

CREATING AND USING SYNTHETIC DATA SETS TO TRAIN MACHINE LEARNING MODELS:

[AN UNDERSTANDABLE MOUTHFUL]

BY: SETH FRAZER





THE WHY Camera Traps -



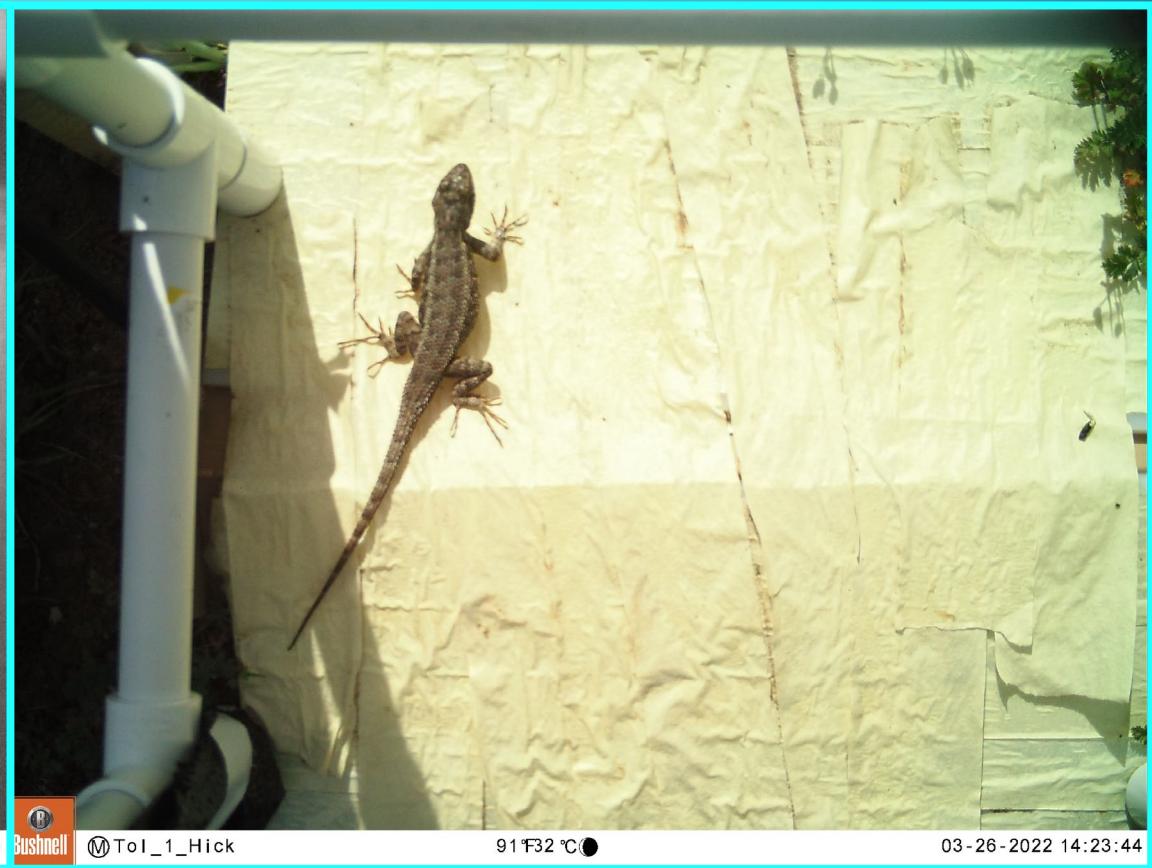
THE WHY

Camera Traps -



THE WHY

Camera Traps -



THE JARGON:

- **Machine Learning (ML) Models** =
“the *thing* that is saved after running a machine learning algorithm on training data and represents the rules, numbers, and any other algorithm-specific data structures required to make predictions on input data. – Jason Brownlee -
<https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/>

Synthetic training data =
“information that's artificially manufactured rather than generated by real-world events and is created algorithmically...and can be used to train machine learning models.” – Nicole Laskowski -
<https://searchcio.techtarget.com/definition/synthetic-data>

THE JARGON:

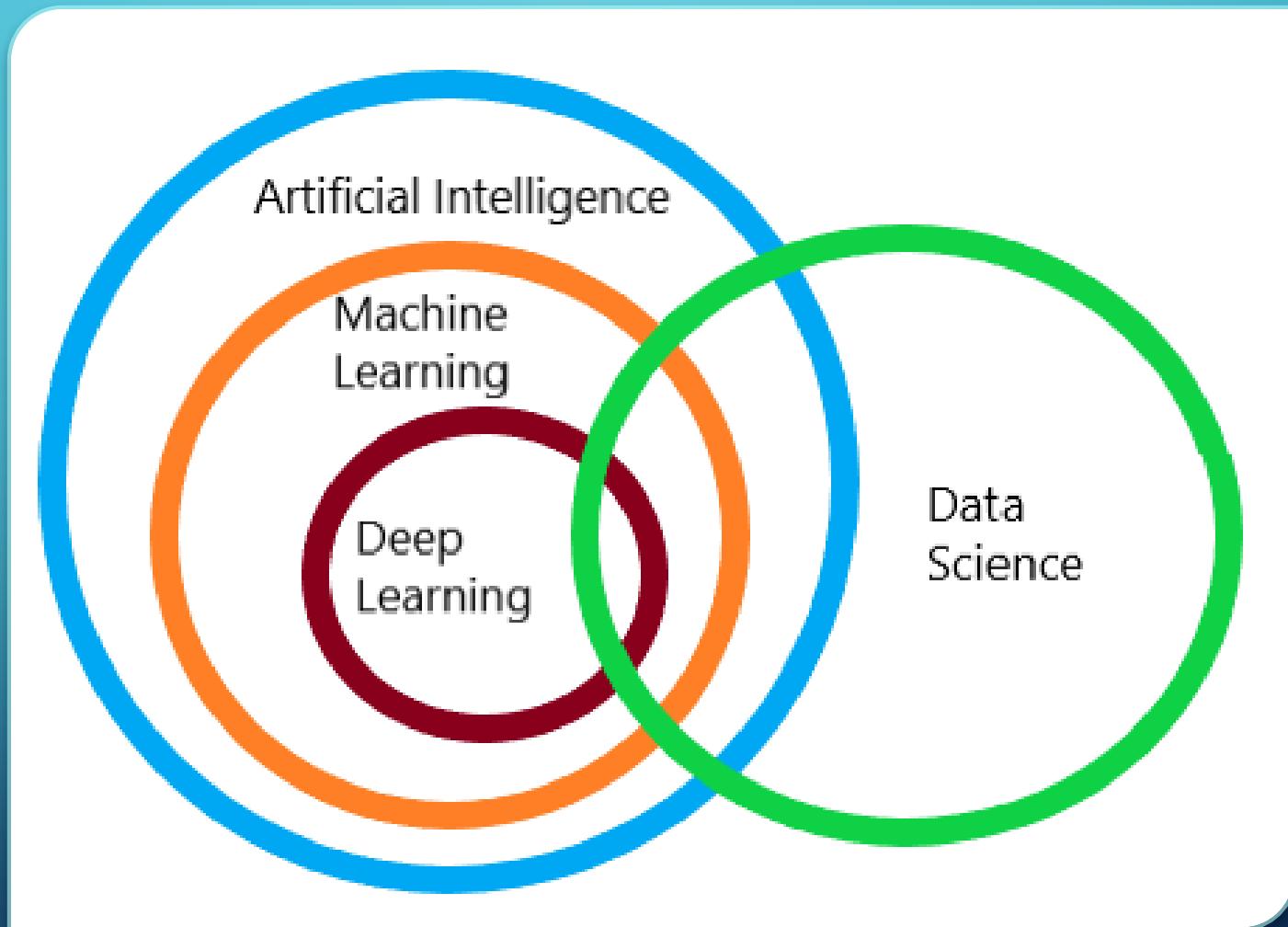


IMAGE CLASSIFIERS - A SUB-CLASS OF ML MODELS



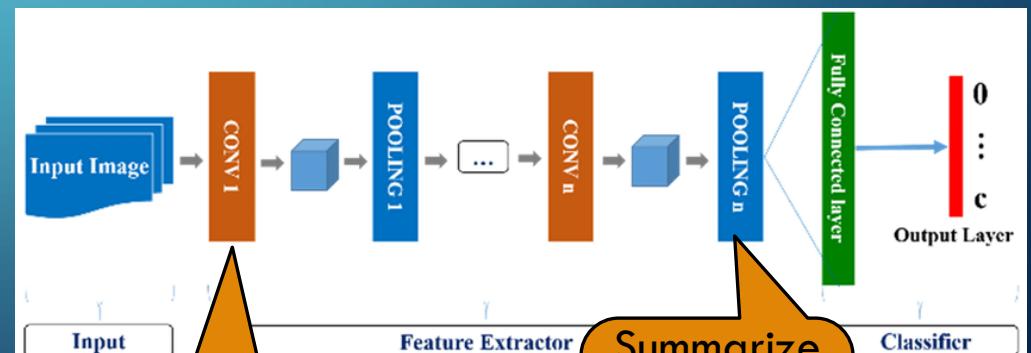
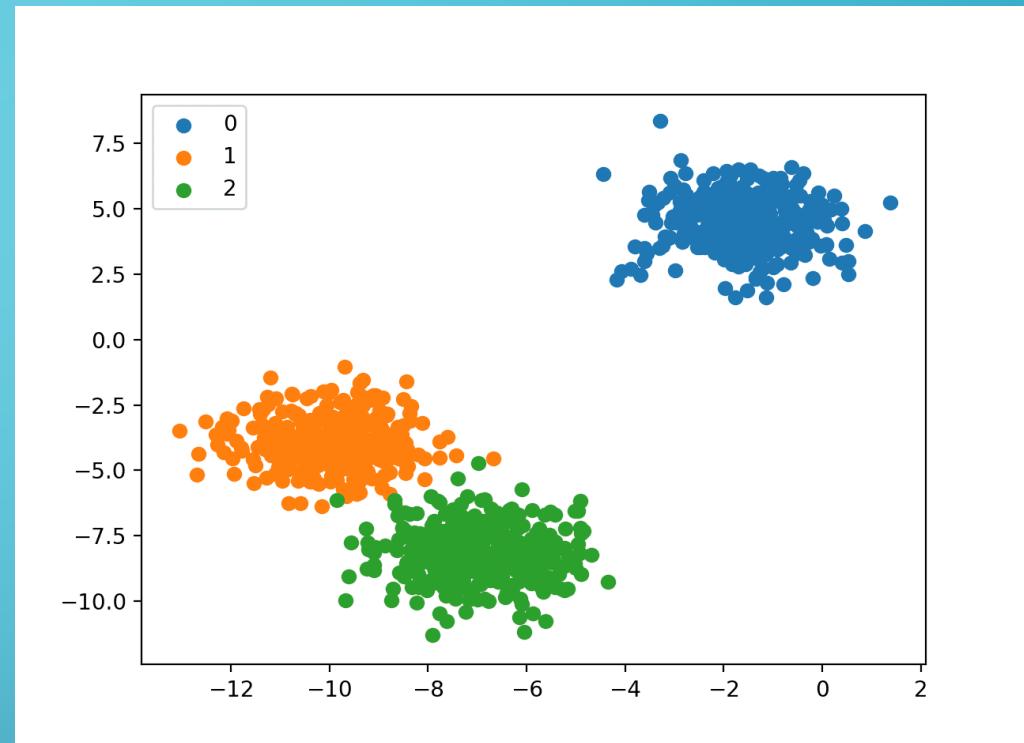
Cat - $n = 300$



Dog - $n = 300$



Cthulhu - $n = 300$



Finds
patterns
in
data

Summarizes
patterns
into
“Features”

IMAGE CLASSIFIERS - A SUB-CLASS OF ML MODELS



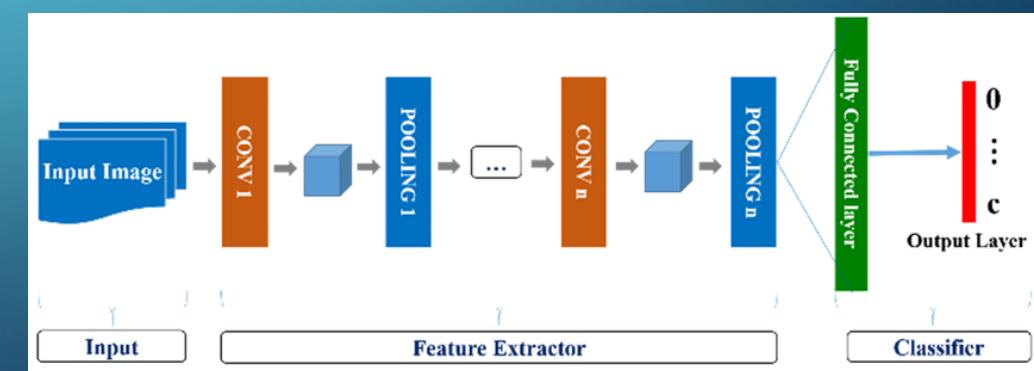
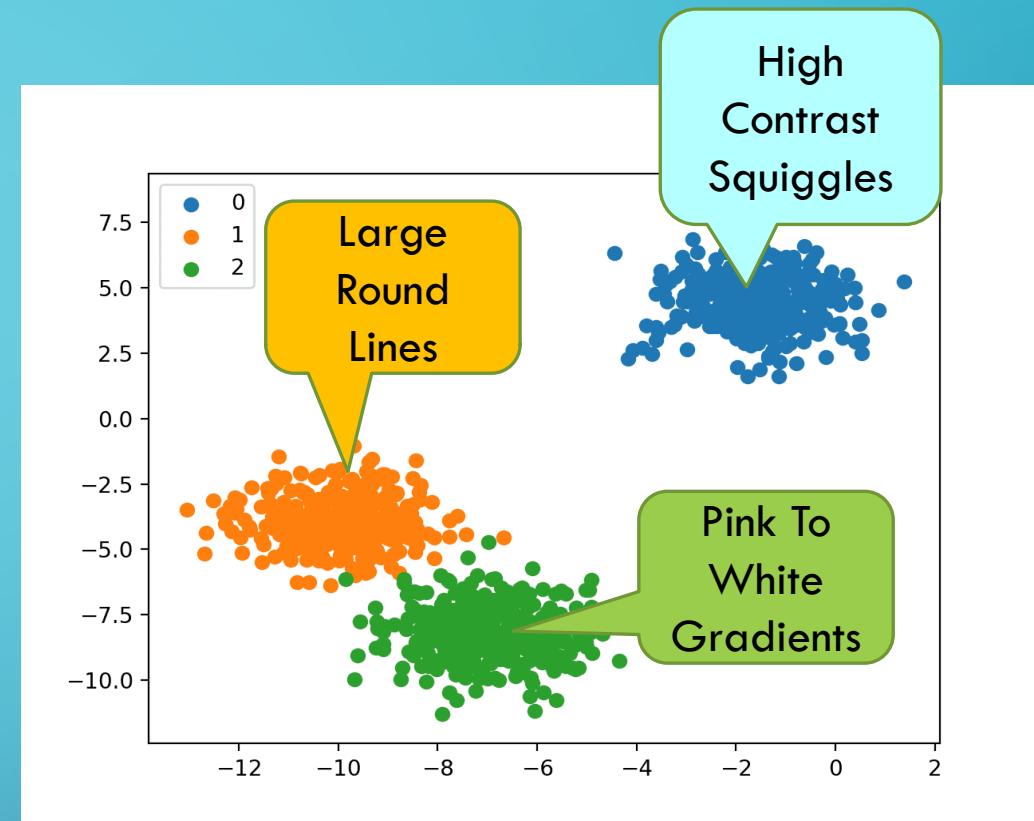
Cat - n = 300



Dog - n = 300



Cthulhu - n = 300



THE WHAT:

We trained a model to recognize
difference species of “Herpetofauna”
[Reptiles and Amphibians], using the
CCBER Museum’s pickled herp. specimen.

THE HOW:

1. Herbarium
Light Box

2. Cutout in
PS

3. Animation

4. Making
BG Types +
Matching

5. Shadows

6. ML Model



E. e. croceater



E. multicarinata



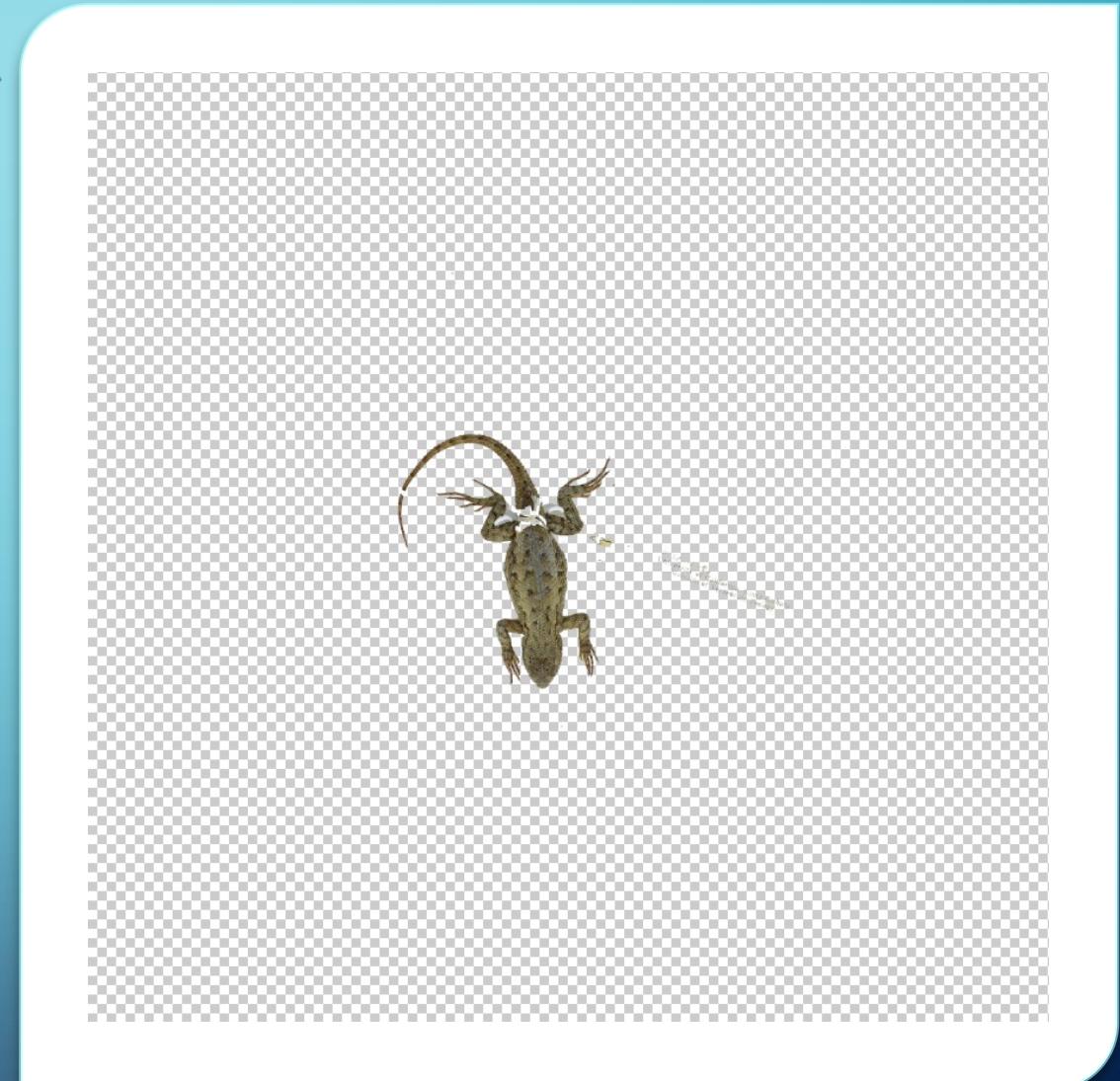
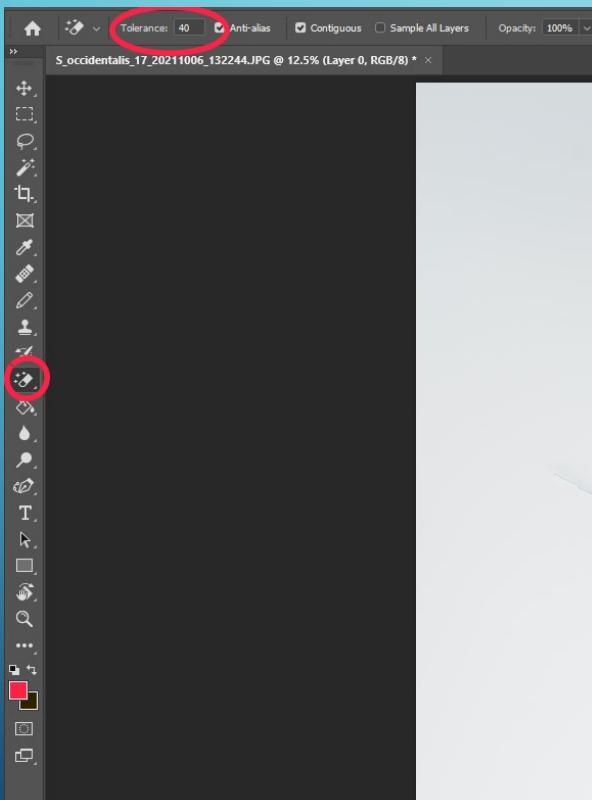
S. occidentalis

THE HOW:

STEP 1 - HERBARIUM LIGHT BOX

THE HOW:

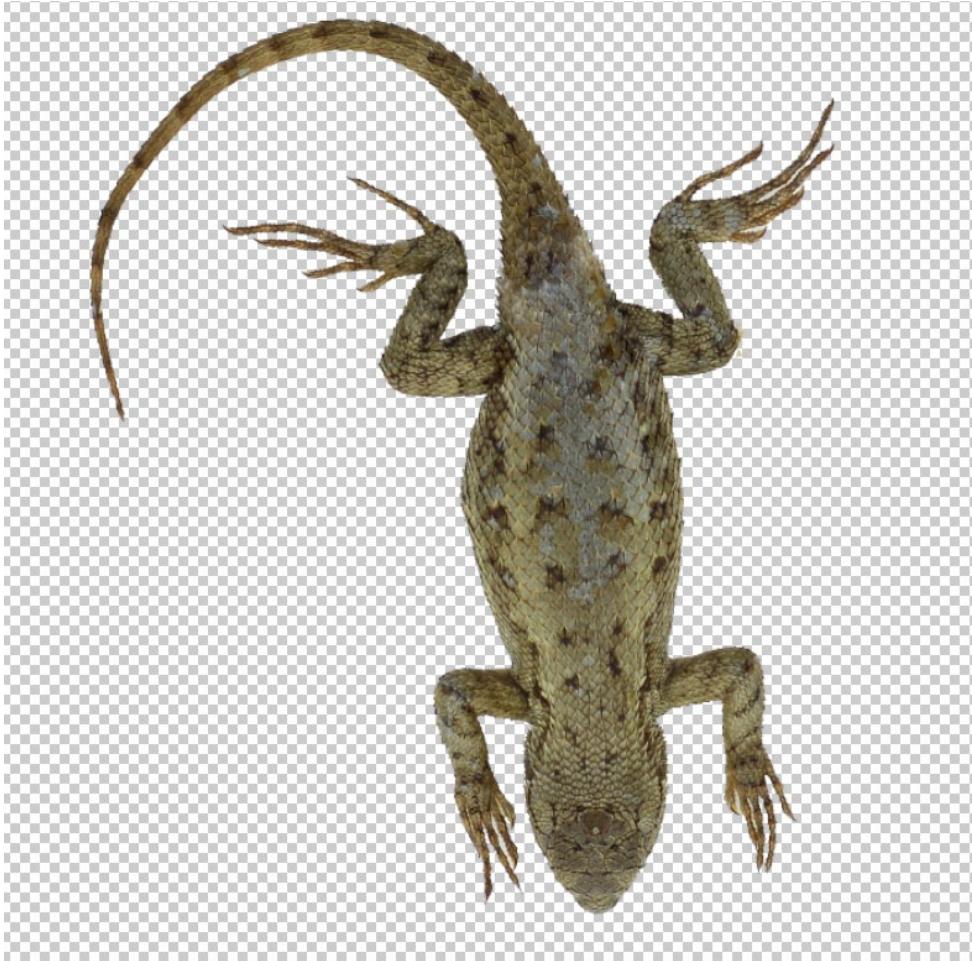
STEP 2 – USING PHOTOSHOP TO
“CUTOUT” SPECIMEN

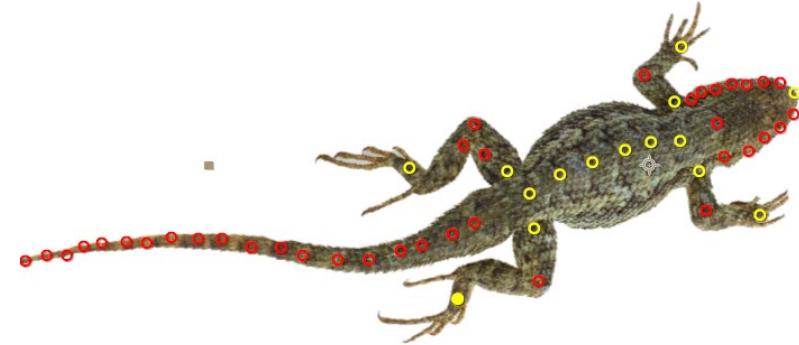


THE HOW:



STEP 2 – USING PHOTOSHOP TO
“CUTOUT” SPECIMEN



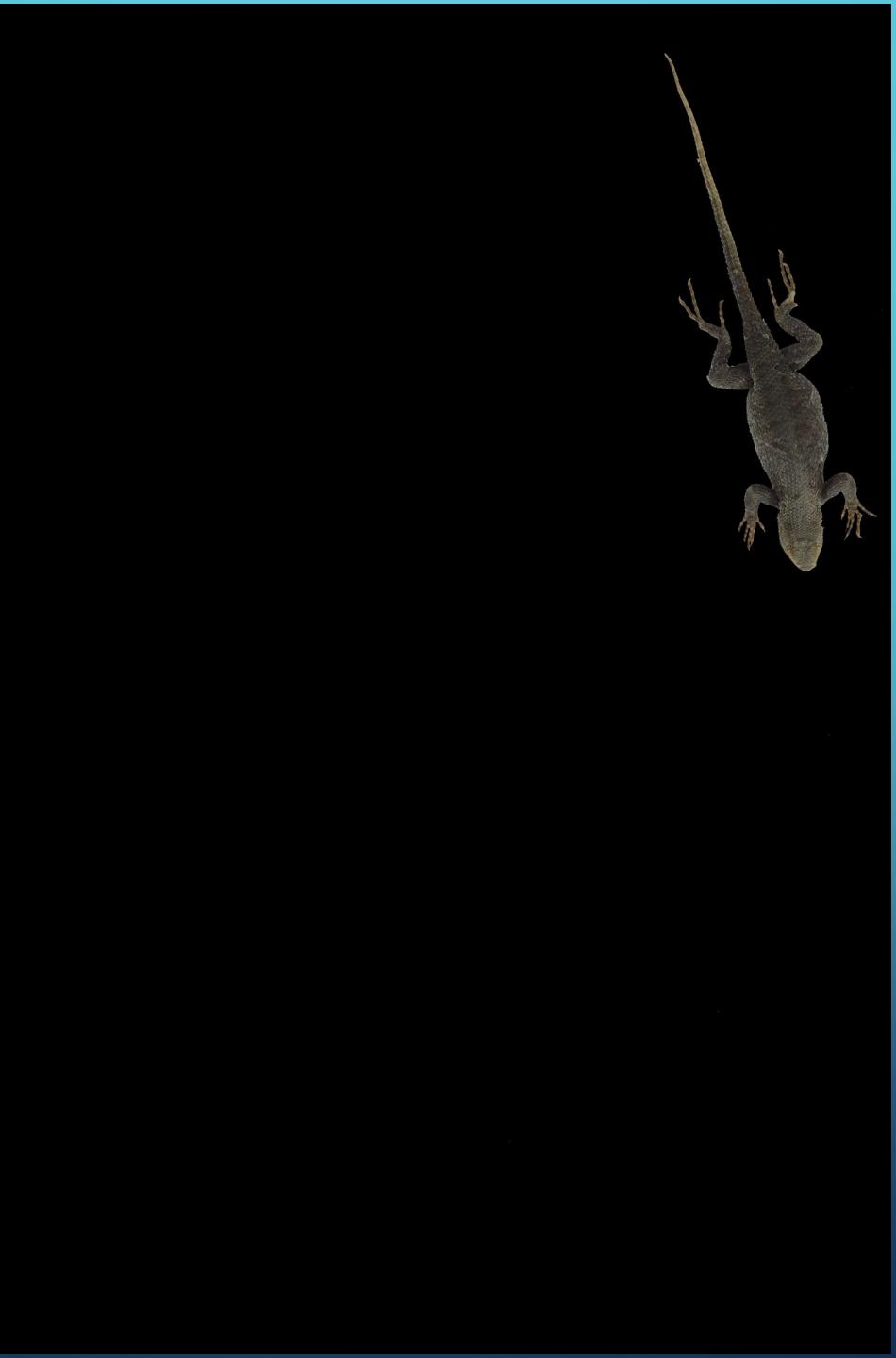


THE HOW:

STEP 3 – USING AFTER EFFECTS TO “ANIMATE” SPECIMEN



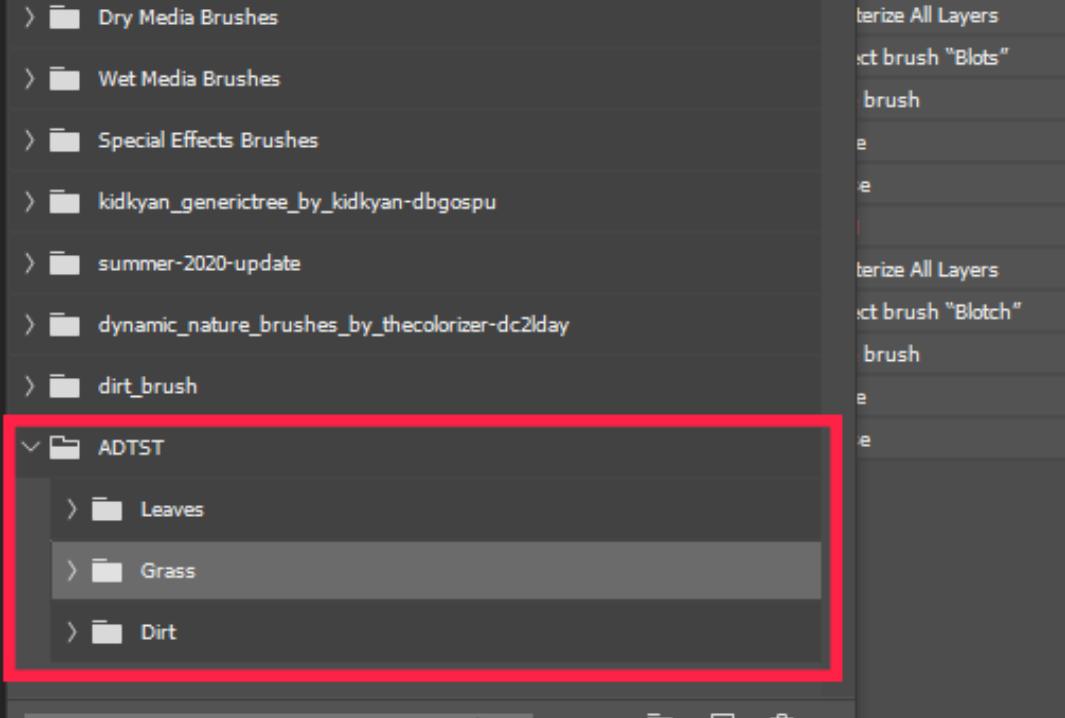
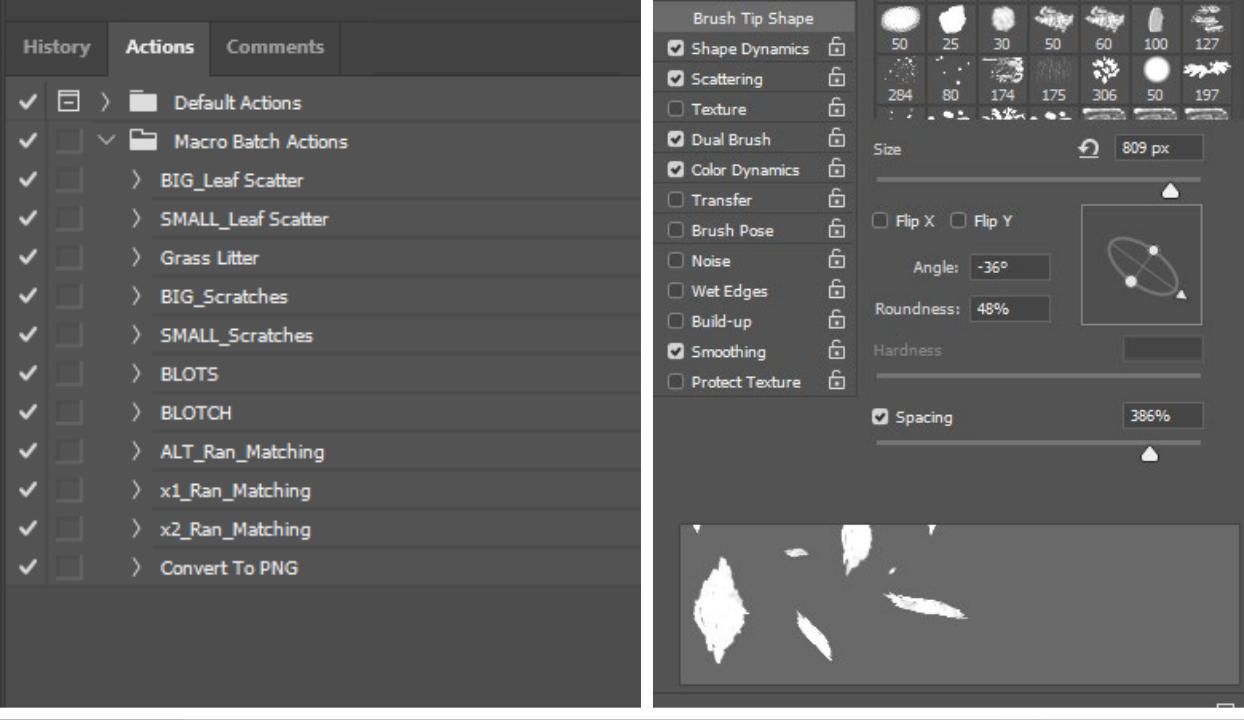






THE HOW:

STEP 4 – USING PHOTOSHOP TO MAKE BACKGROUNDS



STEP 4 – USING PS TO MAKE “BACKGROUNDS”



THE HOW:



STEP 4 – BACKGROUND TYPES -
GRASS



THE HOW: 

STEP 4 – BACKGROUND TYPES - LEAF



THE HOW:

STEP 4 – BACKGROUND TYPES - DIRT



THE HOW: 

STEP 4 – BACKGROUND TYPES - DIRT

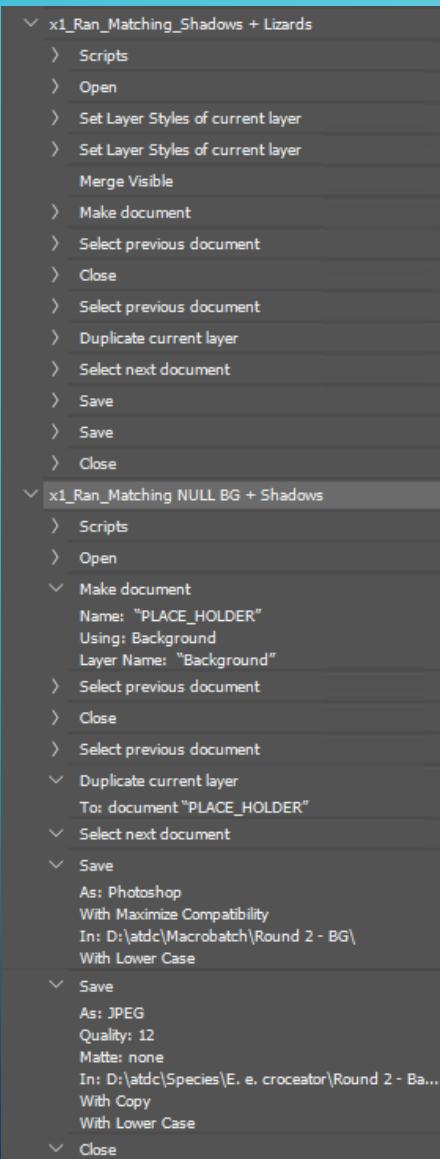


THE HOW:



STEP 4 – BACKGROUND TYPES –
DIRT ROAD

THE HOW: Step 5 – Using PS to Match Backgrounds & Specimen



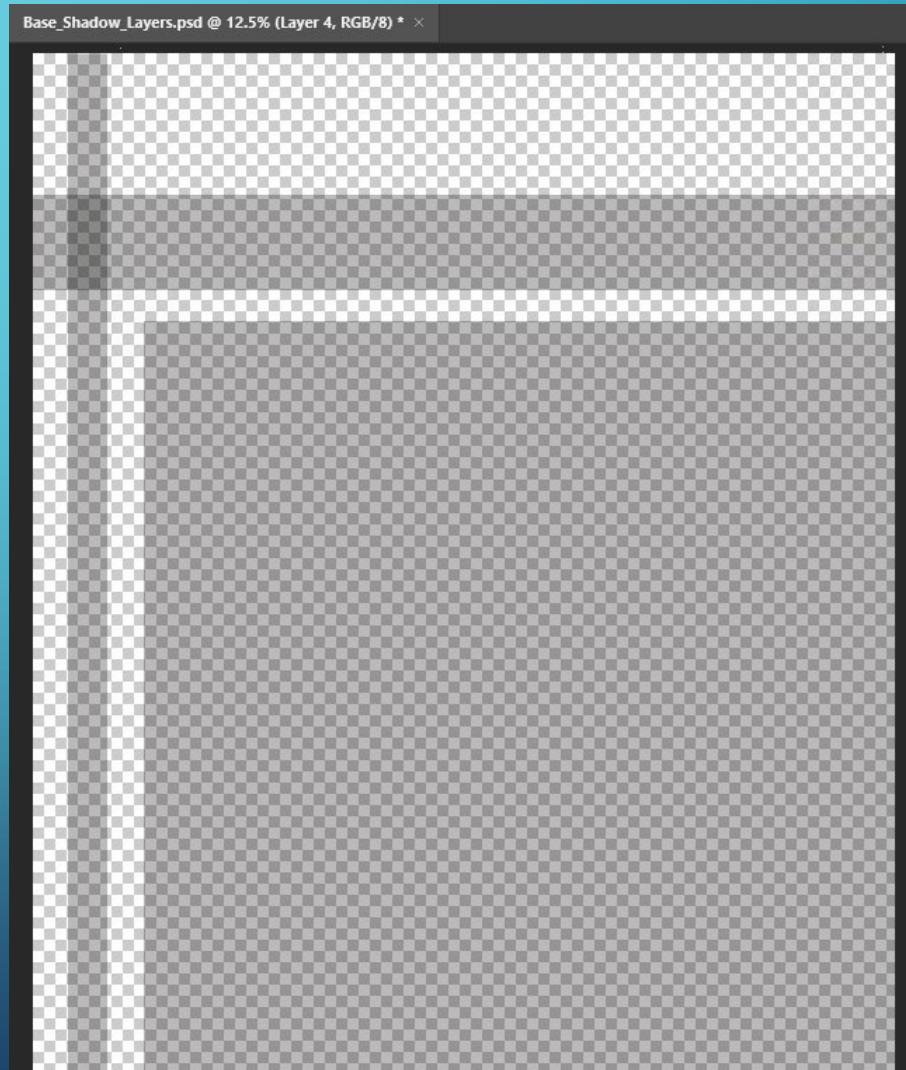
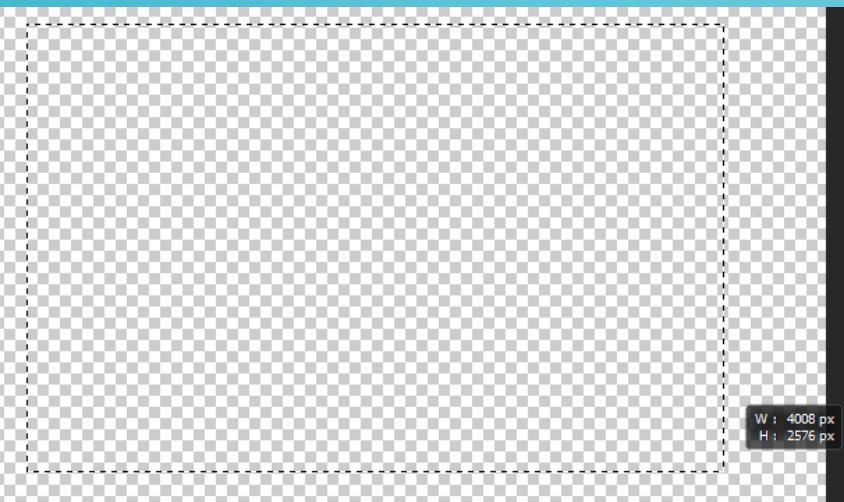
THE HOW: Step 5 – Using PS to Match Backgrounds & Specimen

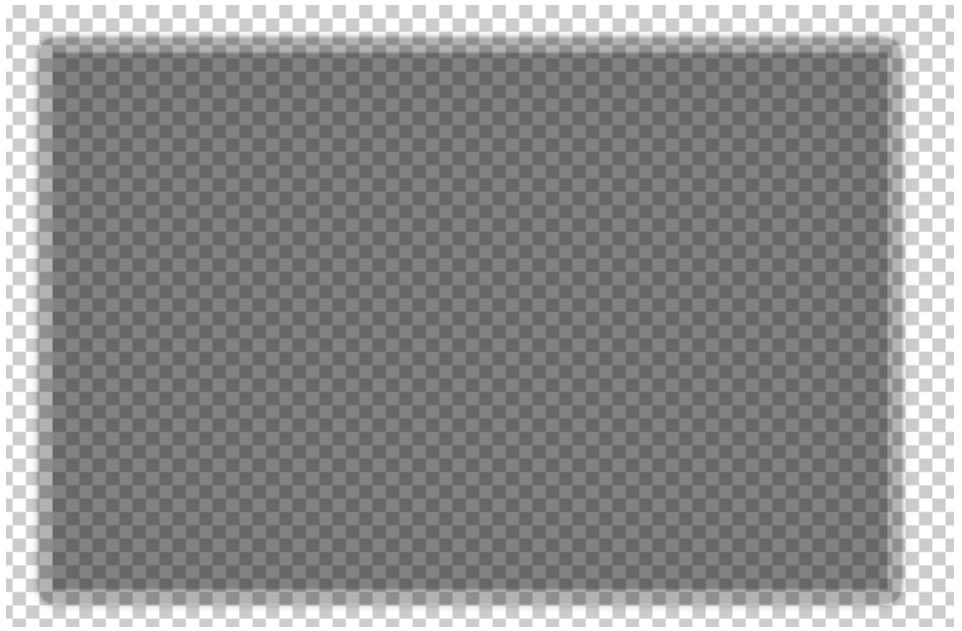


THE HOW: Step 6 – Using PS & AE to Make, Randomize, and Match Shadows To ATD



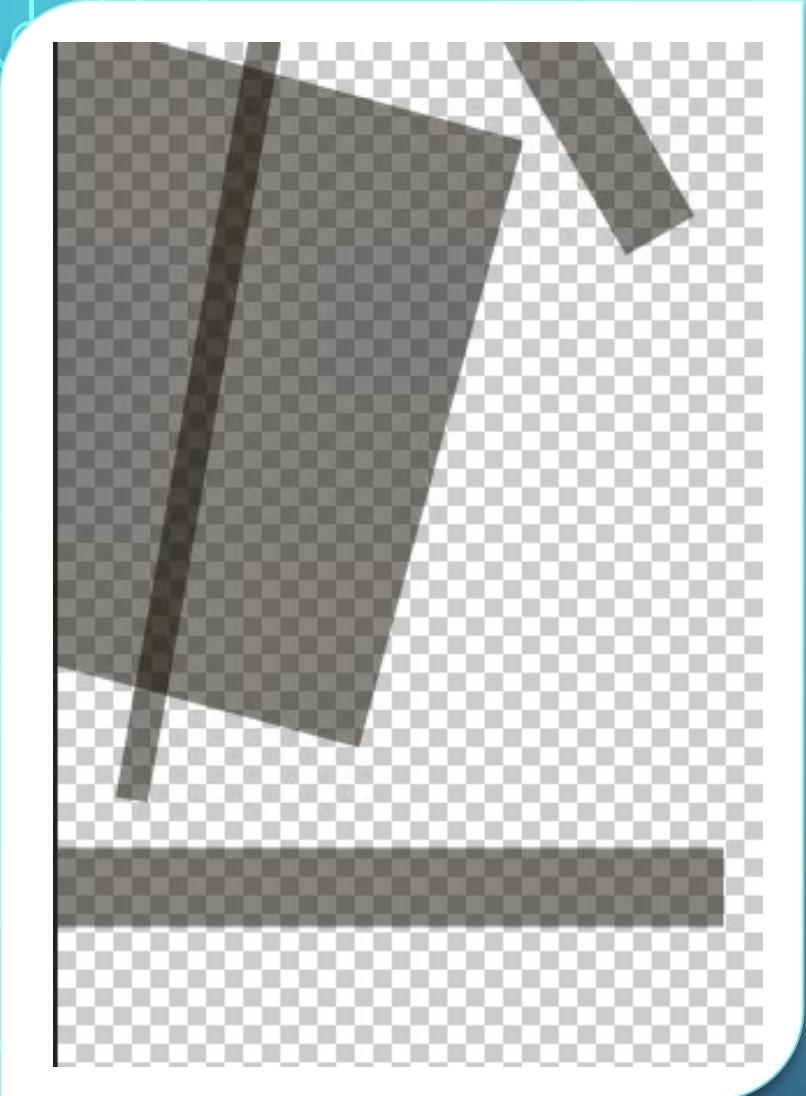
THE HOW: Step 6 – Using PS & AE to Make, Randomize, and Match Shadows To ATD





THE HOW:

STEP 6 – USING PS & AE TO MAKE, RANDOMIZE, AND MATCH SHADOWS TO ATD



THE HOW:



STEP 6 – USING PS & AE TO MAKE, RANDOMIZE, AND
MATCH SHADOWS TO ATD

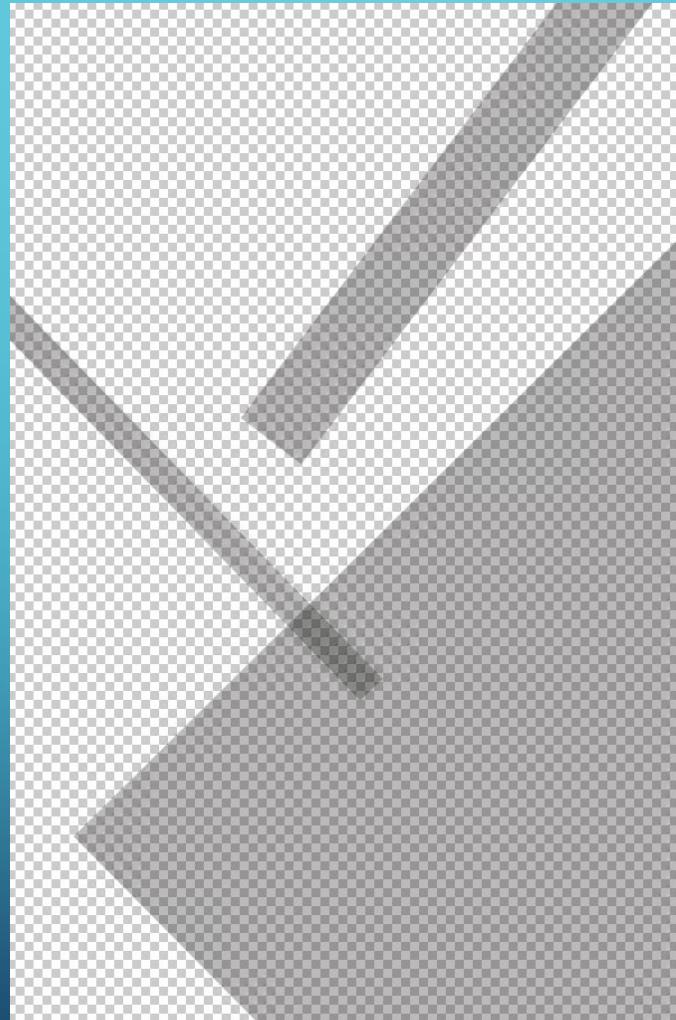
THE HOW:

STEP 6 – USING PS & AE TO MAKE, RANDOMIZE, AND MATCH SHADOWS TO ATD\

```
<ul style="list-style-type: none; padding-left: 0;">- ↳ Scripts
- ↳ Open
- ↳ Set Layer Styles of current layer
- ↳ Set Layer Styles of current layer
- ↳ Merge Visible
- ↳ Make document
- ↳ Select previous document
- ↳ Close
- ↳ Select previous document
- ↳ Duplicate current layer
- ↳ Select next document
- ↳ Save
- ↳ Save
- ↳ Close
  


- ↳ Scripts
- ↳ Open
- ↳ Make document
- Name: "PLACEHOLDER"
- Using: Background
- Layer Name: "Background"
- ↳ Select previous document
- ↳ Close
- ↳ Select previous document
- ↳ Duplicate current layer
- To: document "PLACEHOLDER"
- ↳ Select next document
- ↳ Save
- As: Photoshop
- With Maximize Compatibility
- In: D:\atdc\Macrobatch\Round 2 - BG\
- With Lower Case
- ↳ Save
- As: JPEG
- Quality: 12
- Matte: none
- In: D:\atdc\Species\E. e. croceator\Round 2 - Ba...
- With Copy
- With Lower Case
- ↳ Close

```



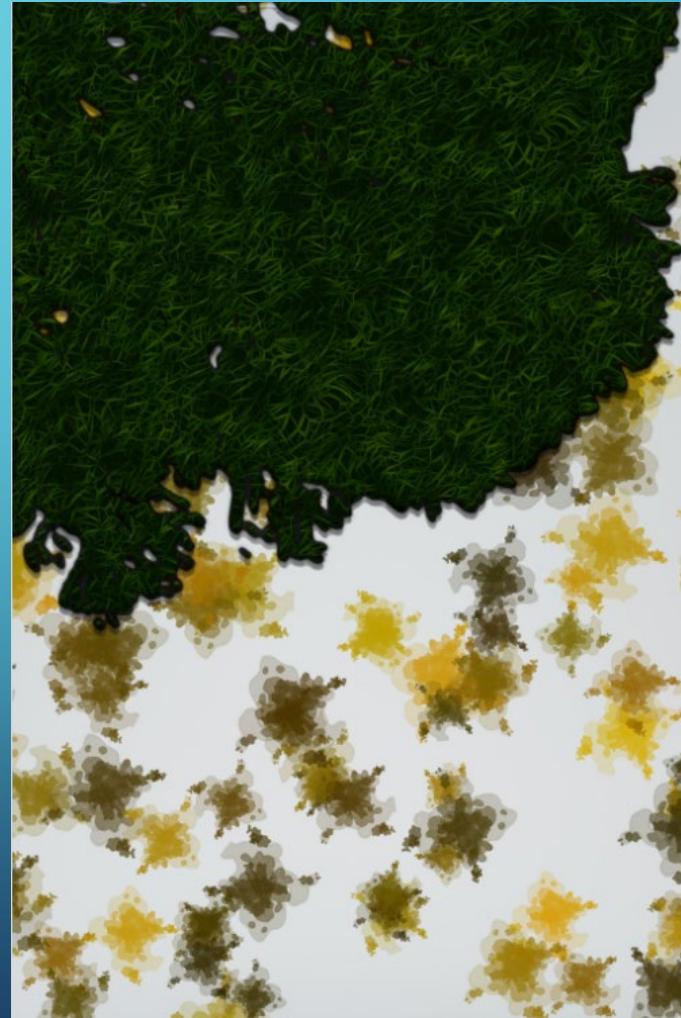
THE HOW:

STEP 6 – USING PS & AE TO MAKE, RANDOMIZE,
AND MATCH SHADOWS TO ATD

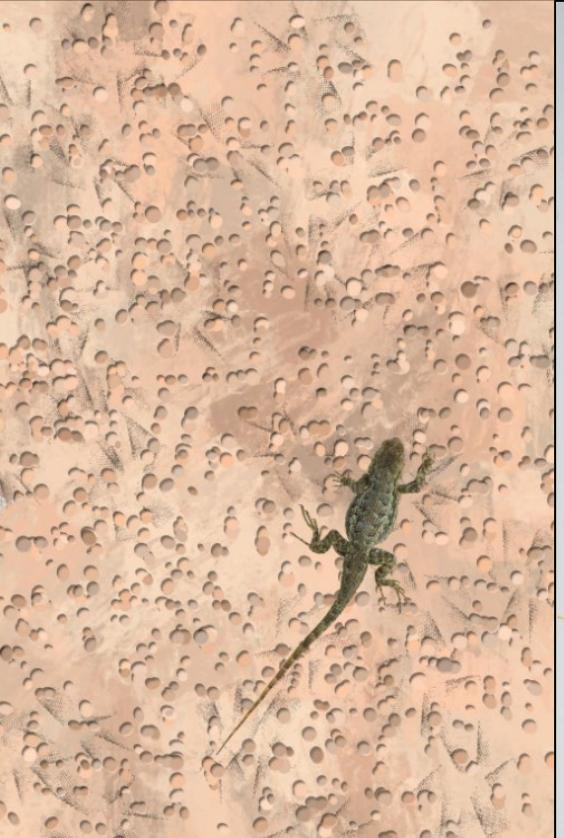


THE HOW: Ps

STEP 7 [OPTIONAL] – USING PS & AE TO MAKE, RANDOMIZE, AND MATCH “MOSS” TO ATD







Training Our Image Classifier Model(s)

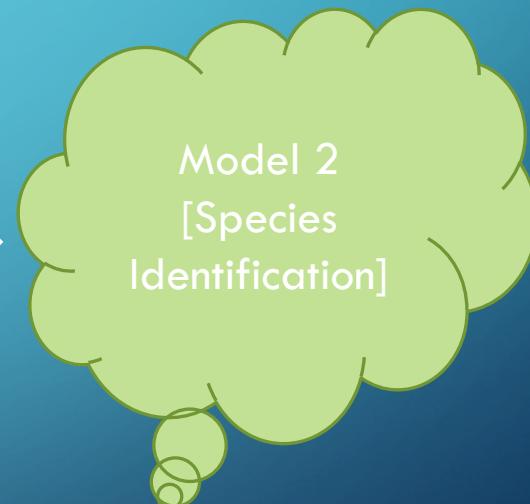
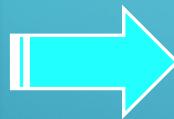
Using Google Cloud Platform's "Vertex AI" Application Program Interface (API)

Google Cloud Platform

ATD-3sp-Model

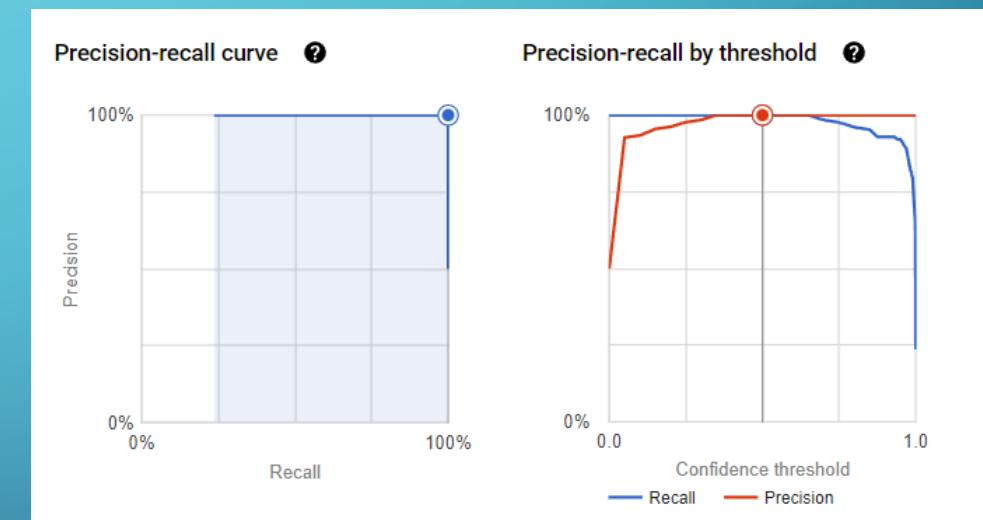
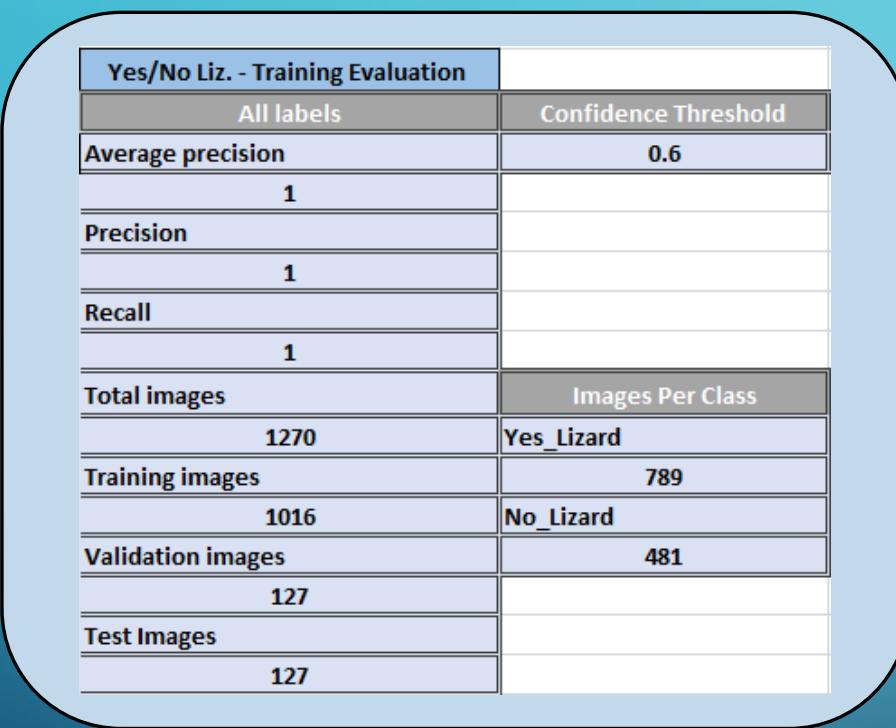


Vertex AI



Training Our Image Classifier Model(s)

Using Google Cloud Platform's "Vertex AI" Application Program Interface (API)



Training Our Image Classifier Model(s)

Using Google Cloud Platform's "Vertex AI" Application Program Interface (API)

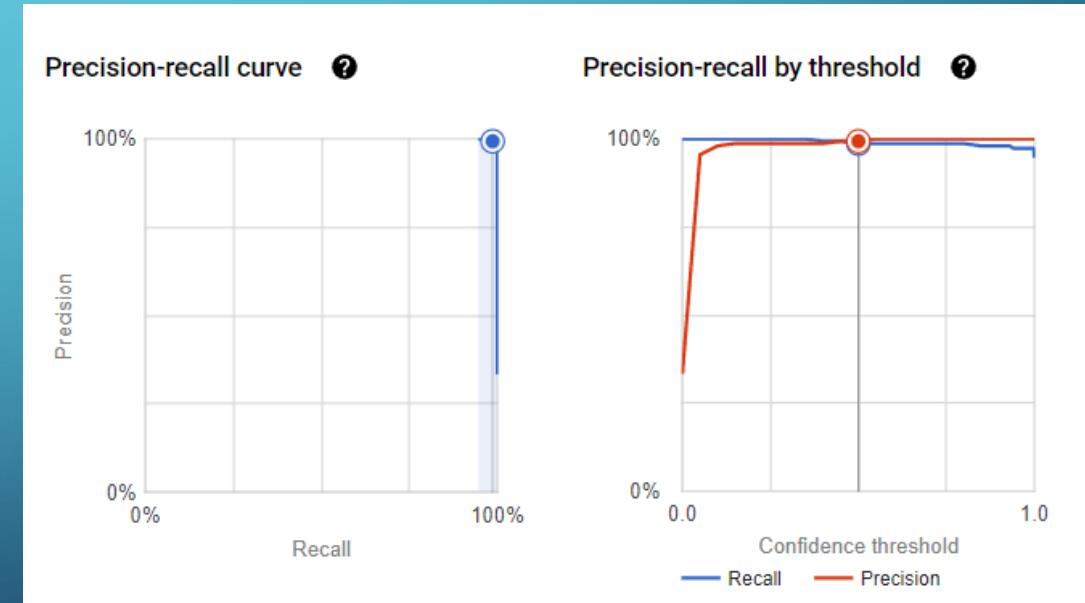
Google Cloud Platform

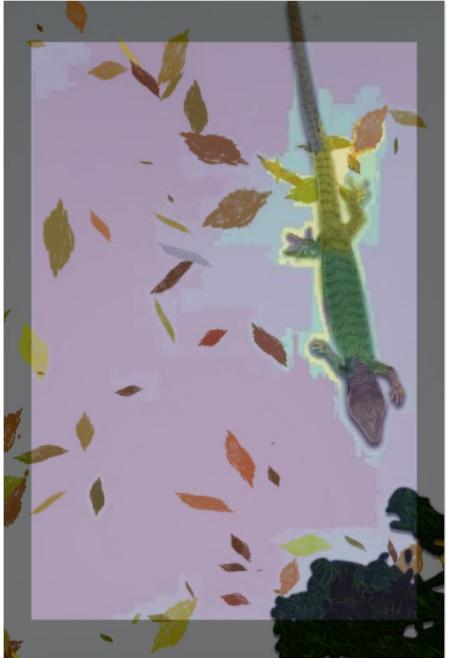
ATD-3sp-Model

Vertex AI

OG 4 - Training Evaluation:	
All labels	Confidence Threshold
Average precision	0.6
1	
Precision	
0.993	
Recall	
0.987	
Total images	Images Per Species
1489	S. Occidentalis
Training images	530
1191	E. Multicarinata
Validation images	489
149	E. e. croceater
Test images	470
149	

OG 4 - Training Confusion Matrix:		PREDICTED LABEL	
TRUE LABEL	S. Occidentalis	E. e. croceater	E. Multicarinata
S. Occidentalis	98%	2%	-
E. e. croceater	-	100%	-
E. Multicarinata	-	-	100%





THE RESULTS [SO FAR...]

XRAI – EXPLAINABILITY BITMAP OVERLAY

The Results [So Far...]

Using Google Cloud Platform's "Vertex AI" Application Program Interface (API)



Base Image

```
{"content": "gs://atd-test-images/Test/Y - Cro/eecroceaterbm05.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceaterbm052.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceaterbm054.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceaterbm05juv.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceaterft01juvdime.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceaterft3032.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceaterkc311.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceaterkc3112.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceatorcfc2.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceatorcfc2c.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/eecroceatorftj.jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/large (1).jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/large (2).jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/large (3).jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/large (7).jpg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/large.jpeg", "mimeType": "image/jpeg"},  
 {"content": "gs://atd-test-images/Test/Y - Cro/large.jpg", "mimeType": "image/jpeg"}
```

Associated JSONL Tag

Yes/No Lizard RL Test Evaluation -	
Yes -	100.0
No -	94.1
Total Correct	144
Total Tested	147
% Correct	98.0
OG5 W/Sticks + Moss V2	
occi	98.0
multi	92.0
croc	96.0
empty backgrounds	66.7
total correct	96
Total Tested	100
% Correct	96.0

BATCH PREDICTIONS ON "REAL LIFE" IMAGES

And Now, A Quick Demonstration

S. occidentalis



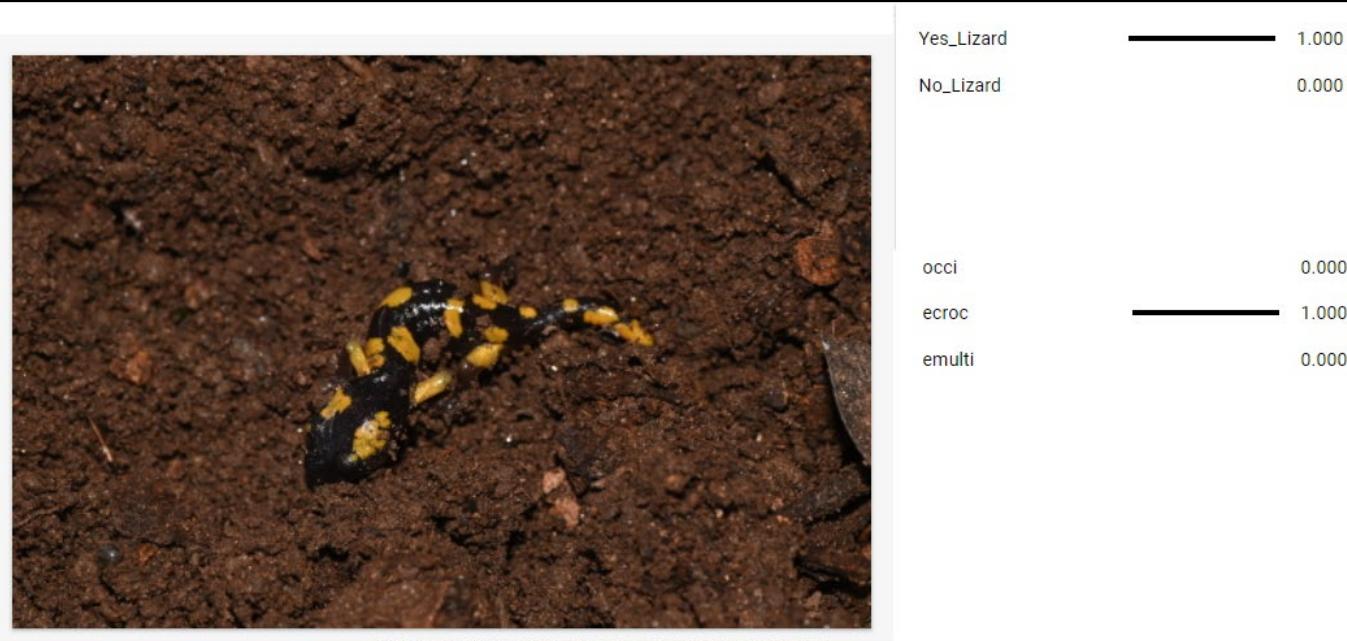
And Now, A Quick Demonstration

E. multicarinata



And Now, A Quick Demonstration

E. e. croceater



And Now, A Quick Demonstration

No Lizard



Yes_Lizard 0.000
No_Lizard 1.000



THE WHY:

- Camara trap technology produce mass amounts of data yield due to time-lapse photos and faulty motion detection.
- A ML Model can “Filter” out false positive detections fed to it and sort true positives to target save directories.
- Replace costly and intrusive manual field surveys with low maintenance long term automated camera trap deployment...
- The ML Models Will Be Customizable By Region For the Detection of Rare and Hard-to-Find Species!



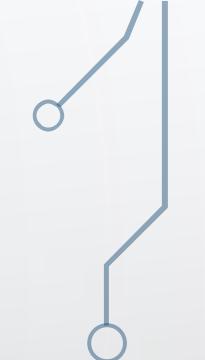
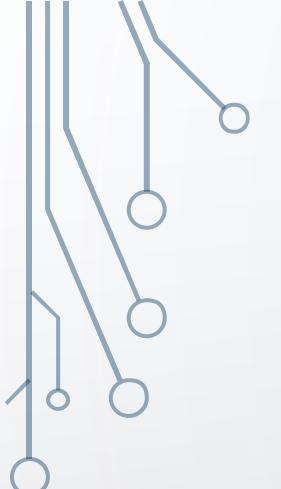
UC SANTA BARBARA
Undergraduate Research
& Creative Activities



ACKNOWLEDGEMENTS

- **Chris Evelyn**
- **Constance Woodman**
- **Connor Foley**
- **Katja Setlmann**
- **Doug Bonham**
- **Dylan Burkey**
- **Malik Elkouby**
- **Charley Lindahl**
- **Don Brightsmith**





**THANK YOU SO MUCH FOR
LISTENING!**

FOR FURTHER QUESTIONS EMAIL ME!



sethfrazer@ucsb.edu

