

# ANimal Deep learning Identification -ANDI

April 22, 2022

Jean-François Baril, Jonathan Beaulieu Emond, Yassine Kassis, Selim Gilon

Supervised by: Victor Schmidt



# Agenda

- 1 **Problem definition & Motivation**
- 2 **The Dataset**
- 3 **Approach**
- 4 **Experiments**
- 5 **Final Results**
- 6 **Conclusions**
- 7 **Future Work**

# 1



## **Problem Definition & Motivation**

# Problem Definition & Motivation

- Automatically taken pictures of animals
- Potential application: Camera Traps, labelling tool
- Goals:
  - Animal classification
  - Animal detection
  - Portability (time and size)
  - Data augmentation



# 2



## The Dataset

# CalTech Camera Trap Subset



## Dataset :

- 62K images from 85 locations in the American SW
- 19 animal classes + empty
- Heat/motion-triggered cameras

## Challenges :

- Luminosity
- Blurriness
- Bad focus
- Small ROI
- Perspective
- label imbalance
- background invariance

# 3



## Approach

# Chosen Models

Models	Classification	Detection	#of parameters (M)	Layers
VGG19	Yes	No	143.7	19
AlexNet	Yes	No	62	11
ResNext50	Yes	No	41.5	50
YOLO v5 (N)	Yes	Yes	1.9	191
YOLO v5 (S)	Yes	Yes	7.2	191
YOLO v5 (M)	Yes	Yes	21.2	191
YOLO v5 (L)	Yes	Yes	46.5	191
YOLO v5 (XL)	Yes	Yes	86.5	191
EfficientDetv2	Yes	Yes	52	247



# Ressources & Hyperparameters

- Computation:
  - Google Colab Pro (Tesla P100)
  - RTX 3060
- Experiment tracker: [Weights & Biases](#)
- Dataset storage & annotating : [Roboflow](#)
- Max 12 hours of training or 50 epochs
- Patience of 5 (consecutive)
- Resized to 320x320 pixels

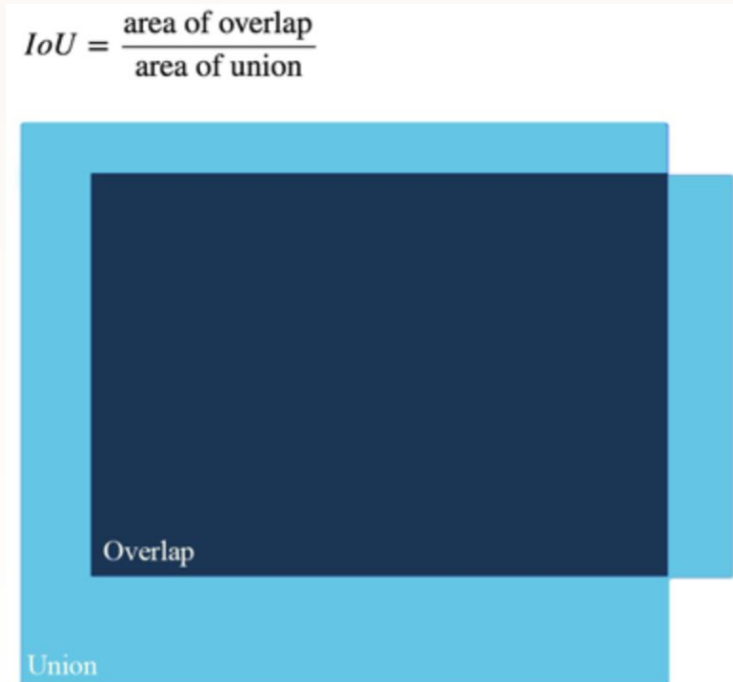
 **Weights & Biases**



# Evaluation Metrics

- F1 macro (= weight to each class)
- mAP (IoU  $\geq 0.5$ )

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$



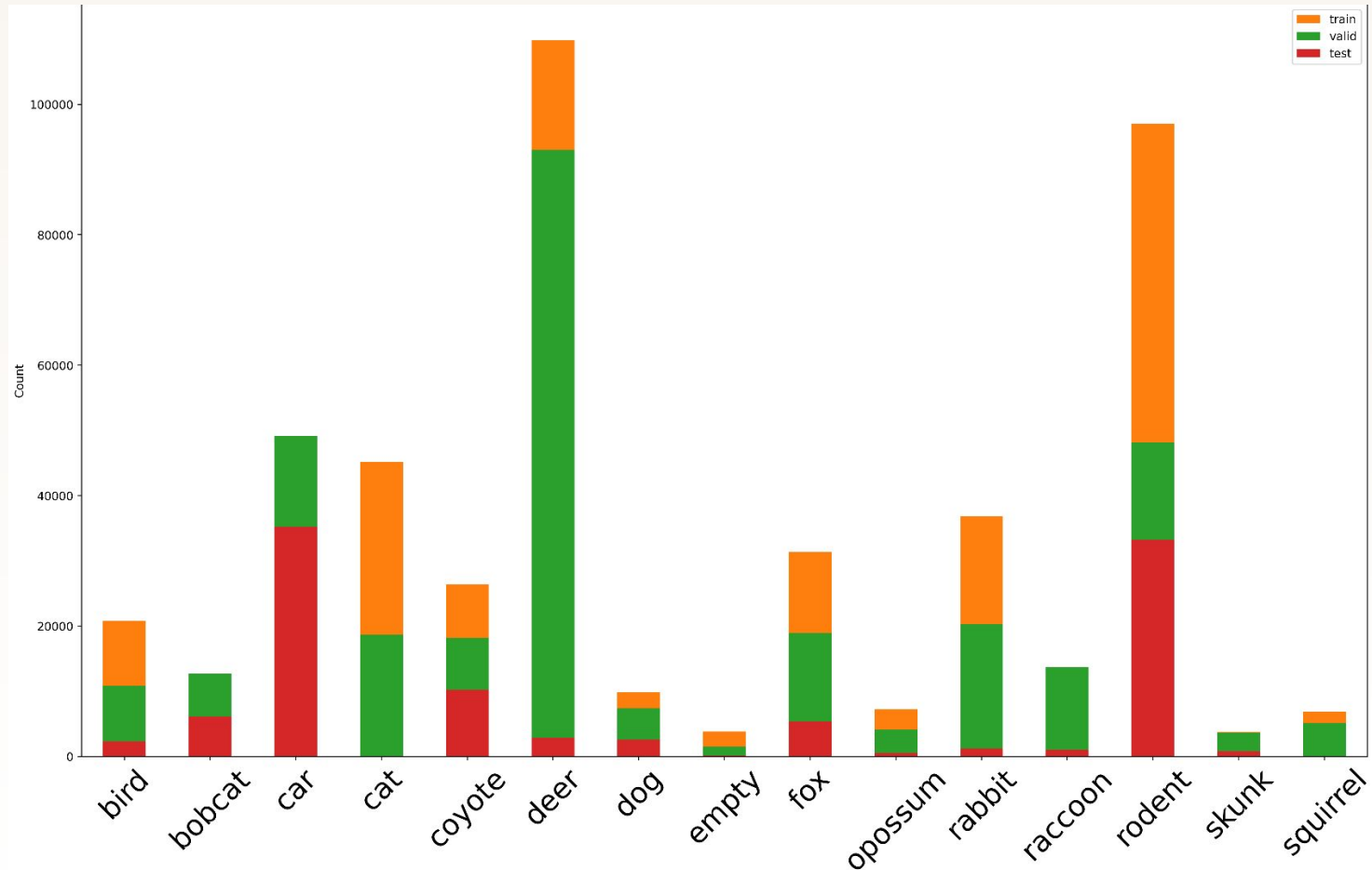
[source](#)

# 4



## Experiments

# Dataset - 1st Split



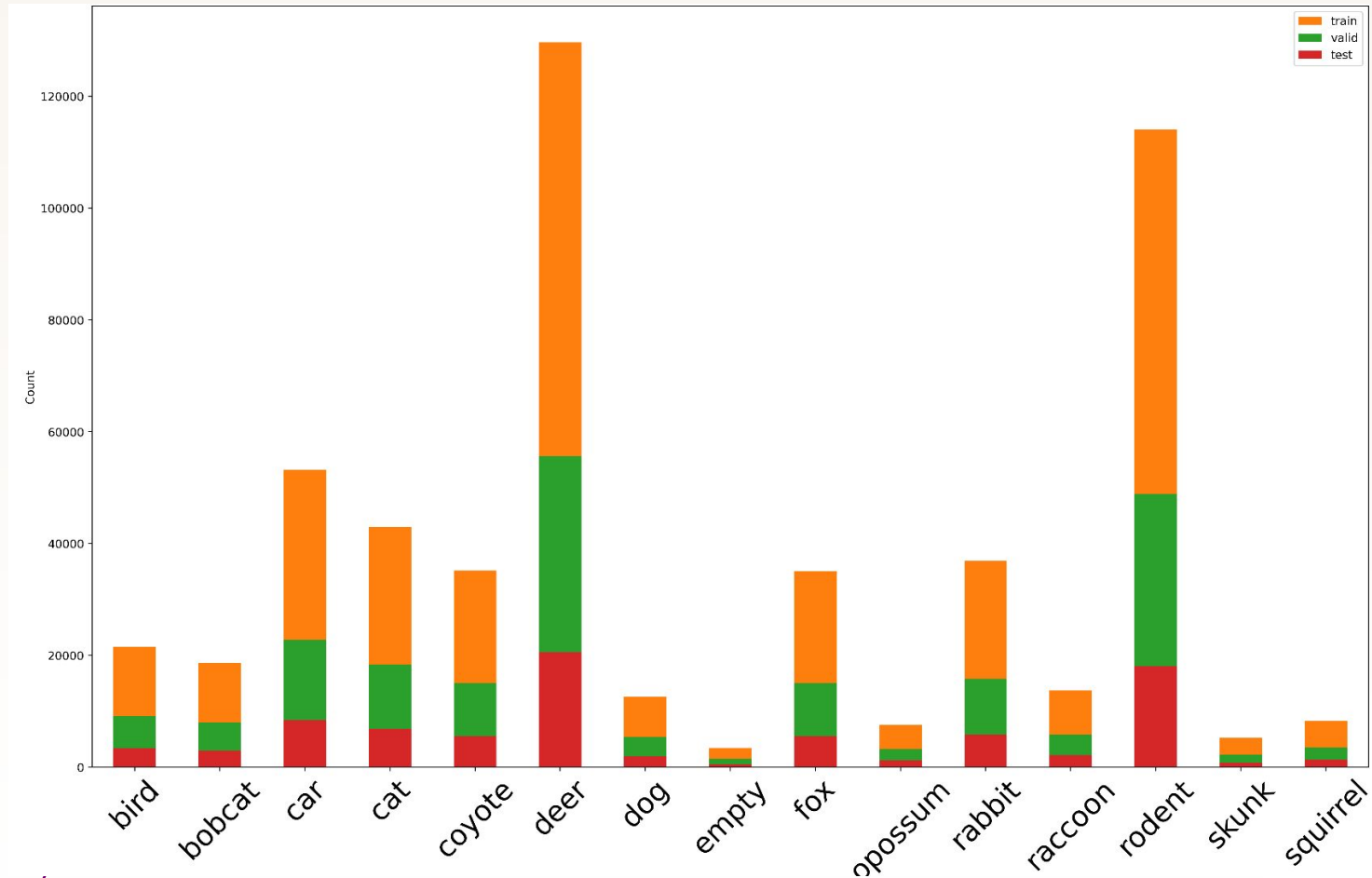
- Split by locations
- Train, valid, test: 46% - 44% - 10%
- Not all classes in training/validation

# Results - Data 1.0

<u>Model</u>	F-score	mAP 0.5	Training-time(h)
Alexnet	0.20	X	<b>0:54</b>
<del>VGG19</del>	<del>0.18</del>	<del>X</del>	<del>5:33</del>
Resnet	<b>0.21</b>	X	2:45 (+)
<del>Efficient Det</del>	<del>0.57</del>	<del>0.18</del>	<del>12:19</del>
Yolov5 N	0.46	0.448	<b>4:15 (+)</b>
Yolov5 S	0.53	0.532	5:02 (+)
Yolov5 M	0.57	0.575	4:34 (+)
<del>Yolov5 L</del>	<del>0.58</del>	<del>0.584</del>	<del>7:17 (+)</del>
<del>Yolov5 X</del>	<del><b>0.61</b></del>	<del><b>0.616</b></del>	<del>12:33 (+)</del>

trained on : (\*) Google Colab (tesla p100); (+) RTX 3060

# Dataset - 2nd Split



- Equal split among classes
- Train, valid, test: 65% - 25% - 10%

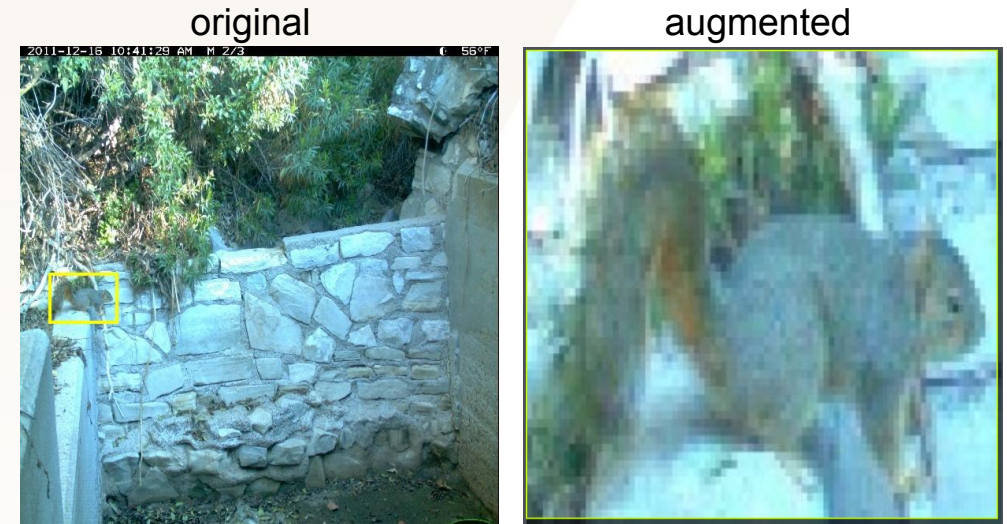
# Results - Data 2.0

<u>Model</u>	F-score	mAP 0.5	Training-time(h)
Alexnet	0.85	X	3:48 (+)
Resnet	<b>0.88</b>	X	5:17 (+)
<del>Yolov5 N</del>	<del>0.86</del>	<del>0.876</del>	<del>4:33 (+)</del>
<del>Yolov5 S</del>	<del>0.90</del>	<del>0.921</del>	<del><b>4:23 (+)</b></del>
Yolov5 M	<b>0.91</b>	<b>0.935</b>	8:07 (+)

trained on : (\*) Google Colab (tesla p100); (+) RTX 3060

# Dataset 2nd Split - Modification

- Upsampling (data 3.0)
  - on least performant classes
  - x 2
- Augmentation (data 4.0)
  - On least performant classes
  - Object Isolation
- Increase the image resolution (data 608)
  - from 320x320 to 608x608





# Results - Data 3.0 & 4.0

	3.0			4.0			608		
<u>Model</u>	F-score	mAP 0.5	Training- time(h)	F-score	mAP 0.5	Training- time(h)	F-score	mAP 0.5	Training- time(h)
Alexnet	<del>0.87</del>	X	<del>1:11 (+)</del>	0.86	X	1:07 (+)	X	X	X
Resnet	<b>0.90</b>	X	11:10(+)	<del>0.84</del>	X	<del>10:10 (+)</del>	X	X	X
Yolov5 M	<b>0.91</b>	<b>0.93</b>	<b>2:37 (+)</b>	<del>0.90</del>	0.93	<del>3:06 (+)</del>	<del>0.88</del>	0.88	<del>5:52</del>

# 5



## Final Results

# Results - Test Set : Seen Locations

<u>Model / Metric</u>	F1	mAP 0.5	Inference time/image
Resnet	0.93	X	0.01 s
Yolov5 M	<b>0.98</b>	<b>0.88</b>	0.01 s

# Unseen locations Test Set

- Collected images from 51 new locations (known animals)
- Around 100 images/class
- Hand Labellisation
- To evaluate on new background
- Shared with other group for performance comparison

Labelled				Unlabelled	
<ul style="list-style-type: none"><li>• 85 locations</li><li>• 62K images</li></ul>				<ul style="list-style-type: none"><li>• 55 unseen locations</li><li>• Hand-labelled</li><li>• 180K images</li></ul>	
<i>Train</i>	<i>Valid</i>	<i>Seen Test</i>	<i>New Test</i>		

# Results - Test Set unseen locations Yolov5M



# Results - Test Set unseen locations ResNext50



# Results - Test Set : unseen Locations

<u>Model / Metric</u>	F1	mAP 0.5	Inference time/image
Resnet50	0.35	X	0.01 s
Yolov5 M	0.76	0.68	0.01 s

# 6



## Conclusions



# Conclusions

- Among all the tested models
  - Best one (F1, mAP, size): YOLOv5 M
  - Quickest classifier: Alexnet
  - Most portable: YOLOv5 N
- Data upsampling on small animals is efficient
- Performant on unseen locations (not only memorized background)
- Could be implemented in embedded systems for live on-site detection

7



**Future Work**

# Future Work

---

- Neural magic (Prune the models)
- Color jittering
- Spurious Correlations
- Image segmentation





Questions

