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



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





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



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)
 - o RTX 3060
- Experiment tracker: <u>Weights & Biases</u>
- Dataset storage & annotating : <u>Roboflow</u>
- Max 12 hours of training or 50 epochs
- Patience of 5 (consecutive)
- Resized to 320x320 pixels

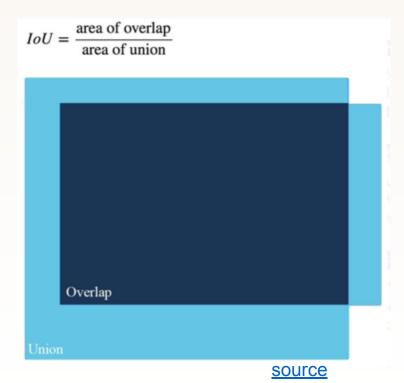






Evaluation Metrics

- F1 macro (= weight to each class)
- mAP (loU >= 0.5)

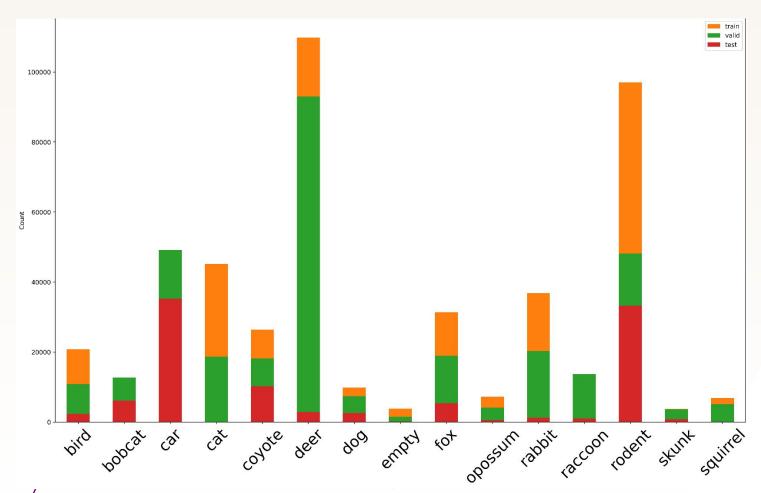


$$F_1 = 2 \cdot rac{ ext{precision} \cdot ext{recall}}{ ext{precision} + ext{recall}}$$



Experiments

Dataset - 1st Split



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

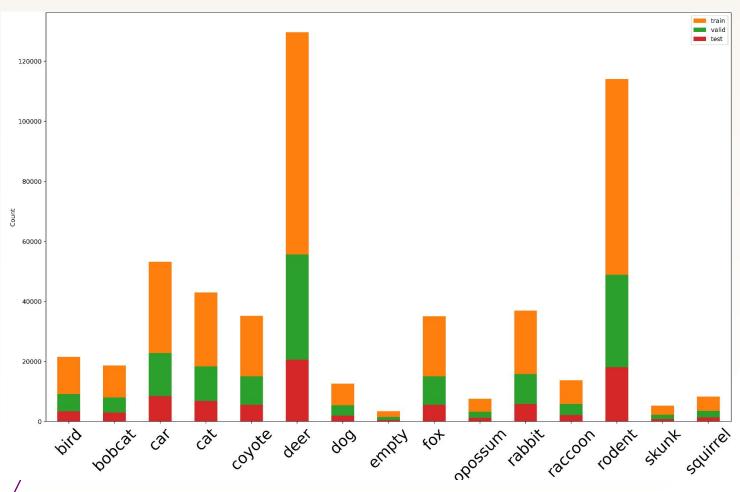


Results - Data 1.0

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



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	0.88	X	5:17 (+)		
Yolov5 N	0.86	0.876	4:33 (+)		
Yolov5 S	0.90	0.921	4:23 (+)		
Yolov5 M	0.91	0.935	8:07 (+)		



Dataset 2nd Split - Modification

- Upsampling (data 3.0)
 - on least performant classes
 - o x 2
- Augmentation (data 4.0)
 - On least performant classes
 - Object Isolation
- Increase the image resolution (data 608)
 - o 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	-0.87	X	1:11 (+)	0.86	X	1:07 (+)	Х	Х	X	
Resnet	0.90	X	11:10(+)	-0.84	X	10:10 (+)	Х	X	Х	
Yolov5 M	0.91	0.93	2:37 (+)	-0.90	0.93	3:06 (+)	-0.88	0.88	5:52	



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	0.98	0.88	0.01 s



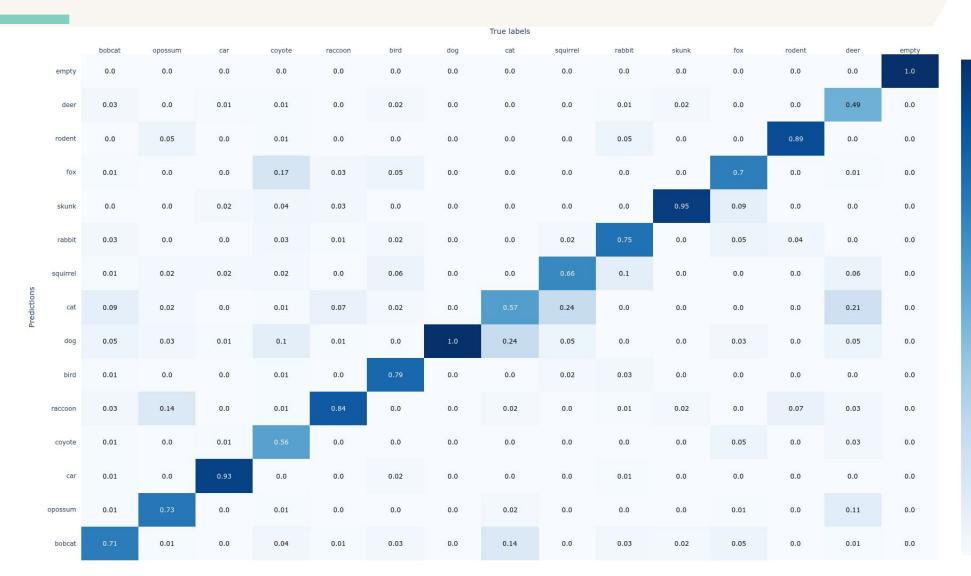
Unseen locations Test Set

- 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

La	abelled		Unlabelled
85 loc62K irTrain		Seen Test	 55 unseen locations Hand-labelled 180K images



Results - Test Set unseen locations Yolov5M





0.4

0.2

Results - Test Set unseen locations ResNext50

								True labels							
	bobcat	opossum	car	coyote	raccoon	bird	dog	cat	squirrel	rabbit	skunk	fox	rodent	deer	empty
empty	0.01	0.0	0.0	0.03	0.0	0.26	0.0	0.0	0.07	0.15	0.01	0.0	0.16	0.04	0.1
deer	0.04	0.0	0.0	0.07	0.0	0.03	0.0	0.0	0.0	0.0	0.01	0.06	0.02	0.39	0.04
rodent	0.01	0.0	0.0	0.02	0.05	0.0	0.0	0.0	0.0	0.1	0.03	0.08	0.37	0.02	0.07
fox	0.07	0.0	0.0	0.16	0.05	0.03	0.0	0.0	0.0	0.02	0.0	0.47	0.05	0.04	0.07
skunk	0.09	0.0	0.0	0.01	0.08	0.0	0.0	0.0	0.0	0.0	0.7	0.06	0.05	0.01	0.03
rabbit	0.01	0.0	0.0	0.07	0.0	0.08	0.14	0.0	0.04	0.21	0.04	0.06	0.06	0.07	0.07
squirrel	0.04	0.01	0.0	0.05	0.0	0.02	0.14	0.0	0.59	0.09	0.0	0.0	0.0	0.03	0.07
Predictions	0.06	0.15	0.0	0.01	0.1	0.02	0.14	0.33	0.0	0.11	0.03	0.0	0.03	0.04	0.12
dog	0.07	0.02	0.0	0.11	0.03	0.08		0.33	0.04	0.1	0.01	0.0	0.02	0.06	0.08
bird	0.03	0.0	0.0	0.01	0.0	0.33	0.0	0.0	0.07	0.12	0.01	0.0	0.05	0.02	0.04
raccoon	0.03	0.25	0.0	0.03	0.61	0.0	0.0	0.17	0.0	0.02	0.06	0.03	0.02	0.01	0.05
coyote	0.06	0.0	0.0	0.34	0.02	0.02	0.0	0.17	0.0	0.02	0.01	0.08	0.0	0.08	0.05
car	0.03	0.0	1.0	0.02	0.0	0.08	0.0	0.0	0.0	0.02	0.06	0.0	0.0	0.15	0.07
opossum	0.07	0.54	0.0	0.03	0.05	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.17	0.01	0.05
bobcat	0.35	0.02	0.0	0.05	0.0	0.05	0.0	0.0	0.19	0.05	0.01	0.17	0.02	0.03	0.08



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



Conclusions

Conclusions

- Among all the tested models
 - Best one (F1, mAP, size): Yolov 5 M
 - Quickest classifier: Alexnet
 - Most portable: YoloV 5 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



Future Work

Future Work

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



Questions

