Using Machine Learning to Identify Reptiles in Jurassic Park Surveillance Images

Jurassic Park Data Science Team (East 3)

Natasha Malik
Gabby Salera
Eric Shin
Steven Villalon

December 2025

University of Notre Dame Data Science M.S. - Advanced Machine Learning Final Project

Main Objective

Our priority is to monitor the location of the dinosaurs in the park at all times. We do this primarily through the use of motion-sensitive cameras.

However, as the ecosystem of Jurassic Park has grown, our cameras are being tripped by other large mammals that inhabit the park.

Our goal is develop a machine learning algorithm to distinguish reptiles from large mammals, focusing on **recall**.

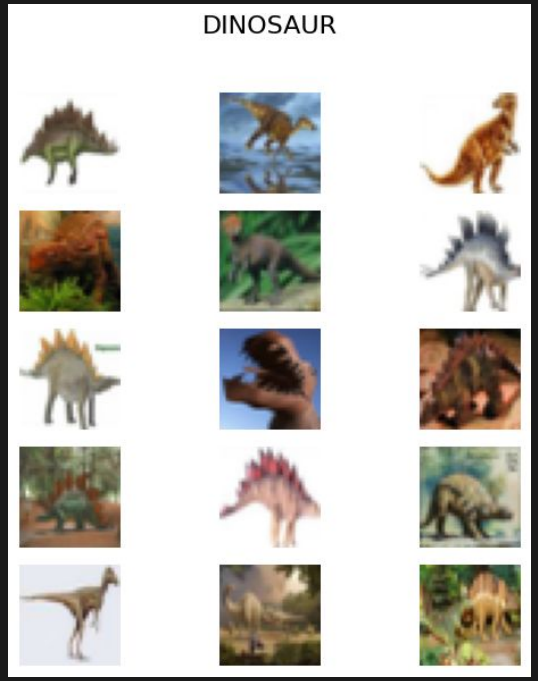
Training Data Class Balance: 33% reptiles, 67% mammals

Training was done using the CIFAR-100 dataset from the University of Toronto. The dataset has 100 classes containing 600 images each. Each image is a 32x32x3 tensor. This means there are 1024 (32^2) pixels per image, and there are 3 layers per image (red, green, blue), so there are 3072 numbers that make up each image.

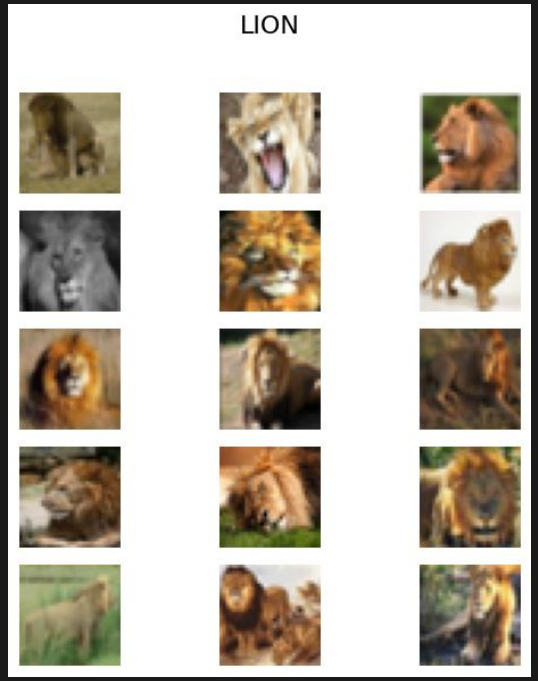
Superclasses used:

- Reptiles (positive class)
- Large carnivores
- Large omnivores/herbivores

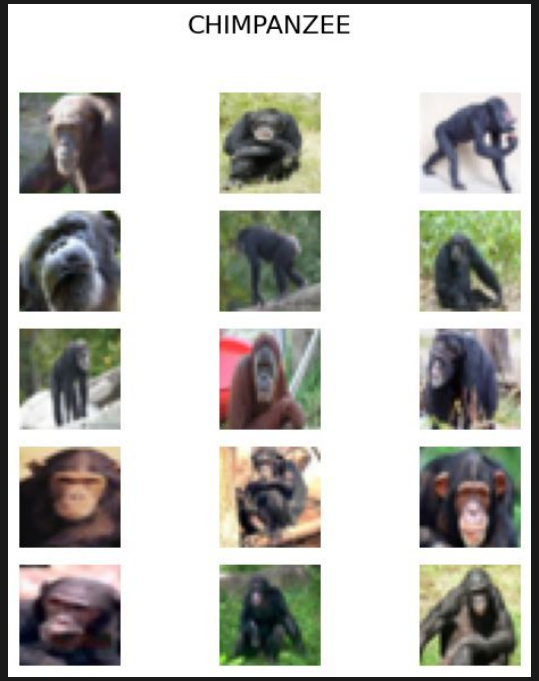
Heterogeneity within classes makes classification challenging



Mix of different dinosaurs and backgrounds.

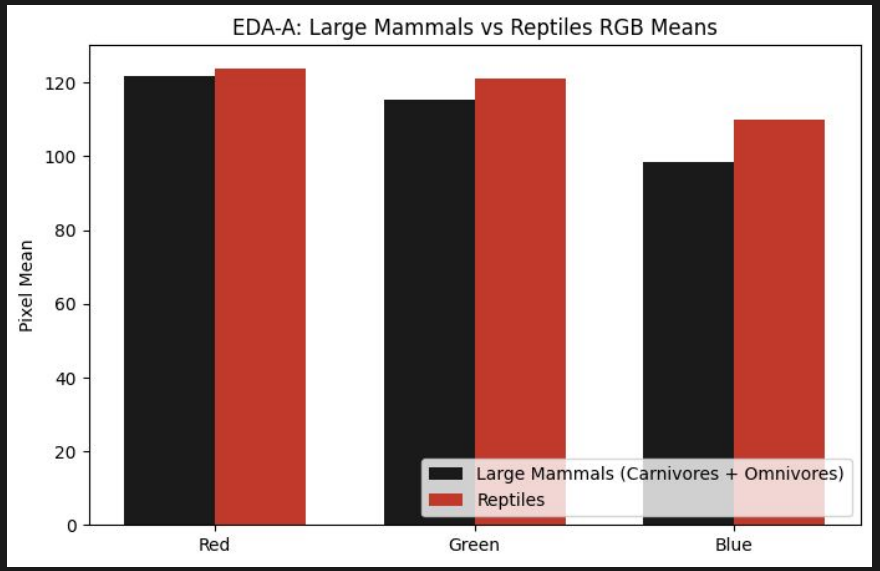


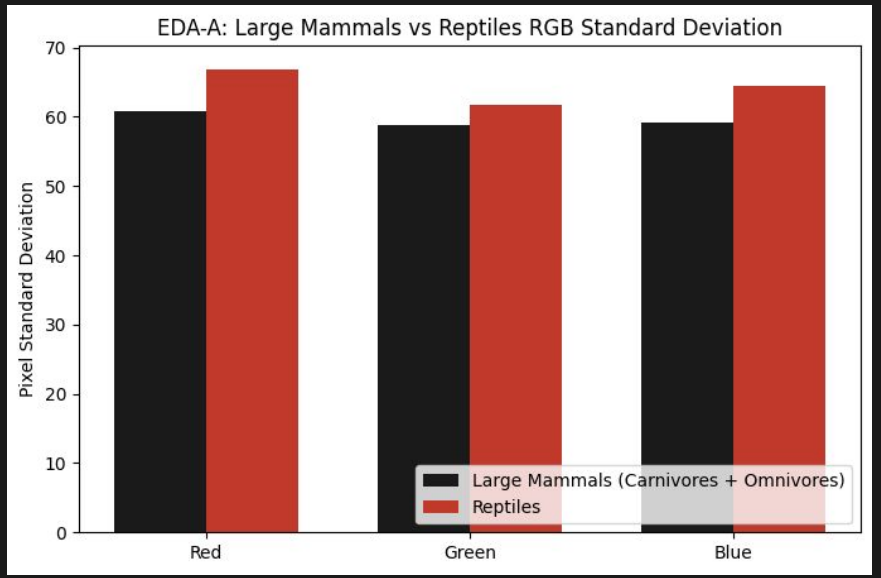
Mostly headshots, but many different angles.



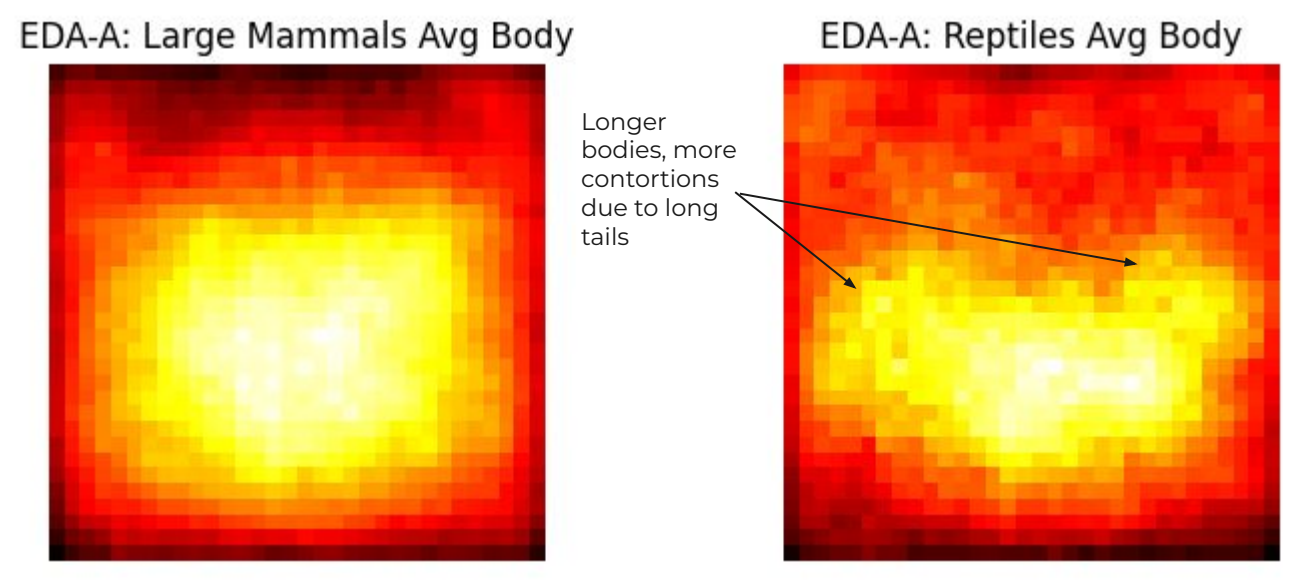
Similar colors and distinct head shapes, but varied backgrounds.

Reptiles have higher color saturation & variability





Reptiles exhibit unique body shape/positioning



Results after removing backgrounds & producing a heatmap of the foreground object using scimage package

Base model accuracy same as guessing mammal in all cases

Logistic Regression Model

Test/validation - 16% split

Reshaped data to 2 dimensions

Standard scaled to 0/1's

Preprocessed with PCA

Cross Validation on Cs

L1 regularization

Liblinear solver

Confusion Matrix

		Predicted	
		Mammals	**Reptiles**
Actual	**Mammals**	758	242
	Reptiles	261	239

	Precision	**Recall**	**F1 Score**	**n**
Mammals	0.7439	0.7580	0.7509	1000
Reptiles	0.4969	0.4780	0.4873	500
Accuracy		0.6647		

Adding more layers had biggest impact on CNN accuracy

Convolutional Neural Network Model

4 convolutional layers (32 to 256 filters)

4 pooling layers (all 2x2)

Early stopping

Batch size = 32

Class weights:

Reptiles = 2

Mammals = 1

Probability threshold for Reptile = 0.4

Confusion Matrix

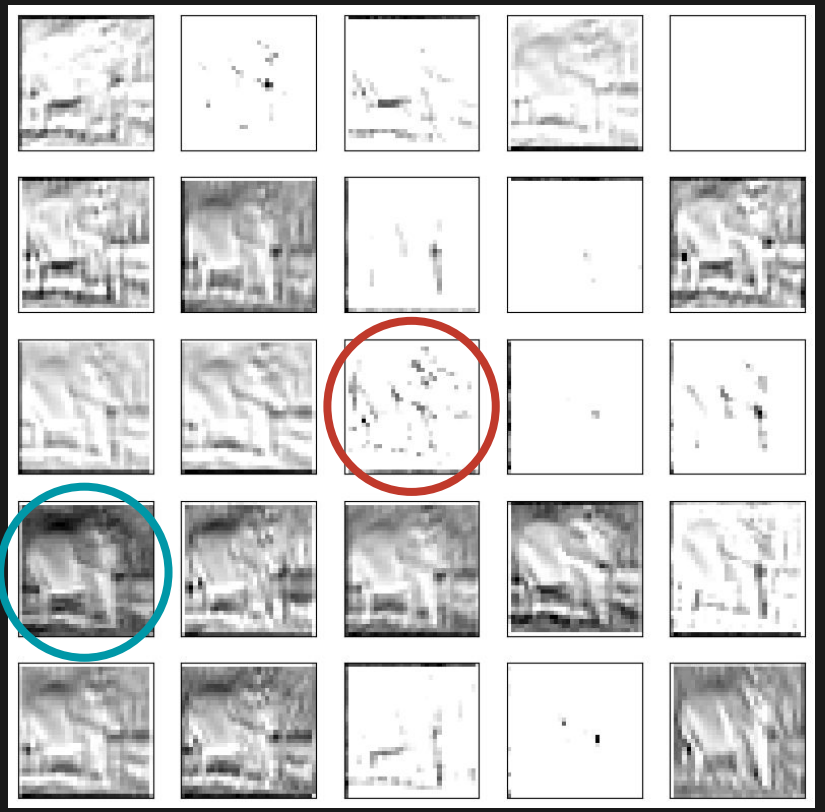
		Predicted	
		Mammals	**Reptiles**
Actual	**Mammals**	763	237
	Reptiles	105	395

	Precision	**Recall**	**F1 Score**	**n**
Mammals	0.8790	0.7630	0.8169	1000
Reptiles	0.6250	0.7900	0.6979	500
Accuracy		0.7720		

Feature maps: first convolutional layer captures big picture

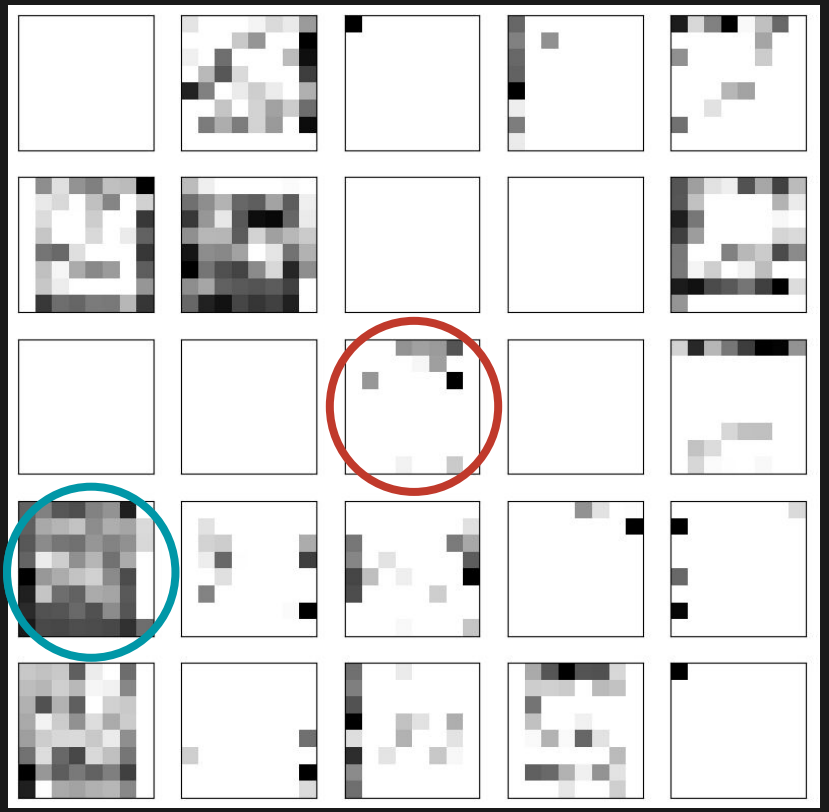
Stegosaurus

Elephant

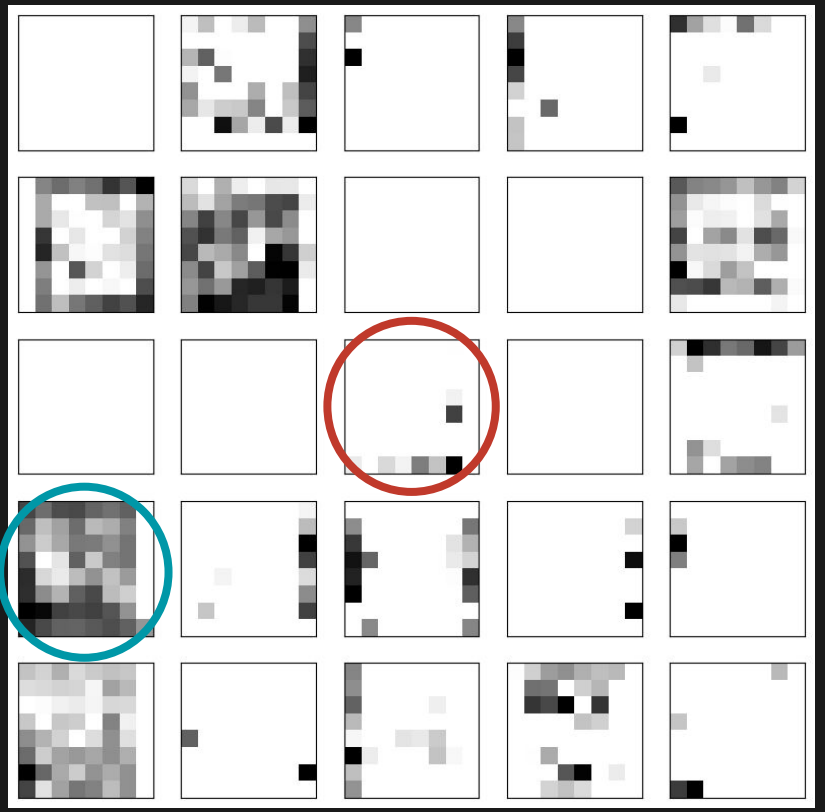


By the third convolutional layer, the feature maps are indecipherable

Stegosaurus



Elephant



Final Results on Test Set

Logistic

Accuracy
68%

Reptile Recall
54%

CNN Initial

Accuracy
71%

Reptile Recall
50%

CNN Final

Accuracy
76%

Reptile Recall
81%

Addressing a few limitations could further improve the model

Higher resolution images could help identify more features

Collect larger dataset (only trained on 6,000 images)

Set up training data to make classes more homogenous

Try running models while removing the backgrounds

Try additional activation functions or stabilization layers

Try other models (e.g., Random Forest)

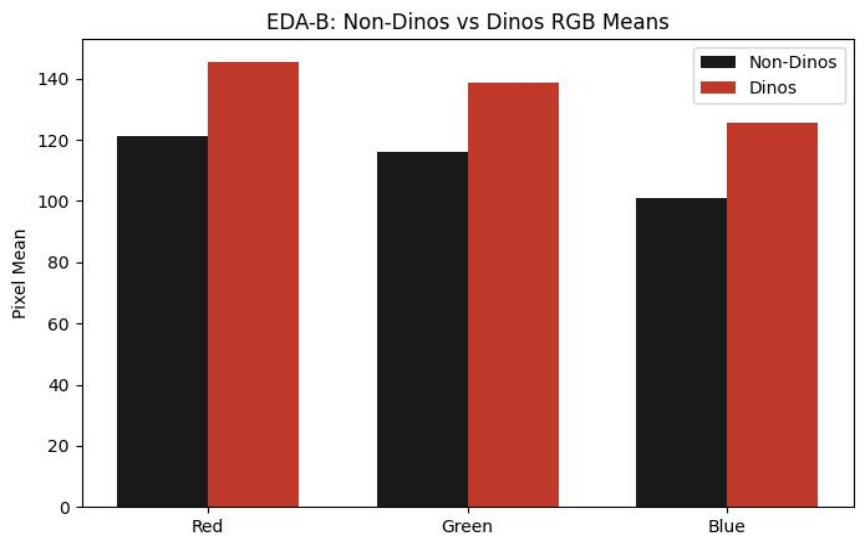
Use a pre-trained model

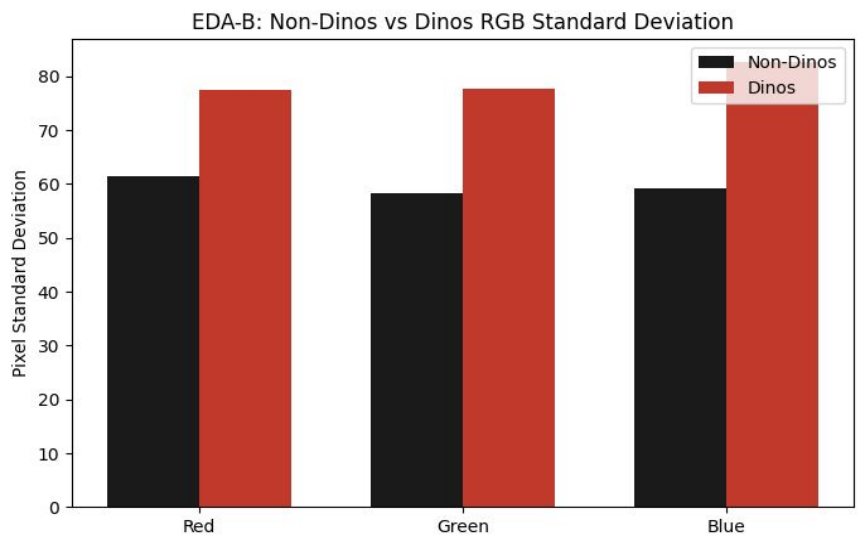
A model 65 million iterations in the making...

Thank you!

Appendix

Dinosaurs alone are even more saturated & color diverse





Dinosaurs are more centrally located in the images