mean Average Precision (mAP) = $\sum_{i=1}^N \text{AP}_i / N$

where N is the number of classes and AP_i is the average precision for class i.

We prioritized two metrics when selecting the optimal detection model. First, recall was considered paramount because missed detections in this initial stage prevent any downstream classification, causing complete pipeline failure for those images. High recall ensures maximum retention of lizard-containing images for subsequent species classification. Second, mAP@50-95 was emphasized because accurate bounding box localization is critical for preserving diagnostic morphological features in cropped images passed to the classification model. Tight bounding boxes minimize extraneous background information that could confound species classification, while loose boxes risk excluding key identifying characteristics.

The following metrics were recorded: precisions of 0.946, 0.933, 0.965, recall of 0.942, 0.932, 0.938, F1-score of 0.944, 0.935, 0.951 and mAP@50-95 of 0.808, 0.800, 0.796 for YOLOv8x, YOLOv11x and RT-DETR(L) respectively. As such, the YOLOv8x model was selected as the first stage of the pipeline.

2.4. Stage 2: Classification model training and evaluation

Following object detection, cropped lizard images were passed to an anole species classification model to assign probability-based identifications to one of five *Anolis* species. We fine-tuned and evaluated the performances of 3 classification models, data-efficient image transformers (DeiT) base model with 87 million parameters (Touvron *et al.*, 2021), Vision Transformer (ViT) base model with 86.9 million parameters (Dosovitskiy *et al.*, 2021) using pre-trained weights from the google/vit-base-patch16-224 checkpoint available on Hugging Face (Deng *et al.*, 2009; Wu *et al.*, 2020), and Swin Transformer base model with 88 million parameters (Liu *et al.*, 2021). For classification model fine-tuning, we created a modified version of the dataset by cropping images of the anoles based on the bounding box annotations in the original dataset. The original species labels were retained. Each model was trained for 30 epochs using Adam optimization, image size of 384, batch size of 128, and HuggingFace package (Wolf *et al.*, 2020). Two H100 GPU with 80GB VRAM were utilized to support this operation.

We evaluated classification performance using top-1 accuracy (Equation 5) together with per-species precision, recall, and F1-score. Top-1 accuracy measures the proportion of test images for which the model's highest-confidence class prediction matches the ground-truth label, providing an overall measure of classification performance. Per-species precision was emphasized as an indicator of prediction reliability for each species.

Equation 5: Top-1 Accuracy = (Number of correct predictions) / (Total number of images in test set)

where a top-1 prediction is the class assigned the highest confidence score by the model, and a prediction is counted as correct if it matches the ground-truth label of the anole instance.

The performance of the models, DeiT, ViT and Swin Transformer were logged respectively: top-1 accuracy of 0.907, 0.919 and 0.939. Furthermore, the precision, recall and F1-score of each class were logged as shown in [Table 2](#). Swin Transformer scored a high precision (above 90%) for almost all classes except for Brown anole. Coupled with a high top-1 accuracy of 0.939, the Swin Transformer was selected to be the classification model for the pipeline.

2.5. Final Model Performance Evaluation

We wanted to compare the performance of LizardLens, the two-stage classification pipeline consist of fine-tuned YOLOv8x detection model and Swin Transformer classification model, against other single-stage detection model. YOLOv8x serves a dual role in our evaluation: it is the detection backbone within our two-stage pipeline, and it is evaluated independently as a single-stage baseline. This controlled comparison isolates the contribution of the dedicated classification stage — since the same detector produces the bounding boxes in both settings, any performance difference is directly attributable to replacing YOLOv8x's built-in classification head with the Swin Transformer backbone. This ablation-style

design answers a core question of this study: whether the added complexity of a two-stage pipeline yields meaningful gains in fine-grained species discrimination over a single-stage approach. Secondly, YOLOv12 introduces an attention-centric architecture that departs from the purely convolutional feature extraction used in prior YOLO variants (Tian, Ye and Doermann, 2025). Because attention mechanisms can capture long-range spatial dependencies between discriminative features — such as dewlap coloration, dorsal patterning, and head scale morphology — this architecture is a compelling baseline for fine-grained species classification tasks where subtle inter-class differences are spatially distributed across the image. As such, the YOLOv12x and YOLOv8x models were chosen as the single-stage detection models of comparison against the two-stage classification pipeline, LizardLens. They were fine-tuned with the original training dataset ($n = 8090$) and their performance was evaluated on the held-out test set using standard classification metrics such as top-1 accuracy and macro-averaged precision, recall, and F1-score (Sokolova and Lapalme, 2009). Top-1 accuracy measures the proportion of images for which the model's highest-probability prediction matched the ground truth species label at the detection level. The evaluation results are shown in Section 3.1.

2.6. LizardLens: Pipeline Deployment and Accessibility

To make our validated classification pipeline accessible to community scientists and educators, we developed a web-based application for LizardLens that processes field photographs and returns immediate species identifications. The application was designed specifically to support non-expert users—including middle school students participating in the Lizards on the Loose initiative—by providing probability-based species classifications that enhance identification accuracy while maintaining user engagement. The complete user workflow is illustrated in [Figure 1](#).

[Figure 1](#). End-to-end web application workflow. Users upload a field image, the backend detects anoles and produces bounding boxes. Cropped lizard detections are passed to the species classifier. The frontend visualizes bounding boxes and per-lizard detection confidence scores before optionally submitting validated observations to iNaturalist.

2.6.1. LizardLens: System Architecture

The application consists of a React user interface and a FastAPI (Ramírez, 2026) python server hosting the trained detection and classification models. The system is deployed on Georgia Tech's College of Computing server and accessible at (*Lizard Lens*, 2026). When users upload photographs, images are transmitted to the server via HTTP POST, processed through the pipeline, and results are returned to the client interface in JSON format. The client application transforms the JSON response into a list of predictions. It then takes these predictions to render Bounding Boxes over the original image, color-coded by confidence level, with labels displaying the species name and confidence percentage.

2.6.2. LizardLens: User Workflow

The application guides users through four sequential stages ([Figure 1](#)):

(1) Image Upload and Automated Inference. Users upload field photographs through the web interface which accepts standard JPG and PNG formats. Upon submission, the back-end detection model (Stage 1) localizes individual lizards and generates bounding boxes, which are then passed to the classification model (Stage 2) for species-level predictions. Inference typically completes within 1.28 seconds per image on a standard CPU (modern Intel Core/Xeon with 4GB+ RAM).

(2) Interactive Detection Review. The interface displays the original photograph with color-coded bounding boxes overlaid on each detected lizard, colored by confidence level (Green for high confidence > 80%, Yellow for medium > 60%, and Red for low $\leq$ 60%). Users can visually assess detection quality and manually adjust bounding boxes by resizing or repositioning to correct incomplete or inaccurate detections (e.g., boxes that exclude diagnostic features or include excessive background). When a bounding box is edited, the system automatically re-extracts the cropped region and re-runs the classification model to update the species prediction based on the corrected detection.

(3) Species Identification and Confidence Assessment. For each individual lizard that LizardLens detects, the application presents: (i) the predicted species common name and scientific name (i.e., Genus, species), (ii) the model's confidence score (ranging from 0–100%), and (iii) a ranked list of alternative species predictions with associated confidence values. This transparency allows users to evaluate prediction certainty and recognize ambiguous classifications that may warrant expert verification.

(4) Data Submission and Community Impact Tracking. Validated observations are submitted directly to the iNaturalist platform. Upon successful submission, the community metrics dashboard reflects the new contribution through real-time counters and species tallies, providing users with immediate scientific validation and visual confirmation of their role within the broader research initiative.

2.6.3. iNaturalist Integration and Real-Time Community Dashboard

Integration with the iNaturalist platform via its public API facilitates streamlined biodiversity data contribution and amplified community impact (Callaghan *et al.*, 2022). Once a user validates an automated detection and species prediction, the application enables a direct-to-iNaturalist submission workflow. This process encapsulates the field photograph and predicted species identification into an iNaturalist observation record. The development of a live community metrics dashboard (Figure 2A) fosters participation and provides immediate feedback. Leveraging React-based animation libraries (e.g., Framer Motion and React-Countup), the dashboard retrieves real-time statistics from the iNaturalist API, including total observations, unique species counts, and contributor totals across the Florida Anolis dataset. The implementation of optimistic UI updates ensures that these counters increment instantly upon a successful upload request, providing participants with a responsive experience that bridges the gap between individual contribution and collective scientific impact (Jennett *et al.*, 2016).

Figure 2. Screenshots of the LizardLens web application. (A) Homepage of the web application – a user-friendly interface that displays recent iNaturalist submissions made through the application. (B) Displays

the captured image and options to change image, classify the species or manually annotate the bounding box. (C) The LizardLens prediction of the anole bounding box, confidence score, number of species and species classification are displayed. User can edit the predicted bounding box or choose to upload this observation to iNaturalist.

7. Model Interpretability Analysis

To understand which morphological features the classification model relies on for species discrimination, we applied Gradient-weighted Class Activation Mapping (Grad-CAM) to visualize spatial attention patterns during inference. Grad-CAM generates heatmaps highlighting image regions that strongly influence the model's predictions, enabling post-hoc interpretation of the classification decision process (Selvaraju *et al.*, 2017; Chattopadhyay *et al.*, 2018).

We implemented Grad-CAM using the fine-tuned Swin Transformer as mentioned in Section 2.4. Because Swin Transformer employs a hierarchical encoder architecture incompatible with standard Grad-CAM implementations, we developed a custom reshape function to convert the model's multi-scale feature representations into activation maps suitable for gradient-based attention visualization. We generated Grad-CAM heatmaps at multiple encoder layers and evaluated which layer produced the most biologically interpretable attention pattern. Heatmaps derived from earlier layers exhibited diffuse attention patterns dominated by low-level texture, while later layers often produced overly coarse activations. We selected the third encoder stage for detailed analysis because it consistently produced spatially localized heatmaps centered on the lizard body while preserving sufficient anatomical detail.

To assess whether the model attended to species-diagnostic morphological features, we randomly sampled 25 test images (5 per species) and J.T.S. qualitatively evaluated their corresponding Layer 3 Grad-CAM heatmaps against established identification criteria.

3. Results

3.1. LizardLens: Overall Pipeline Performance

The LizardLens two-stage pipeline outperformed both single-stage baselines across all evaluation metrics (Table 3). LizardLens achieved an F1-score of 0.890 and top-1 accuracy of 83.0%, compared with 0.802 and 74.0% for YOLOv8x and 0.724 and 68.0% for YOLOv12x. Relative improvements over the next-best model (YOLOv8x) ranged from 10.5% (precision) to 13.2% (mAP@50-95), with consistent gains across all metrics including localization accuracy (mAP@50: 0.891 vs. 0.794; mAP@50-95: 0.704 vs. 0.622).

These gains can be attributed directly to the dedicated classification stage. Because YOLOv8x serves as both the detection backbone within LizardLens and as an independent single-stage baseline, the two pipelines share identical bounding box predictions — any performance difference is therefore

attributable solely to replacing YOLOv8x's built-in classification head with the Swin Transformer. The magnitude of improvement across all metrics confirms that decoupling detection from classification and optimizing each task independently yields substantial gains in fine-grained species discrimination. Notably, YOLOv12x underperformed even the convolutional YOLOv8x despite its attention-centric architecture, which can in principle capture long-range spatial dependencies between discriminative features such as dewlap coloration and dorsal patterning (Tian, Ye and Doermann, 2025). This result suggests that architectural capacity for modeling spatial relationships does not, on its own, compensate for the lack of a dedicated classification stage when targets are small and interspecific differences are subtle.

3.2. Species-Level Performance

LizardLens's performance varied among the five *Anolis* species but remained consistently high (Table 4). All species achieved precision, recall, and F1-scores exceeding 0.85, indicating robust classification regardless of species identity. In contrast, the single-stage YOLOv8x and YOLOv12x models exhibited substantially lower and more variable performance across species, with several species falling below 0.80 for multiple metrics. The pipeline performed best for *A. distichus* (F1-score = 0.930; Table 4), but was strong for all species (range = 0.850 [*A. sagrei*] - 0.920 [*A. cristatellus*]). Differences between species likely represent both biological (high intraspecific variation) and technical (bad photograph quality) features.

3.3. LizardLens: Error Analysis

Our analysis of pipeline errors revealed specific failure modes that inform both the interpretation of performance metrics and potential avenues for improvement. We categorized errors into three primary types: missed detections (false negatives), spurious detections (false positives), and species misclassifications among correctly detected lizards.

3.3.1. Overview of Error Sources

The normalized confusion matrix for the LizardLens pipeline illustrates the distribution of these error types (Figure 3). Missed detections (false negatives, Background column) occurred when the Stage 1 detection model failed to localize lizards present in images, preventing downstream classification. These errors reduced recall, F1-score, and mAP values. Spurious detections (false positives, Background row) occurred when the detection model incorrectly identified non-lizard image regions as lizards, which were subsequently passed to the classifier and assigned species labels. These errors reduced precision, F1-score, and mAP. Species misclassifications (non-diagonal elements) occurred when the Stage 2 classification model assigned incorrect species identities to correctly detected lizards. Together, these three error sources contributed to the reduction in overall Top-1 accuracy from perfect classification.

Quantitatively, out of the 1011 anole lizard detections in the test image dataset, 67 errors (6.63%) resulted from missed detections, 81 errors (8.01%) from detections, and 39 errors (3.86%) from species misclassifications among correctly detected lizards as shown in the confusion matrix (Figure 4). Among species-level classification errors, the most frequent confusions ($n = 10$) occurred between brown anoles (*A. sagrei*) and crested anoles (*A. cristatellus*), which exhibit exceptional similarity in body shape, coloration, dorsal patterning, and habitat use (Stroud *et al.*, 2024) and can be difficult for even expert biologists to distinguish (Stroud pers. obs.). Notably, errors involving misclassification of non-lizard background objects as lizards outnumbered confusions between any pair of lizard species, indicating that distinguishing lizards from visually similar environmental features represents a greater challenge than discriminating among *Anolis* species.

Figure 3. The normalized confusion matrix of the classification pipeline predictions of each anole species.

Figure 4. Confusion matrix, in absolute counts, of the classification pipeline's species predictions for each anole class. Out of 1,011 anole lizard detections in the test image dataset, 81 errors (8.01%) were false positive detections, 67 errors (6.63%) were missed detections, and 39 errors (3.86%) were species misclassifications among correctly detected lizards.

3.3.2. False Negative Errors: Missed Detections

False negatives occurred when the Stage 1 detection model either failed to generate bounding boxes around lizards or produced boxes with insufficient overlap ($\text{IoU} < 0.7$) with ground truth annotations. Through qualitative analysis of error cases, we identified two primary causes:

Partial Occlusion. When lizards were partially obscured by vegetation, substrate, or other environmental features, the detection model frequently failed to localize individuals (Figure 5A). Images where only the head or a small body portion was visible proved particularly challenging, likely because the majority of training images depicted largely unobstructed lizards with most body regions visible.

Multiple Proximate Individuals. When multiple lizards appeared in close spatial proximity in the same photograph, the detection model often generated a single bounding box encompassing multiple individuals rather than separate boxes for each lizard (Figure 5B). This error likely stems from the Non-Maximum Suppression (NMS) post-processing step, which removes overlapping predictions above a defined IoU threshold (currently set at 0.7). When two lizards are closely positioned, their predicted bounding boxes may exceed this overlap threshold, causing NMS to retain only the highest-confidence prediction and suppress the others.

For our primary use case—species-level classification for community science—this error mode has minimal practical impact because the co-occurring lizards are conspecific. As mentioned in Section 2.1, all of the images with multiple anole lizard detections contain only the same species class. This suggests

that missing one of two conspecific individuals does not substantially reduce observation quality. Importantly, most online community science initiatives, such as iNaturalist, do not allow multi-species photographic records, so photographs with more than one species are unable to be contributed.

3.3.3. False Positive Errors: Spurious Detections

False positives occurred when the detection model incorrectly identified non-lizard image regions as lizards (Figure 5C), which were subsequently passed to the classification model and assigned species labels. Post-hoc analysis of confidence distributions (Figure 6) revealed that false positive predictions exhibited substantially lower and more variable detection confidence (median 0.54, IQR 0.39–0.73) compared with true positive detections, which were tightly clustered at high confidence (median 0.91, with 75% of cases exceeding 0.86). The spuriously detected objects often resembled lizards in color, texture, or shape—such as sticks, bark, dried leaves, or dappled lighting patterns on vegetation—suggesting the model may rely partially on low-level visual features (e.g., texture, color patches) rather than holistic morphological recognition. In Section 5.3, we discuss approaches to reduce spurious detections

Figure 5. Representative error modes in the detection model. (A) False-negative due to partial occlusion (n=41 images, 4.06% of test set): the model fails to detect a lizard when only a small portion of the body (e.g., head) is visible and obscured by surrounding vegetation or substrate. (B) False-negative due to multiple proximate individuals (n=9 images, 0.89% of test set): the model generates a single bounding box encompassing two closely positioned lizards rather than detecting each individual separately. (C) False-positive detection: the model incorrectly identifies non-lizard objects (e.g., sticks, bark, or lighting patterns) as lizards, often with low confidence. Blue boxes indicate predicted detections, and red boxes indicate ground truth annotations.

Figure 6. Distribution of prediction confidence across outcome categories. Boxes span the IQR (Q1–Q3), white lines mark medians, yellow diamonds mark means, and grey dots denote outliers. True positives (n = 896, median 0.91, IQR 0.86–0.94) and misclassifications (n = 39, median 0.90, IQR 0.86–0.93) are tightly clustered at high confidence, whereas false positives (n = 81, median 0.54, IQR 0.39–0.73) occupy a much broader, lower range. Missed-detection statistics (median 0.73, IQR 0.42–0.86) are computed over the 44 of 67 cases for which a low-IoU prediction was produced; the remaining 23 are true missed detections with no prediction and thus no confidence value.

3.3.4. Trade-Offs and Deployment Considerations

The optimal balance between false negatives and false positives depends on the intended application context. For the Lizards on the Loose educational initiative, minimizing false negatives (ensuring detected lizards are not missed) may be prioritized to maintain student engagement and maximize capture of valid observations, even if this increases false positives that students can readily identify and dismiss during interactive review (see Section 2.6.2). Conversely, for fully automated biodiversity

monitoring pipelines requiring minimal human verification, minimizing false positives should be prioritized to reduce manual review burden and prevent erroneous records from entering databases.

Our LizardLens web application currently uses confidence threshold of 0.3 and NMS IoU threshold of 0.7 values optimized for retaining true positives, but these parameters can be dynamically adjusted based on user feedback, application context, and emerging deployment priorities. The interactive review interface ([Figure 2C](#)) provides an additional quality control layer that mitigates both error types by allowing users to validate, correct, or reject automated predictions before submitting observations to online community science databases such as iNaturalist.

3.4. Grad-CAM: Model Interpretability

3.4.1. Alignment With Known Diagnostic Features

We conducted a body-part level analysis of the Grad-CAM visualizations (n=25 images, 2.5% of the test set) to investigate the area of the image that the classification model attended to. The area of focus included the diagnostic features of the species ([Table 5](#)), as well as other body parts (e.g. front legs, hind legs, toe pads, tail) and background noises. The analysis revealed that the classification model frequently attended (n=21 images, 84% of analysed images) to morphological features known to be species-diagnostic ([Table 5](#)). However, the model did not consistently utilize all recognized diagnostic characters and, in some cases, attended to non-diagnostic features. For example, although *A. distichus* exhibits distinctive mottled body patterning (light and dark flecks), the Grad-CAM heatmap ([Figure 7](#)) attended primarily on the eyes and legs rather than body coloration. Similarly, for *A. cristatellus* (Crested anole), the model did not focus on dorso-lateral stripes that are prominent field marks in juveniles, females, and subadult males.

The Grad-CAM analysis provides preliminary insights into the classification model's decision-making process but is limited by its qualitative nature and small sample size. The sample size was intentionally limited to enable careful qualitative interpretation across species, as large-scale Grad-CAM visualization is computationally straightforward but requires manual inspection to assess biological relevance. Additionally, Grad-CAM visualizes where the model focuses but does not definitively establish causal relationships between highlighted features and classification decisions (Selvaraju *et al.*, 2017).

[Figure 7](#). Grad-CAM visualizations for a Brown Anole (*Anolis sagrei*) across model stages. From left to right: original image, followed by Grad-CAM heatmaps from the four stages (Stage 1-Stage 4) of the Swin Transformer. Warmer colors (red/yellow) indicate regions with higher contribution to the model's prediction. The model primarily attends to the head and limb regions rather than the dorsal patterning, indicating partial alignment with known diagnostic features. Original image (left) followed by Grad-CAM heatmap from the Swin Transformer (right). Warmer colors (red/yellow) indicate regions with higher contribution to the model's prediction. The model primarily attends to the toepad and limb regions rather than the dorsal patterning, indicating partial alignment with known diagnostic features.

4. Discussion

Accurate species identification is a persistent bottleneck limiting the scientific utility of community science biodiversity data (Kosmala *et al.*, 2016; Kelling *et al.*, 2019). Here, we developed LizardLens, a two-stage machine learning pipeline that decouples object detection from species classification to identify morphologically similar *Anolis* lizard species in visually complex field photographs. Our pipeline substantially outperformed state-of-the-art single-stage architectures across all evaluation metrics, achieving 83.0% Top-1 accuracy and relative improvements of 10.5–13.2% over the best-performing single-stage model (YOLOv8x; Tables 3 and 4). The Swin Transformer classification stage achieved high precision and recall for all five species ($F1 \geq 0.85$; Table 2), and Grad-CAM interpretability analysis revealed that the model independently converged on many of the same morphological features used by expert taxonomists for species discrimination (Figure 2, Table 5). Deployed as an accessible web application, LizardLens directly supports the *Lizards on the Loose* community science initiative by providing immediate, probability-based species identifications that enhance both data quality and student engagement. Below, we discuss the ecological and methodological significance of these findings and outline directions for future development.

4.1. Advantages of task decoupling for ecological image classification

The enhanced performance of our two-stage pipeline relative to single-stage models highlights the benefits of decoupling detection from classification in tasks requiring fine-grained visual discrimination, particularly when organisms occur in visually heterogeneous environments or when species exhibit subtle morphological differences. Prior work in fine-grained recognition has shown that accurate classification often depends on isolating discriminative regions or parts, as global image features may obscure subtle inter-class variation. In single-stage architectures such as YOLOv8x and YOLOv12x, localization and classification are learned jointly within a single model architecture, which can limit the model's ability to focus on fine-scale features when targets are small relative to image dimensions or embedded in structurally complex backgrounds (Krause *et al.*, 2014; Li *et al.*, 2018; Dong, 2025). By contrast, our two-stage pipeline enables specialization of each component: the YOLOv8x detector is optimized to localize lizards independent of species identity, while the Swin Transformer classifier operates on cropped, centered representations of the target organism. This spatial decoupling reduces background interference and emphasizes diagnostic morphological features, thereby improving discrimination among visually similar species.

An additional advantage of the two-stage architecture is that it enables user interaction during the detection stage. Because localization and classification are decoupled, non-expert users can manually adjust bounding boxes even when they lack confidence in species identification. This interactive correction process allows users to improve model inputs without needing taxonomic expertise and is particularly valuable in community science and educational settings.

These advantages are particularly relevant for ecological applications, as community science photographs (e.g., iNaturalist) frequently depict small organisms in natural environments. *Anolis* lizards exemplify this challenge: they typically occupy a small proportion of a photograph and are photographed against backgrounds of bark, foliage, branches, and other structurally complex substrates that share color and textural properties with the organisms themselves, and that generate complex shadow patterns. The cropping step effectively increases the signal-to-noise ratio for the classifier by isolating the organism from its surroundings—a preprocessing strategy that mirrors the approach a field biologist would take when examining a photograph for species identification.

Notably, YOLOv12x—a more recent architecture incorporating attention-centric feature extraction designed to capture long-range spatial dependencies (Tian, Ye and Doermann, 2025)—underperformed the older YOLOv8x model across all metrics. This result suggests that architectural novelty and increased representational capacity do not necessarily translate into improved performance in ecological classification tasks that require subtle interspecific discrimination. The attention mechanisms in YOLOv12x may be better suited to tasks involving larger, more distinct objects rather than the subtle phenotypic differences characteristic of closely related species. This finding reinforces the principle that model selection for ecological applications should be guided by empirical performance on domain-specific datasets rather than general benchmark rankings (Horn *et al.*, 2021; Sterkenburg and Grünwald, 2021; Tredennick *et al.*, 2021).

4.2. Species-level variation in classification performance reflects morphological distinctiveness

LizardLens performed best for *A. distichus* (Bark anole; $F1 = 0.93$), a species characterized by a distinctive light-colored ocular ring and compact body proportions that distinguish it from all other Florida anoles. Similarly strong performance for *A. carolinensis* (Green anole; $F1 = 0.89$) and *A. equestris* (Knight anole; $F1 = 0.88$) likely reflects the morphological distinctiveness of these species: an elongated pointed snout and long tail in the former, and large body size with distinctive white shoulder stripes and pale perioral markings in the latter.

Conversely, the lowest performance was observed for *A. sagrei* (Brown anole; $F1 = 0.85$; [Table 4](#)), and the most frequent interspecific confusion ($n = 16$ misclassifications; [Figure 4](#)) occurred between *A. sagrei* and *A. cristatellus* (Crested anole). Rather than reflecting a simple model limitation, this confusion pattern mirrors the genuine difficulty that biology students—and occasionally lizard experts—experience when distinguishing these two species (Stroud pers. obs.). *Anolis sagrei* and *A. cristatellus* exhibit remarkable phenotypic similarity in body shape, overall coloration, and dorsal patterning, as well as semi-terrestrial habitat use (Hall *et al.*, 2020; Stroud *et al.*, 2020, 2024; Hall, Thawley and Stroud, 2024).

It is worth noting that *A. sagrei* also exhibits considerable sexual dimorphism in dorsal coloration and patterning, most pronounced among females (Moon and Kamath, 2019), which may further reduce classification precision by broadening the feature space the model must learn for this species. Future

work could investigate whether incorporating additional information on intraspecific pattern variation as an auxiliary input improves discrimination between *A. sagrei* and *A. cristatellus*.

4.3. Error modes and their ecological significance

Interspecific misclassification was not the dominant source of pipeline failure (n=39 images, 3.86% of test set); rather, detection-stage errors—missed detections (n=67 images, 6.63% of test set) and spurious detections of non-lizard objects (n = 81 images, 8.01% of test set)—accounted for the majority of errors. This hierarchy of error sources carries important ecological implications. The finding that distinguishing lizards from visually similar environmental features (sticks, bark, dappled light) represents a greater challenge than discriminating among *Anolis* species, which highlights the fundamental difficulty of organism detection in structurally complex habitats—a challenge that extends well beyond *Anolis* to any small-bodied organism that relies on crypsis or occupies visually heterogeneous microhabitats (Baling *et al.*, 2020; Yu, Lin and Xiao, 2024).

The two primary causes of missed detections—partial occlusion by vegetation (n=41 images, 4.06% of test set) and failure to resolve multiple proximate individuals (n=9 images, 0.89% of test set) —are both ecologically predictable. *Anolis* lizards frequently perch among branches and foliage where partial body concealment is common, and high population densities or behavioral interactions (e.g., reproductive behaviors) typically mean that conspecifics are frequently in close spatial proximity (Stroud *et al.*, 2023, 2024). The Non-Maximum Suppression (NMS) post-processing step, which suppresses overlapping bounding boxes above a defined IoU threshold, appears particularly problematic when conspecifics are in close proximity. However, the practical impact of this error mode for community science applications is minimal, as most community science platforms including iNaturalist require single-species photographic records, and all multi-individual images in our dataset contained conspecifics, so missing one of two individuals of the same species does not degrade observation quality.

Spurious detections are clustered at lower confidence (n = 81, 8.01% of test set, median 0.54, IQR 0.39–0.73) as compared to confidences of true positives (n = 896, 88.6% of test set, median 0.91, IQR 0.86–0.94) (Figure 6). We use a 0.3 threshold to retain nearly all true positives, including those in the low-confidence tail, while admitting only a small residual of false positives (8.01% of the test set). These residual false positives are addressed by the interactive review interface, in which users visually inspect each bounding box and dismiss spurious detections before submission. The spuriously detected objects—sticks, bark textures, dried leaves, and dappled lighting patterns—share low-level visual features (color, texture, elongated shape) with *Anolis* lizards, indicating that the detection model relies partially on these coarse features rather than exclusively on holistic morphological recognition. This parallels the broader challenge of automated wildlife detection in camera trap and field photography studies, where environmental features frequently trigger false positives (Tabak, Norouzzadeh, Wolfson, Sweeney, Vercauteren, Snow, Halseth, Di Salvo, J. S. Lewis, *et al.*, 2019; Maile, Duggan and Mousseau, 2023).

4.4. Model interpretability: convergence on biologically meaningful features

A key finding from our Grad-CAM analysis is that the Swin Transformer classification model independently learned to attend to morphological features that expert taxonomists use for *Anolis* species identification (Losos, 2009). The model attended to species-specific diagnostic features—including snout morphology, limb proportions, toepad size, ocular ring coloration, body patterning, and shoulder markings (Table 5)—that correspond closely to characters used by expert taxonomists for field identification (Krysko *et al.*, 2011; Stroud *et al.*, 2025). This convergence between algorithmic and expert feature use is notable because the model received no explicit guidance about which morphological characters are taxonomically informative; these features were learned entirely from the statistical structure of the training images. This result adds to a growing body of evidence that deep learning models can recover biologically meaningful signal from ecological image data (Akagi *et al.*, 2020; Wei *et al.*, 2022; Mostafa *et al.*, 2023), and it enhances confidence that the model’s predictions are grounded in taxonomically relevant morphology rather than spurious correlations with background features or image artifacts.

4.5. Applications for community science and conservation

LizardLens was designed not only as a classification tool but as an integrated platform that bridges machine learning with community science education and biodiversity monitoring. The web application’s design prioritizes accessibility for non-expert users through several features: probability-ranked species predictions allow users to evaluate classification certainty, interactive bounding box correction enables users to refine automated detections, and confidence-coded color visualization provides immediate visual feedback on prediction reliability. These features are particularly important for the Lizards on the Loose initiative, where middle school students serve as primary data collectors and may lack the taxonomic expertise to independently verify species identifications (Aristeidou *et al.*, 2021).

From a conservation perspective, the pipeline addresses a pressing need for high-quality occurrence data on invasive *Anolis* species in Florida. Four of the five species in our study (*A. sagrei*, *A. cristatellus*, *A. distichus*, and *A. equestris*) are non-native to Florida (Krysko *et al.*, 2011; Stroud *et al.*, 2019), and their ongoing range expansions have significant ecological implications for native species, including negative interactions with native *A. carolinensis* (Campbell, 2000; Stuart *et al.*, 2014; Kamath and Stuart, 2015; Bush, Ellison and Simberloff, 2022). Automated tools that improve the accuracy and volume of occurrence records for these species can directly inform management strategies and contribute to invasion biology research, such as more informed invasion range forecasting (Mothes *et al.*, 2019). The adjustable confidence and NMS thresholds in our pipeline allow deployment parameters to be tuned to specific use cases: lower confidence thresholds maximize detection sensitivity for engagement-focused educational settings, while higher thresholds reduce false positive rates for automated biodiversity monitoring pipelines where manual review is impractical. More broadly, by providing immediate feedback on species identifications, *LizardLens* transforms passive data collection into an interactive learning experience in which students engage directly with taxonomic decision-making—demonstrating

that research-quality biodiversity data and authentic scientific experiences for young students need not be competing objectives (Shah and Martinez, 2016).

4.6. Limitations

Two limitations of this study should be acknowledged. First, our dataset is restricted to five *Anolis* species in Florida, and pipeline performance may not generalize to other geographic regions, taxonomic groups, or species assemblages. Second, our training data were sourced exclusively from iNaturalist “Research Grade” observations, which represent a potentially biased subsection of photographs that exclude low-quality images in which iNaturalist users were unable to collectively agree on a species’ identity. Therefore, *LizardLens* performance may degrade on images that differ substantially from those used for model training.

4.7. Future directions

The two-stage pipeline framework demonstrated here is readily extensible to other ecological classification challenges involving small-bodied organisms in complex habitats. For example, automated identification of frogs, insects, small birds, and small mammals from field photographs and camera trap imagery, where the same challenges of small target size, complex backgrounds, and fine-grained interspecific similarity apply (Schneider, Taylor and Kremer, 2018; Tabak, Norouzzadeh, Wolfson, Sweeney, Vercauteren, Snow, Halseth, Di Salvo, J. Lewis, *et al.*, 2019; Bjerger *et al.*, 2023). There are also several specific avenues for improving *LizardLens*. First, expanding the species set to include additional common non-native lizard species in Florida (e.g., *Agama picticauda*, *Basiliscus vittatus*, *Leiocephalus carinatus*, *Hemidactylus* sp., *Iguana iguana*, and *Ctenosaura similis*) would increase the pipeline’s utility for comprehensive biodiversity monitoring in the region (Krysko, Enge and Moler, 2019; Mothes *et al.*, 2019).

Integrating geographic locality data as auxiliary model input represents a particularly promising direction for improving discrimination between morphologically similar species with partially non-overlapping ranges, such as *A. sagrei* and *A. cristatellus* (Kolbe *et al.*, 2016; Stroud *et al.*, 2024). Developing edge-deployed versions of the pipeline for offline use would expand accessibility to field settings with limited connectivity, while direct integration with community science platforms such as iNaturalist could streamline data submission workflows. Similarly, important insights on ecological interactions may be derived from iNaturalist images if tools are able to identify multiple species or predator-prey interactions, which are documented within our five species study system (Stroud, 2013; Ljustina and Stroud, 2016).

Finally, formal user studies with student participants in the *Lizards on the Loose* program would provide valuable data on educational outcomes, user experience, and the extent to which interactive machine learning tools enhance taxonomic learning and engagement with biodiversity science.

5. Conclusion

LizardLens demonstrates that decoupling object detection from species classification into a two-stage pipeline yields substantial performance improvements for fine-grained identification of morphologically similar organisms in visually complex field photographs. By achieving robust classification accuracy across five *Anolis* species while learning to attend to biologically meaningful diagnostic features, our pipeline bridges the gap between machine learning and expert taxonomic knowledge. The integration of this technical capability into an accessible, user-centered web application illustrates how computational tools can simultaneously enhance biodiversity data quality and create authentic scientific learning experiences for community science participants. As automated identification tools become increasingly central to biodiversity monitoring at scale, approaches that combine high classification performance with interpretability and accessibility will be essential for maintaining the scientific rigor and public engagement that underpin effective community science and conservation.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and Code Availability Statement

The curated iNaturalist image dataset supporting this study, including bounding box annotations and train, validation and test splits, is archived in the Dryad Digital Repository (<https://doi.org/10.5061/dryad.xksn02vx2>).

All code for model training, evaluation, Grad-CAM analysis, and the deployed web application is archived on Zenodo (<https://doi.org/10.5281/zenodo.20366694>) and developed openly at https://github.com/wenhan9604/Anole_classifier. The trained model weights are included in the Zenodo archive.

Acknowledgements

This work was supported by the David and Lucille Packard Foundation and the Maxwell/Hanrahan Foundation (both to J.T.S.). We thank Nabil Abdullah for assistance during the early development of this project. We also thank the Georgia Tech College of Computing for hosting and supporting the web application resources used in this study. This research was supported in part through research cyberinfrastructure resources and services provided by the Partnership for an Advanced Computing Environment (PACE) at the Georgia Institute of Technology, Atlanta, Georgia, USA (RRID:SCR_027619).

References

- Akagi, T. *et al.* (2020) "Explainable Deep Learning Reproduces a 'Professional Eye' on the Diagnosis of Internal Disorders in Persimmon Fruit," *Plant and Cell Physiology*, 61(11), pp. 1967–1973. Available at: <https://doi.org/10.1093/pcp/pcaa111>.
- Aristeidou, M. *et al.* (2021) "How Do Young Community and Citizen Science Volunteers Support Scientific Research on Biodiversity? The Case of iNaturalist," *Diversity*, 13(7), p. 318. Available at: <https://doi.org/10.3390/d13070318>.
- Baling, M. *et al.* (2020) "Spatial and temporal variation in prey color patterns for background matching across a continuous heterogeneous environment," *Ecology and Evolution*, 10(5), pp. 2310–2319. Available at: <https://doi.org/10.1002/ece3.6024>.
- Binley, A.D. *et al.* (2023) "Patterns of community science data use in peer-reviewed research on biodiversity," *Biological Conservation*, 280, p. 109985. Available at: <https://doi.org/10.1016/j.biocon.2023.109985>.
- Bjerge, K. *et al.* (2023) "Accurate detection and identification of insects from camera trap images with deep learning," *PLOS Sustainability and Transformation*, 2(3), p. e0000051. Available at: <https://doi.org/10.1371/journal.pstr.0000051>.
- Bush, J.M., Ellison, M. and Simberloff, D. (2022) "Impacts of an invasive species (*Anolis sagrei*) on social and spatial behaviours of a native congener (*Anolis carolinensis*)," *Animal Behaviour*, 183, pp. 177–188. Available at: <https://doi.org/10.1016/j.anbehav.2021.06.006>.
- Callaghan, C.T. *et al.* (2022) "The benefits of contributing to the citizen science platform iNaturalist as an identifier," *PLOS Biology*, 20(11), p. e3001843. Available at: <https://doi.org/10.1371/journal.pbio.3001843>.
- Campbell, T. (2000) "Analyses of the Effects of an Exotic Lizard (*Anolis sagrei*) on a Native Lizard (*Anolis carolinensis*) in Florida, Using Islands as Experimental Units." Available at: [https://www.semanticscholar.org/paper/Analyses-of-the-Effects-of-an-Exotic-Lizard-\(-\)-on-Campbell/96e219426292cbd06d4f904151d4113ddbe0c047](https://www.semanticscholar.org/paper/Analyses-of-the-Effects-of-an-Exotic-Lizard-(-)-on-Campbell/96e219426292cbd06d4f904151d4113ddbe0c047) (Accessed: April 19, 2026).
- Chandler, M. *et al.* (2017) "Contribution of citizen science towards international biodiversity monitoring," *Biological Conservation*, 213, pp. 280–294. Available at: <https://doi.org/10.1016/j.biocon.2016.09.004>.
- Chattopadhyay, A. *et al.* (2018) "Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks," *2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 839–847. Available at: <https://doi.org/10.1109/WACV.2018.00097>.
- Coca-de-la-Iglesia, M. *et al.* (2024) "High rate of species misidentification reduces the taxonomic certainty of European biodiversity databases of ivies (*Hedera L.*)," *Scientific Reports*, 14(1), p. 4876. Available at: <https://doi.org/10.1038/s41598-024-54735-0>.
- Deng, J. *et al.* (2009) "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255. Available at: <https://doi.org/10.1109/CVPR.2009.5206848>.
- Di Cecco, G.J. *et al.* (2021) "Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science," *BioScience*, 71(11), pp. 1179–1188. Available at: <https://doi.org/10.1093/biosci/biab093>.

- 817 Dong, Z. (2025) "A Review of Fine-Grained Image Recognition Techniques Based on Deep Learning." *International*
818 *Conference on Data Science and Engineering*, SCITEPRESS, pp. 188–193. Available at:
819 <https://doi.org/10.5220/0013680800004670>.
- 820 Dosovitskiy, A. *et al.* (2021) "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv.
821 Available at: <https://doi.org/10.48550/arXiv.2010.11929>.
- 822 Dwyer, B., Nelson, J. and Hansen, T. (2026) "Roboflow." Available at: <https://roboflow.com> (Accessed: April 25,
823 2026).
- 824 Encarnação, J., Teodósio, M.A. and Morais, P. (2021) "Citizen Science and Biological Invasions: A Review," *Frontiers*
825 *in Environmental Science*, 8. Available at: <https://doi.org/10.3389/fenvs.2020.602980>.
- 826 *Export Observations, iNaturalist* (2026) *iNaturalist*. Available at: <https://www.inaturalist.org/observations/export>
827 (Accessed: April 25, 2026).
- 828 Hall, J.M. *et al.* (2020) "Adaptive seasonal shift towards investment in fewer, larger offspring: Evidence from field
829 and laboratory studies," *Journal of Animal Ecology*, 89(5), pp. 1242–1253. Available at:
830 <https://doi.org/10.1111/1365-2656.13182>.
- 831 Hall, J.M., Thawley, C.J. and Stroud, J.T. (2024) "Life-history strategies are decoupled from ecomorphological
832 convergence in two Anolis lizards," *Biological Journal of the Linnean Society*, 143(2). Available at:
833 <https://doi.org/10.1093/biolinnea/blae088>.
- 834 Hooykaas, M.J.D. *et al.* (2019) "Identification skills in biodiversity professionals and laypeople: A gap in species
835 literacy," *Biological Conservation*, 238, p. 108202. Available at: <https://doi.org/10.1016/j.biocon.2019.108202>.
- 836 Horn, G.V. *et al.* (2021) "Benchmarking Representation Learning for Natural World Image Collections." arXiv.
837 Available at: <https://doi.org/10.48550/arXiv.2103.16483>.
- 838 Jäckel, D. *et al.* (2023) "Community engagement and data quality: best practices and lessons learned from a citizen
839 science project on birdsong," *Journal of Ornithology*, 164(1), pp. 233–244. Available at:
840 <https://doi.org/10.1007/s10336-022-02018-8>.
- 841 Jaskuła, R. *et al.* (2021) "From Phenology and Habitat Preferences to Climate Change: Importance of Citizen
842 Science in Studying Insect Ecology in the Continental Scale with American Red Flat Bark Beetle, *Cucujus clavipes*, as
843 a Model Species," *Insects*, 12(4), p. 369. Available at: <https://doi.org/10.3390/insects12040369>.
- 844 Jennett, C. *et al.* (2016) "Motivations, learning and creativity in online citizen science," *Journal of Science*
845 *Communication*, 15(03). Available at: <https://doi.org/10.22323/2.15030205>.
- 846 Jocher, G., Qiu, J. and Chaurasia, A. (2023) "Ultralytics." Available at: <https://www.ultralytics.com/> (Accessed: April
847 25, 2026).
- 848 Kamath, A. and Stuart, Y.E. (2015) "Movement Rates of the Lizard *Anolis carolinensis* (Squamata: Dactyloidae) in
849 the Presence and Absence of *Anolis sagrei* (Squamata: Dactyloidae)," *Breviora*, 546(1), pp. 1–7. Available at:
850 <https://doi.org/10.3099/0006-9698-546.00.1>.
- 851 Kelling, S. *et al.* (2019) "Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity,"
852 *BioScience*, 69(3), pp. 170–179. Available at: <https://doi.org/10.1093/biosci/biz010>.
- 853 Khanam, R. and Hussain, M. (2024) "YOLOv11: An Overview of the Key Architectural Enhancements." arXiv.
854 Available at: <https://doi.org/10.48550/arXiv.2410.17725>.

- 855 Kolbe, J.J. *et al.* (2016) "Determinants of spread in an urban landscape by an introduced lizard," *Landscape Ecology*,
856 31(8), pp. 1795–1813. Available at: <https://doi.org/10.1007/s10980-016-0362-1>.
- 857 Koo, K.-S. *et al.* (2022) "Assessing the Accuracy of Citizen Science Data Based on iNaturalist Data," *Diversity*, 14(5),
858 p. 316. Available at: <https://doi.org/10.3390/d14050316>.
- 859 Kosmala, M. *et al.* (2016) "Assessing data quality in citizen science," *Frontiers in Ecology and the Environment*,
860 14(10), pp. 551–560. Available at: <https://doi.org/10.1002/fee.1436>.
- 861 Krause, J. *et al.* (2014) "Learning Features and Parts for Fine-Grained Recognition," *Proceedings of the 2014 22nd*
862 *International Conference on Pattern Recognition*. USA: IEEE Computer Society (ICPR '14), pp. 26–33. Available at:
863 <https://doi.org/10.1109/ICPR.2014.15>.
- 864 Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) "ImageNet Classification with Deep Convolutional Neural
865 Networks," *Advances in Neural Information Processing Systems*. Curran Associates, Inc. Available at:
866 https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html
867 (Accessed: April 25, 2026).
- 868 Krysko, K.L. *et al.* (2011) "<p>Verified non-indigenous amphibians and reptiles in Florida from 1863
869 through 2010: Outlining the invasion process and identifying invasion pathways and stages</p>,"
870 *Zootaxa*, 3028(1), pp. 1–64. Available at: <https://doi.org/10.11646/zootaxa.3028.1.1>.
- 871 Krysko, K.L., Enge, K.M. and Moler, P.E. (2019) *Amphibians and reptiles of Florida*. University of Florida Press.
- 872 Li, W. *et al.* (2018) "Region based Ensemble Learning Network for Fine-grained Classification," *2018 Chinese*
873 *Automation Congress (CAC)*, pp. 4173–4177. Available at: <https://doi.org/10.1109/CAC.2018.8623687>.
- 874 Liu, S. *et al.* (2024) "Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection."
875 arXiv. Available at: <https://doi.org/10.48550/arXiv.2303.05499>.
- 876 Liu, Z. *et al.* (2021) "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows." arXiv. Available at:
877 <https://doi.org/10.48550/arXiv.2103.14030>.
- 878 *Lizard Lens* (2026) *Lizard Lens*. Available at: <https://lizard-class.cc.gatech.edu/> (Accessed: April 25, 2026).
- 879 Ljustina, O. and Stroud, J.T. (2016) "Anolis equestris (Cuban knight anole): novel predator–prey interaction,"
880 *Herpetological Review*, 47(3), pp. 459–460.
- 881 Losos, J.B. (2009) *Lizards in an evolutionary tree: ecology and adaptive radiation of anoles*. Berkeley: University of
882 California Press (Organisms and environments, No. 10).
- 883 Maile, R.E., Duggan, M.T. and Mousseau, T.A. (2023) "The successes and pitfalls: Deep-learning effectiveness in a
884 Chernobyl field camera trap application," *Ecology and Evolution*, 13(9), p. e10454. Available at:
885 <https://doi.org/10.1002/ece3.10454>.
- 886 Mason, B.M. *et al.* (2025) "iNaturalist accelerates biodiversity research," *BioScience*, 75(11), pp. 953–965. Available
887 at: <https://doi.org/10.1093/biosci/biaf104>.
- 888 Mesaglio, T. *et al.* (2025) "Expert identification blitz: A rapid high value approach for assessing and improving
889 iNaturalist identification accuracy and data precision and confidence," *PLANTS, PEOPLE, PLANET*, 7(5), pp. 1469–
890 1484. Available at: <https://doi.org/10.1002/ppp3.70005>.

891 Moon, R.M. and Kamath, A. (2019) "Re-examining escape behaviour and habitat use as correlates of dorsal pattern
892 variation in female brown anole lizards, *Anolis sagrei* (Squamata: Dactyloidae)," *Biological Journal of the Linnean*
893 *Society*, 126(4), pp. 783–795. Available at: <https://doi.org/10.1093/biolinnean/blz006>.

894 Mostafa, S. *et al.* (2023) "Explainable deep learning in plant phenotyping," *Frontiers in Artificial Intelligence*, 6.
895 Available at: <https://doi.org/10.3389/frai.2023.1203546>.

896 Mothes, C.C. *et al.* (2019) "Evaluating ecological niche model accuracy in predicting biotic invasions using South
897 Florida's exotic lizard community," *Journal of Biogeography*, 46(2), pp. 432–441. Available at:
898 <https://doi.org/10.1111/jbi.13511>.

899 Mulero-Pázmány, M. *et al.* (2025a) "Addressing significant challenges for animal detection in camera trap images:
900 a novel deep learning-based approach," *Scientific Reports*, 15(1), p. 16191. Available at:
901 <https://doi.org/10.1038/s41598-025-90249-z>.

902 Mulero-Pázmány, M. *et al.* (2025b) "Addressing significant challenges for animal detection in camera trap images:
903 a novel deep learning-based approach," *Scientific Reports*, 15(1), p. 16191. Available at:
904 <https://doi.org/10.1038/s41598-025-90249-z>.

905 Murat, A.A. and Kiran, M.S. (2025) "A comprehensive review on YOLO versions for object detection," *Engineering*
906 *Science and Technology, an International Journal*, 70, p. 102161. Available at:
907 <https://doi.org/10.1016/j.jestch.2025.102161>.

908 Padilla, R. *et al.* (2021) "A Comparative Analysis of Object Detection Metrics with a Companion Open-Source
909 Toolkit," *Electronics*, 10(3), p. 279. Available at: <https://doi.org/10.3390/electronics10030279>.

910 Powers, D.M.W. (2020) "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and
911 correlation." arXiv. Available at: <https://doi.org/10.48550/arXiv.2010.16061>.

912 Price, S. *et al.* (2025) "Using large-scale community science data and computer vision to evaluate thermoregulation
913 as an adaptive driver of physiological color change in *Anolis carolinensis*," *Frontiers in Zoology*, 22(1), p. 31.
914 Available at: <https://doi.org/10.1186/s12983-025-00580-4>.

915 Queiruga-Dios, M.Á. *et al.* (2020) "Citizen Science for Scientific Literacy and the Attainment of Sustainable
916 Development Goals in Formal Education," *Sustainability*, 12(10), p. 4283. Available at:
917 <https://doi.org/10.3390/su12104283>.

918 Ramírez, S. (2026) "FastAPI." Available at: <https://github.com/fastapi/fastapi> (Accessed: April 25, 2026).

919 Redmon, J. *et al.* (2016) "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on*
920 *Computer Vision and Pattern Recognition (CVPR)*. *2016 IEEE Conference on Computer Vision and Pattern*
921 *Recognition (CVPR)*, pp. 779–788. Available at: <https://doi.org/10.1109/CVPR.2016.91>.

922 Ren, S. *et al.* (2016) "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." arXiv.
923 Available at: <https://doi.org/10.48550/arXiv.1506.01497>.

924 Schneider, S., Taylor, G.W. and Kremer, S.C. (2018) "Deep Learning Object Detection Methods for Ecological
925 Camera Trap Data." arXiv. Available at: <https://doi.org/10.48550/arXiv.1803.10842>.

926 Selvaraju, R.R. *et al.* (2017) "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,"
927 *2017 IEEE International Conference on Computer Vision (ICCV)*. *2017 IEEE International Conference on Computer*
928 *Vision (ICCV)*, pp. 618–626. Available at: <https://doi.org/10.1109/ICCV.2017.74>.

- 929 Shah, H.R. and Martinez, L.R. (2016) "Current Approaches in Implementing Citizen Science in the Classroom,"
930 *Journal of Microbiology & Biology Education*, 17(1), pp. 17–22. Available at:
931 <https://doi.org/10.1128/jmbe.v17i1.1032>.
- 932 Shen, L. *et al.* (2025) "TransFGVC: transformer-based fine-grained visual classification," *The Visual Computer*, 41(4),
933 pp. 2439–2459. Available at: <https://doi.org/10.1007/s00371-024-03545-6>.
- 934 Sokolova, M. and Lapalme, G. (2009) "A systematic analysis of performance measures for classification tasks,"
935 *Information Processing & Management*, 45(4), pp. 427–437. Available at:
936 <https://doi.org/10.1016/j.ipm.2009.03.002>.
- 937 Sterkenburg, T.F. and Grünwald, P.D. (2021) "The no-free-lunch theorems of supervised learning," *Synthese*,
938 199(3–4), pp. 9979–10015. Available at: <https://doi.org/10.1007/s11229-021-03233-1>.
- 939 Stroud, J.T. (2013) "Exotic intraguild predation: *Anolis equestris* (Cuban Knight Anole) and *Anolis distichus*
940 (Hispaniolian Bark Anole).," *Herpetological Review*, 44, p. 4.
- 941 Stroud, J.T. *et al.* (2019) "Behavioral shifts with urbanization may facilitate biological invasion of a widespread
942 lizard," *Urban Ecosystems*, 22(3), pp. 425–434. Available at: <https://doi.org/10.1007/s11252-019-0831-9>.
- 943 Stroud, J.T. *et al.* (2020) "An extreme cold event leads to community-wide convergence in lower temperature
944 tolerance in a lizard community," *Biology Letters*, 16(10), p. 20200625. Available at:
945 <https://doi.org/10.1098/rsbl.2020.0625>.
- 946 Stroud, J.T. *et al.* (2023) "Fluctuating selection maintains distinct species phenotypes in an ecological community in
947 the wild," *Proceedings of the National Academy of Sciences*, 120(42), p. e2222071120. Available at:
948 <https://doi.org/10.1073/pnas.2222071120>.
- 949 Stroud, J.T. *et al.* (2024) "Observing character displacement from process to pattern in a novel vertebrate
950 community," *Nature Communications*, 15(1), p. 9862. Available at: <https://doi.org/10.1038/s41467-024-54302-1>.
- 951 Stroud, J.T. *et al.* (2025) "Pirates of the Caribbean (and Elsewhere): Three-Legged Lizards and the Study of
952 Evolutionary Adaptation," *The American Naturalist*, 206(5), pp. 403–417. Available at:
953 <https://doi.org/10.1086/737525>.
- 954 Stroud, J.T. and Losos, J.B. (2020) "Bridging the process-pattern divide to understand the origins and early stages of
955 adaptive radiation: a review of approaches with insights from studies of *Anolis* lizards," *Journal of Heredity*, 111(1),
956 pp. 33–42.
- 957 Stuart, Y.E. *et al.* (2014) "Rapid evolution of a native species following invasion by a congener," *Science*, 346(6208),
958 pp. 463–466. Available at: <https://doi.org/10.1126/science.1257008>.
- 959 Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Halseth, J.M., Di
960 Salvo, P.A., Lewis, J.S., *et al.* (2019) "Machine learning to classify animal species in camera trap images:
961 Applications in ecology," *Methods in Ecology and Evolution*. Edited by T. Photopoulou, 10(4), pp. 585–590.
962 Available at: <https://doi.org/10.1111/2041-210X.13120>.
- 963 Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Vercauteren, K.C., Snow, N.P., Halseth, J.M., Di
964 Salvo, P.A., Lewis, J., *et al.* (2019) "Machine learning to classify animal species in camera trap images: Applications
965 in ecology," *Methods in Ecology and Evolution*, 10(4), pp. 585–590. Available at: <https://doi.org/10.1111/2041-210X.13120>.
- 966

967 Thawley, C. and Stroud, J. (2026) *Lizards On The Loose! Anole Lizards of South Florida*, iNaturalist, iNaturalist.
968 Available at: <https://www.inaturalist.org/guides/5439> (Accessed: April 25, 2026).

969 Tian, Y., Ye, Q. and Doermann, D. (2025) "YOLOv12: Attention-Centric Real-Time Object Detectors." arXiv. Available
970 at: <https://doi.org/10.48550/arXiv.2502.12524>.

971 Touvron, H. *et al.* (2021) "Training data-efficient image transformers & distillation through attention." arXiv.
972 Available at: <https://doi.org/10.48550/arXiv.2012.12877>.

973 Tredennick, A.T. *et al.* (2021) "A practical guide to selecting models for exploration, inference, and prediction in
974 ecology," *Ecology*, 102(6), p. e03336. Available at: <https://doi.org/10.1002/ecy.3336>.

975 Van Horn, G. *et al.* (2018) "The iNaturalist Species Classification and Detection Dataset," *2018 IEEE/CVF Conference*
976 *on Computer Vision and Pattern Recognition. 2018 IEEE/CVF Conference on Computer Vision and Pattern*
977 *Recognition*, pp. 8769–8778. Available at: <https://doi.org/10.1109/CVPR.2018.00914>.

978 Varghese, R. and M., S. (2024) "YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and
979 Robustness," *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems*
980 *(ADICS). 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems*
981 *(ADICS)*, pp. 1–6. Available at: <https://doi.org/10.1109/ADICS58448.2024.10533619>.

982 Wei, K. *et al.* (2022) "Explainable Deep Learning Study for Leaf Disease Classification," *Agronomy*, 12(5), p. 1035.
983 Available at: <https://doi.org/10.3390/agronomy12051035>.

984 White, E. *et al.* (2023) "Quantifying error in occurrence data: Comparing the data quality of iNaturalist and digitized
985 herbarium specimen data in flowering plant families of the southeastern United States," *PLOS ONE*, 18(12), p.
986 e0295298. Available at: <https://doi.org/10.1371/journal.pone.0295298>.

987 Wolf, T. *et al.* (2020) "HuggingFace's Transformers: State-of-the-art Natural Language Processing." arXiv. Available
988 at: <https://doi.org/10.48550/arXiv.1910.03771>.

989 Wu, B. *et al.* (2020) "Visual Transformers: Token-based Image Representation and Processing for Computer
990 Vision," *arXiv e-prints*, p. arXiv:2006.03677. Available at: <https://doi.org/10.48550/arXiv.2006.03677>.

991 Yu, H., Lin, Z. and Xiao, F. (2024) "Role of body size and shape in animal camouflage," *Ecology and Evolution*, 14(5),
992 p. e11434. Available at: <https://doi.org/10.1002/ece3.11434>.

993 Zhao, Y. *et al.* (2024) "DETRs Beat YOLOs on Real-time Object Detection." arXiv. Available at:
994 <https://doi.org/10.48550/arXiv.2304.08069>.

995

996

997

Table 1. Dataset composition by species. Images were curated to 2,000 per species and partitioned into training (80%; n = 1600), validation (10%; n = 200), and test (10%; n = 200) image subsets. The number of lizards detected (Lizards detected Total) exceeds the number of images analyzed (Images Total) because some images contain more than one anole in them (all images contain at least one). Training, Validation, and Test columns report lizard detection counts per split.

Dataset	Images (Total)	Lizards detected (Total)	Lizards detected in Training images	Lizards detected in Validation images	Lizards detected in Test images
Bark Anole	2000	2018	1614	202	202
Brown Anole	2000	2009	1606	203	200
Crested Anole	2000	2018	1609	205	204
Green Anole	2000	2041	1636	202	203
Knight Anole	2000	2028	1625	201	202
Total	10000	10114	8090	1013	1011

Table 2: The performance of each classification model fine-tuned with a dataset consisting of cropped anole lizards detected in test images. The precision, recall and F1-score of each anole class were reported for each model.

Metric		DeiT	ViT	Swin Transformer
Parameters (millions)		87m	86.9m	88m
Bark Anole	P	0.945	0.951	0.944
	R	0.855	0.870	0.925
	F1	0.900	0.909	0.934
Brown Anole	P	0.846	0.879	0.891
	R	0.905	0.910	0.940
	F1	0.874	0.894	0.915
Crested Anole	P	0.833	0.833	0.928
	R	0.870	0.900	0.895
	F1	0.851	0.865	0.911
Green Anole	P	0.959	0.969	0.956
	R	0.935	0.940	0.970
	F1	0.947	0.954	0.963
Knight Anole	P	0.959	0.975	0.980
	R	0.935	0.975	0.965
	F1	0.947	0.975	0.972
Overall	Top-1 Accuracy	0.907	0.919	0.939
	P	0.908	0.921	0.940
	R	0.900	0.919	0.939
	F1	0.904	0.919	0.939

Table 3. Performance comparison of the LizardLens two-stage pipeline against single-stage YOLOv8x and YOLOv12x models on the test dataset (n = 1,011, Table 1). Asterisks (*) indicate best model performance for each metric. The final column represents the relative performance improvement of LizardLens over the next-best model.

Performance metric	YOLOv8x	YOLOv12x	LizardLens	LizardLens improvement
Parameters	68.2M	59.1M	156.2M	-
Top-1 Accuracy	0.740	0.680	0.830*	+12.2%
F1-score	0.802	0.724	0.890*	+11.0%
Precision	0.798	0.736	0.882*	+10.5%
Recall	0.806	0.720	0.896*	+11.2%
mAP@50	0.794	0.676	0.891*	+12.2%
mAP@75	0.679	0.580	0.758*	+11.6%
mAP@50–95	0.622	0.533	0.704*	+13.2%

Table 4. Species-level performance comparison of the LizardLens two-stage pipeline against single-stage YOLOv8x and YOLOv12x models on the test dataset (n = 1,011; Table 1). Asterisks (*) indicate best model performance for each metric. Precision, recall, and F1-score are represented as P, R, and F1 respectively. The overall precision, recall, and F1-scores are calculated using macro-average.

Species	Metric	YOLOv8x	YOLOv12x	LizardLens
Bark Anole	P	0.790	0.780	0.930*
	R	0.800	0.710	0.930*
	F1	0.790	0.740	0.930*
Brown Anole	P	0.820	0.730	0.850*
	R	0.740	0.630	0.850*
	F1	0.780	0.680	0.850*
Crested Anole	P	0.810	0.690	0.870*
	R	0.830	0.740	0.920*
	F1	0.820	0.710	0.900*
Green Anole	P	0.810	0.760	0.890*
	R	0.820	0.720	0.890*
	F1	0.820	0.740	0.890*
Knight Anole	P	0.760	0.720	0.870*
	R	0.840	0.800	0.890*
	F1	0.800	0.750	0.880*
Overall	P	0.798	0.736	0.882*
	R	0.806	0.720	0.896*
	F1	0.802	0.724	0.890*

Table 5. Diagnostic features used by LizardLens identify *Anolis* lizard species.

Species	Grad-CAM diagnostic features
Bark anole (<i>A. distichus</i>)	Light-colored ocular ring; limb lengths
Brown anole (<i>A. sagrei</i>)	Variable dorsal body patterning; large hindfeet and hind toes
Crested anole (<i>A. cristatellus</i>)	Hindlimb length; toepad size
Green anole (<i>A. carolinensis</i>)	Elongated (pointed) snout; long tail; short hindlimbs
Knight anole (<i>A. equestris</i>)	Distinctive white shoulder stripe; pale perioral markings

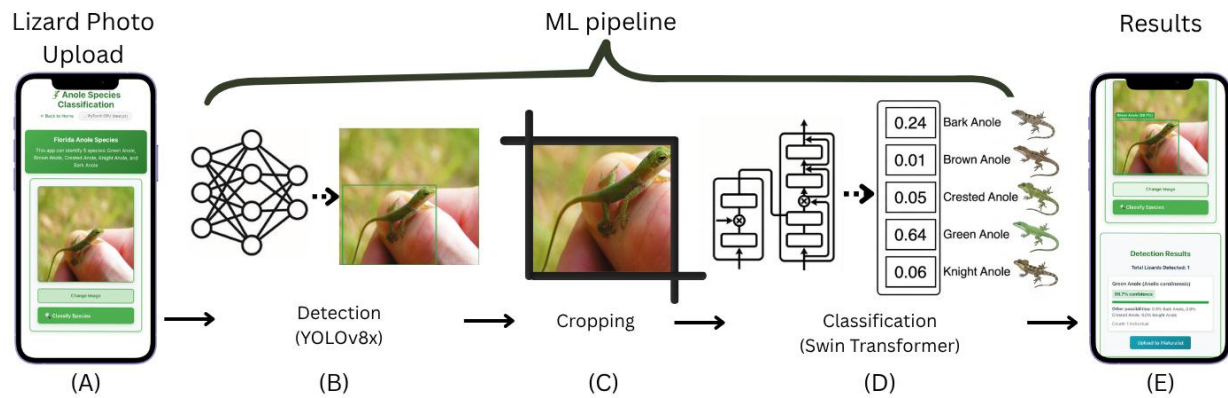


Figure 1. End-to-end web application workflow. (A) Users upload a field image. (B) The backend detects anoles and produces bounding boxes. (C) Cropped lizard detections are passed to the species classifier. (D) The Swin Transformer performs species classification. (E) The frontend visualizes bounding boxes and per-lizard detection confidence scores before optionally submitting validated observations to iNaturalist.

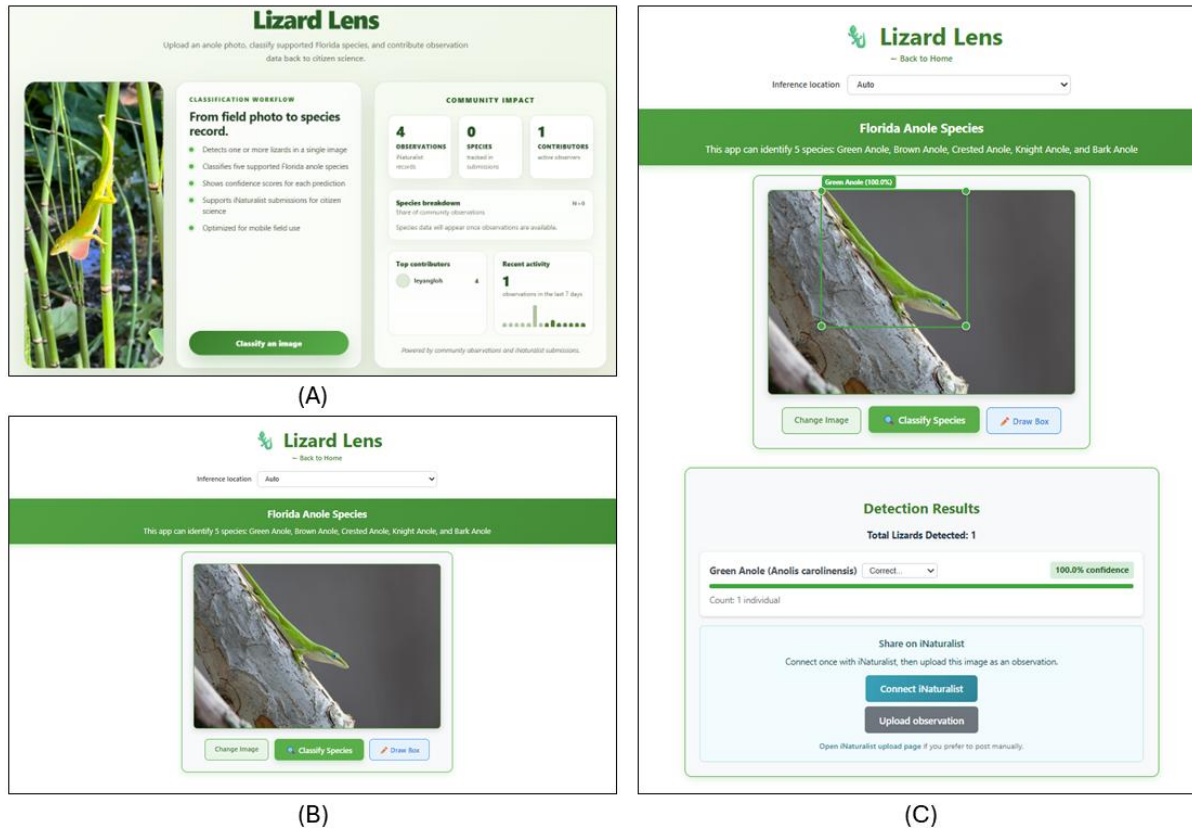


Figure 2. Screenshots of the LizardLens web application. (A) Homepage of the web application – a user-friendly interface that displays recent iNaturalist submissions made through the application. (B) Displays the captured image and options to change image, classify the species or manually annotate the bounding box. (C) The LizardLens prediction of the anole bounding box, confidence score, number of species and species classification are displayed. User can edit the predicted bounding box or choose to upload this observation to iNaturalist.

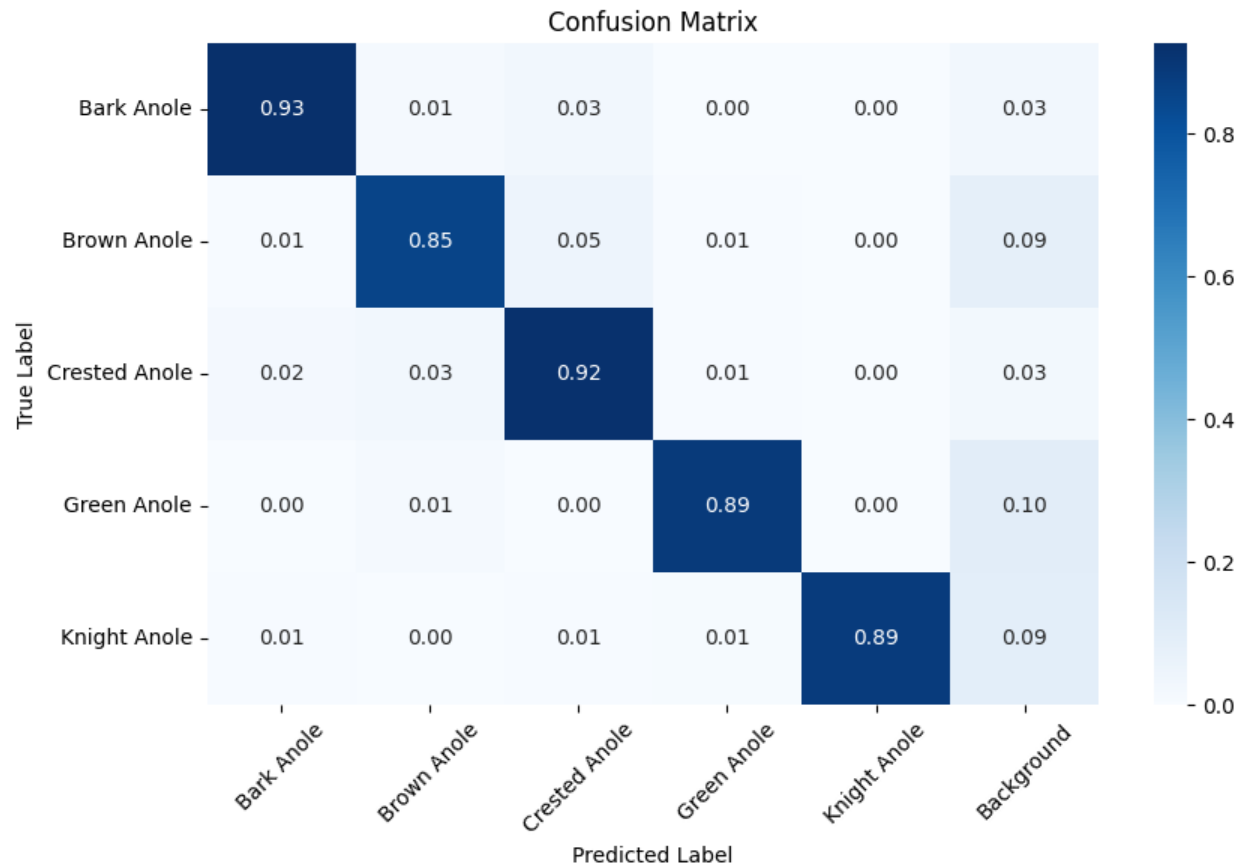


Figure 3. The normalized confusion matrix of the classification pipeline predictions of each anole species.

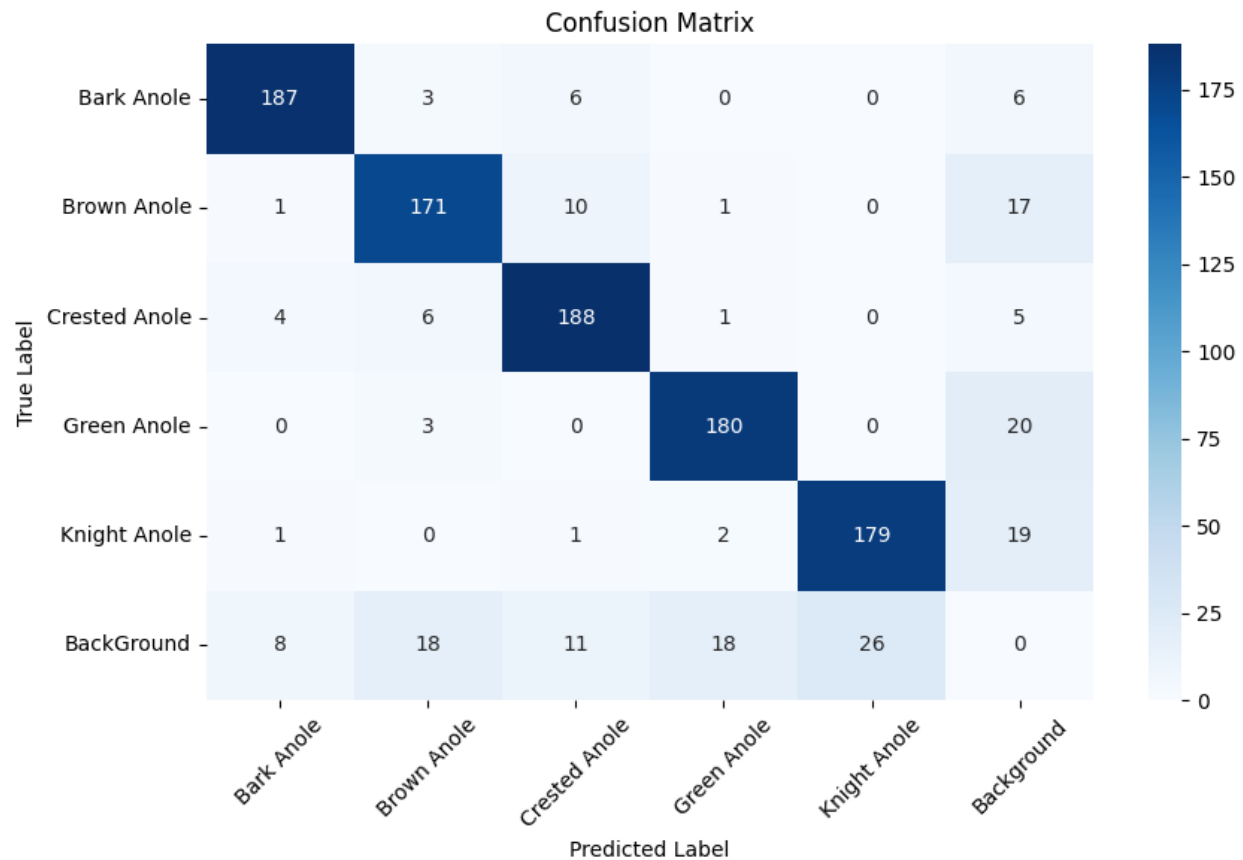


Figure 4. Confusion matrix, in absolute counts, of the classification pipeline's species predictions for each anole class. Out of 1,011 anole lizard detections in the test image dataset, 81 errors (8.01%) were false positive detections, 67 errors (6.63%) were missed detections, and 39 errors (3.86%) were species misclassifications among correctly detected lizards.

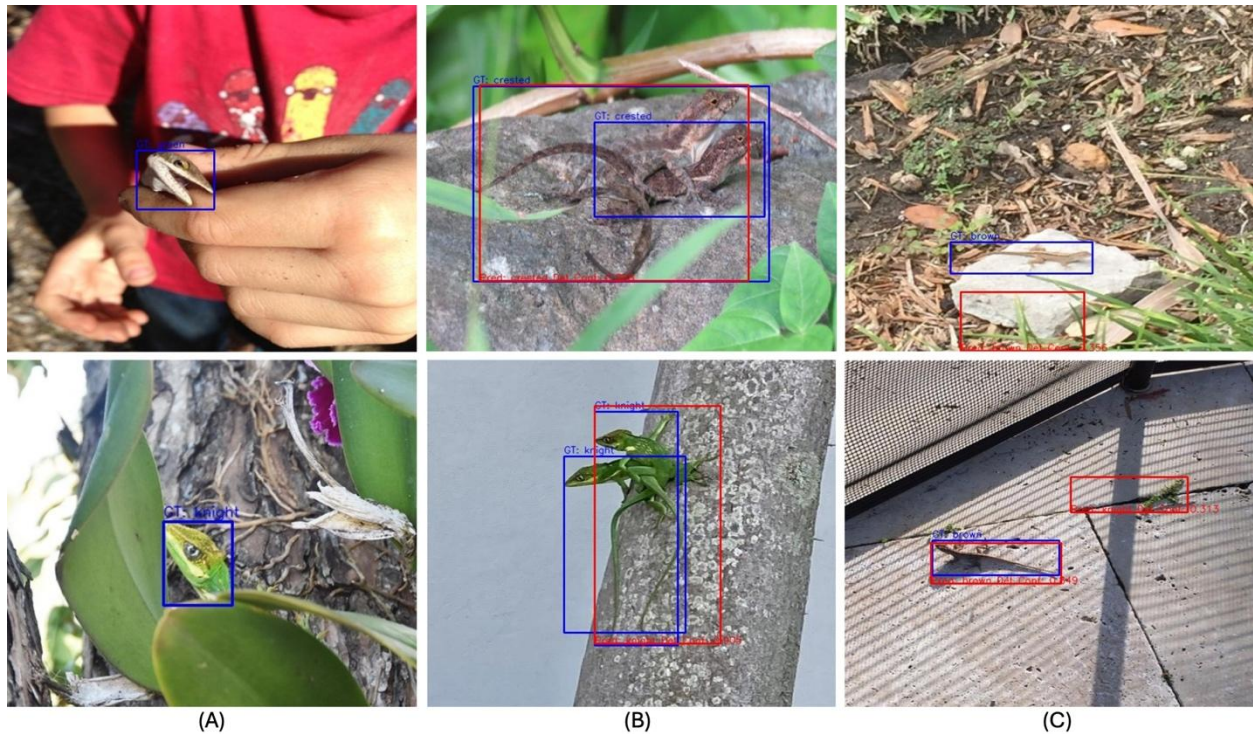


Figure 5. Representative error modes in the detection model. (A) False-negative due to partial occlusion (n=41 images, 4.06% of test set): the model fails to detect a lizard when only a small portion of the body (e.g., head) is visible and obscured by surrounding vegetation or substrate. (B) False-negative due to multiple proximate individuals (n=9 images, 0.89% of test set): the model generates a single bounding box encompassing two closely positioned lizards rather than detecting each individual separately. (C) False-positive detection: the model incorrectly identifies non-lizard objects (e.g., sticks, bark, or lighting patterns) as lizards, often with low confidence. Blue boxes indicate predicted detections, and red boxes indicate ground truth annotations.

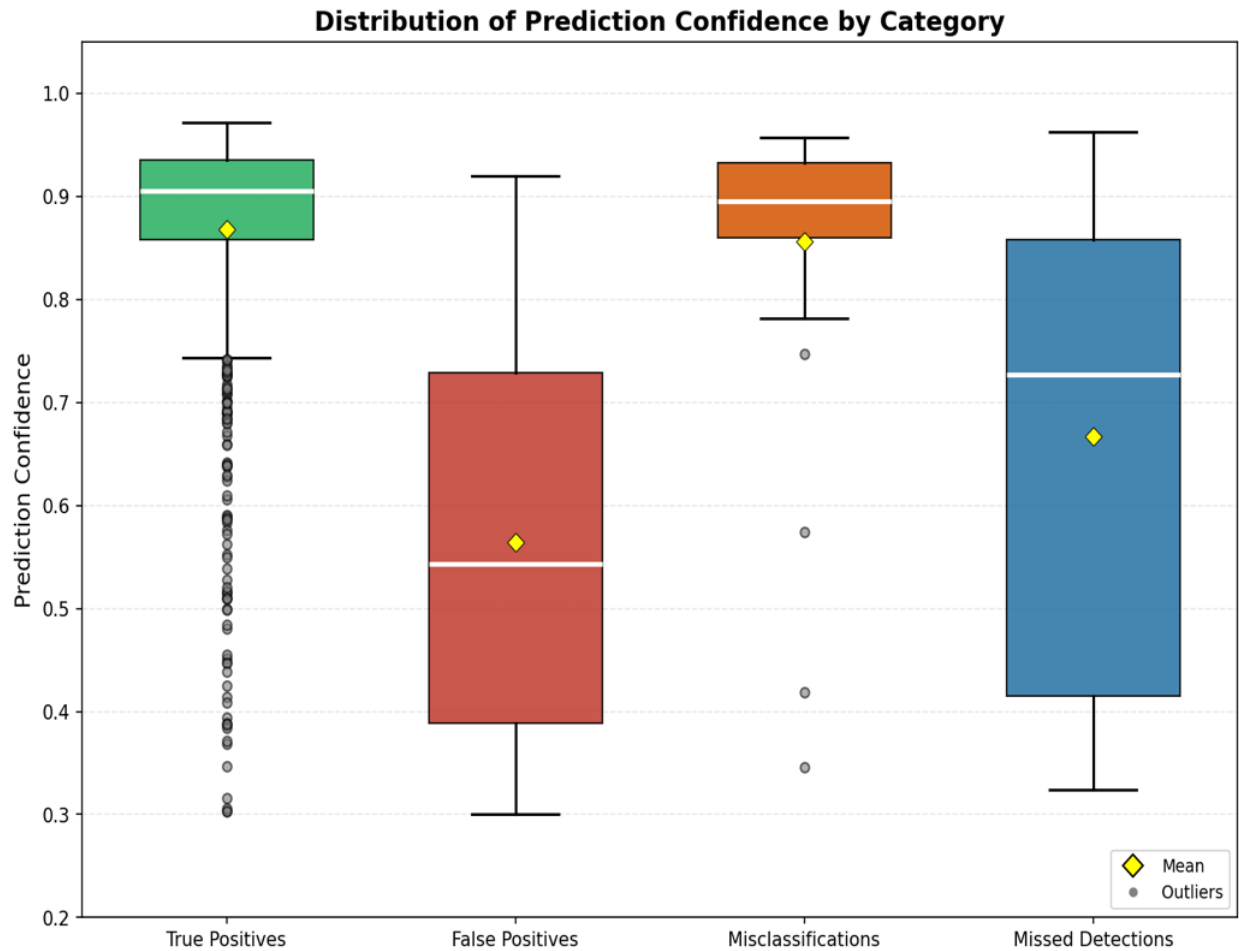


Figure 6. Distribution of prediction confidence across outcome categories (test set lizard instance count = 1011). Boxes span the IQR (Q1–Q3), white lines mark medians, yellow diamonds mark means, and grey dots denote outliers, defined as points falling more than $1.5 \times \text{IQR}$ below Q1 or above Q3. True positives ($n = 896$, 88.6% of test set, median 0.91, IQR 0.86–0.94) and misclassifications ($n = 39$, 3.86% of test set, median 0.90, IQR 0.86–0.93) are tightly clustered at high confidence, whereas false positives ($n = 81$, 8.01% of test set, median 0.54, IQR 0.39–0.73) occupy a much broader, lower range. Missed detection statistics (median 0.73, IQR 0.42–0.86) are computed over the 44 of 67 cases for which a low-loU prediction was produced; the remaining 23 cases are true missed detections with no prediction and thus have no confidence value.

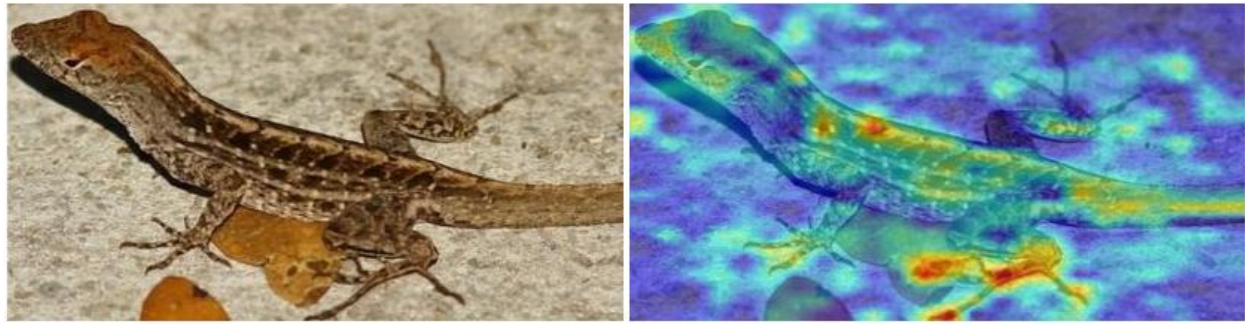


Figure 7. Grad-CAM visualization for a Brown Anole (*Anolis sagrei*). Original image (left) followed by the Grad-CAM heatmap from the Swin Transformer model (right). Warmer colors (red/yellow) indicate regions with higher contribution to the model’s prediction. The model primarily attends to the head, toepad, and limb regions rather than the dorsal patterning, indicating partial alignment with known diagnostic features.