WILDS - Distribution shifts - iwildcam

Vijaya Lakshmi Kuruba
Saif Kurdi-Teylouni
Chaitanya Kothapalli

Agenda:

- Problem Definition & Motivation
- Dataset
- Domain generalisation algorithms
- Data augment techniques
- Model architecture
- Results
- Results Interpretation
- Future works

Problem definition :

General assumption in Machine learning :

Train Data Distribution = Test Data Distribution

Distribution shifts can cause models to fail :

Train Data Distribution # Test Data Distribution

Train - Images from some locations

Test - Images from the same location

(High accuracy)

Test - Images from other location

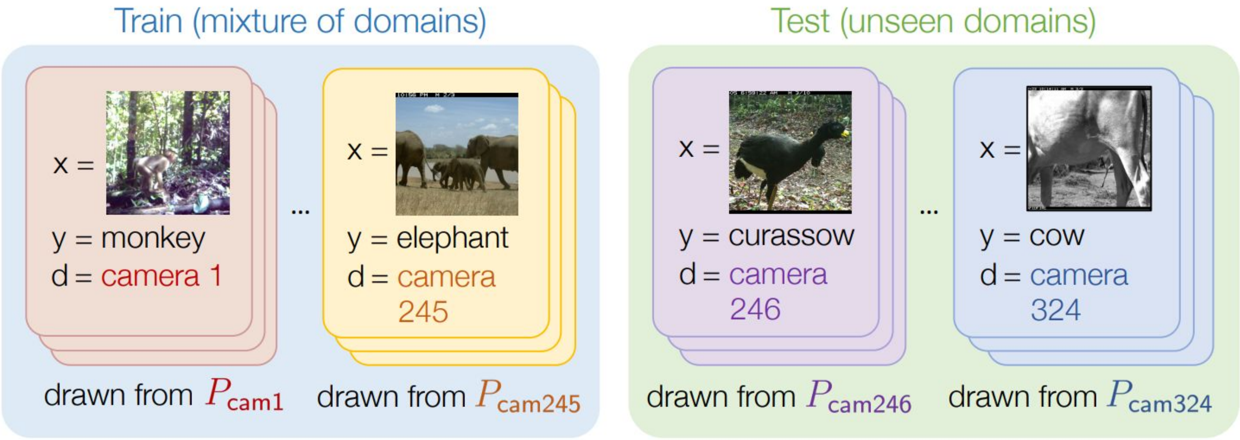
(Low accuracy)

Animal Classification in Wildlife monitoring & conservation :

- ML models are used to process, understand and monitor wildlife biodiversity loss with data from deployed camera traps.

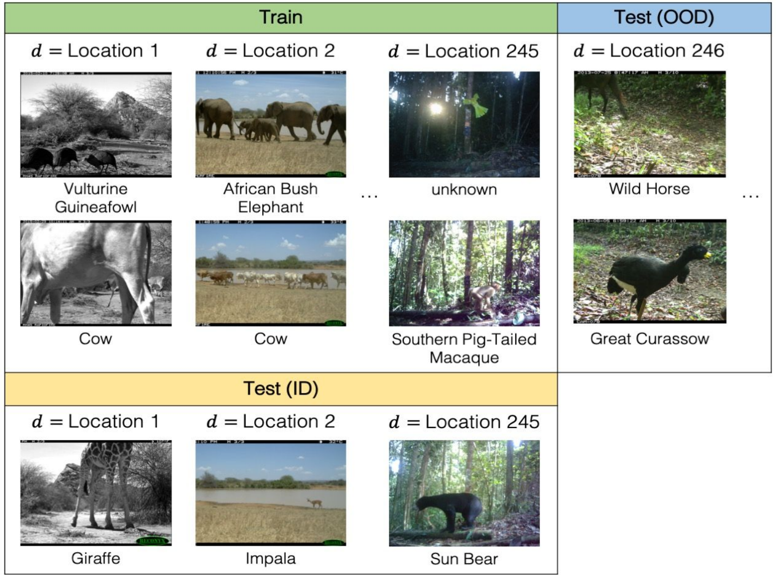
- Typical ML models would generalise poorly to new camera Trap deployments due to variations in illumination, camera angle, background, vegetation, color, and relative animal frequencies.

- This distribution shift is a example of domain generalisation.



Dataset:

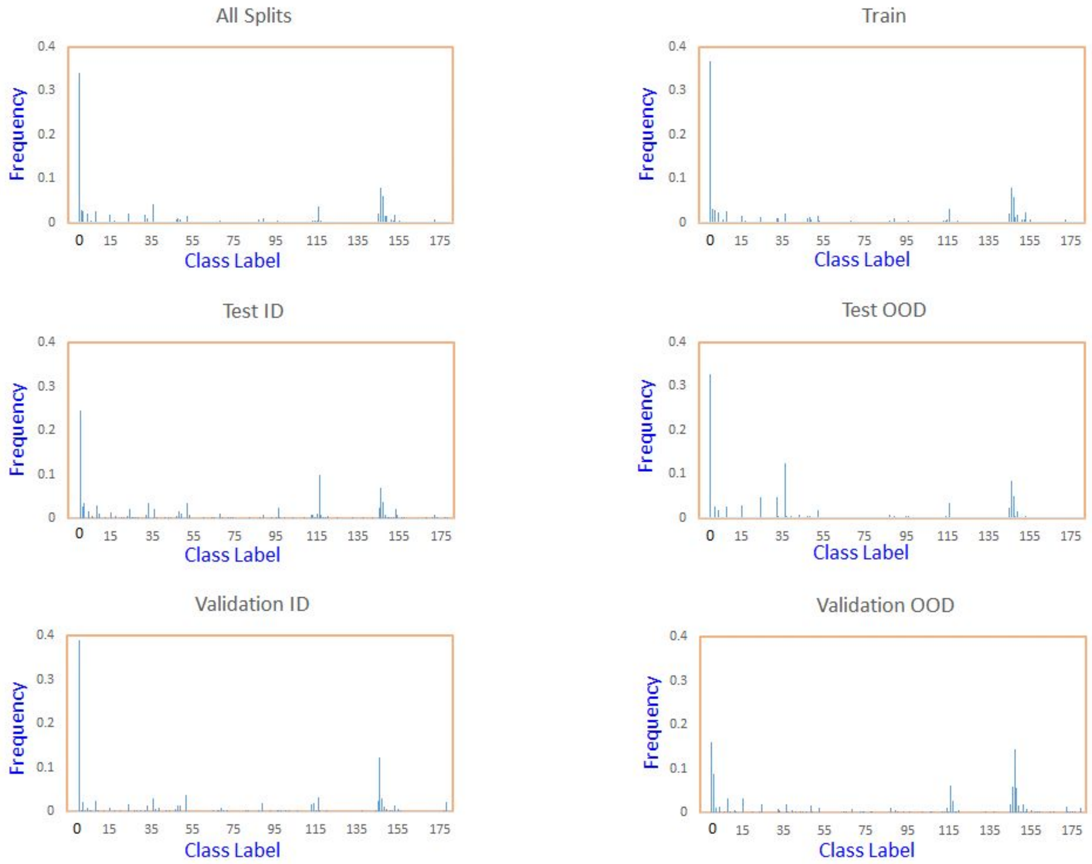
- Dataset - iWildCam2020
- OOD - Measured performance on default wild splits
- ID - Measured without distribution shift on held out dataset



* In-Distribution (ID), Out-of-Distribution (OOD)

Dataset	Train		Validation (OOD)		Test (OOD)		Validation (ID)		Test (ID)	
	N_Img	N_traps	N_Img	N_traps	N_Img	N_traps	N_Img	N_traps	N_Img	N_traps
Full	129,809	243	14,961	32	42,791	48	7,314	Same traps as training set but on different days from train and test (ID) images	8,514	Same traps as training set but on different days from train and val (ID) images
Partial	51,924	238	5,984	32	17,116	48	2,926		3,262	

Label distribution for each iWildCam2020 split experimental data set



- Approximately 35% of the total number of images are empty - Class 0

Metrics:

- Highly Imbalanced dataset
- Accuracy is not a good metric to weigh models with unbalanced datasets.
- Macro F1 score to better capture model performance on rare species.

Project Setup & Approach :

Phase 1 : Experiments with Full dataset

- Reproduce baseline results
- Use Data augmentation techniques

Phase 2 : Experiments with Partial dataset

- Create new baseline results
- Use Data augmentation techniques
- Use different Model Architecture
- Comparison and Interpretation

- 182 classes - full data set vs 169 classes in partial dataset. Same number of camera traps are retained.

Domain Generalisation Algorithms :

- ERM - Minimizes the average loss across dataset

Domain Generalisation algorithms : ERM + Penalty to capture invariance , Uses Domain annotations during training

- CORAL -Penalizes differences in the means and covariances of the feature distributions (i.e., the distribution of last layer activations in a neural network) for each domain.

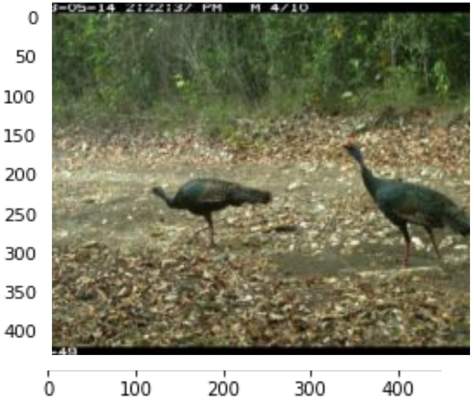
- IRM - To learn invariances across environments, find a data representation such that the optimal classifier on top of that representation matches for all environments.

Data Augmentation :

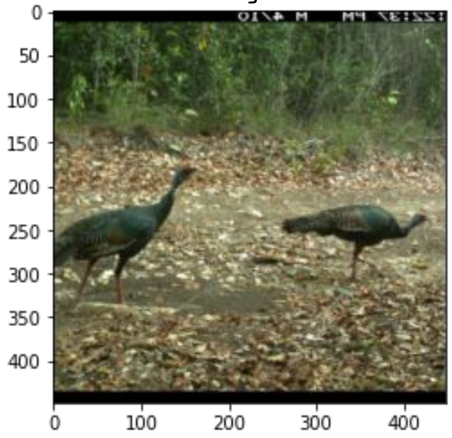
Original Image (W x 448)

* W > 448

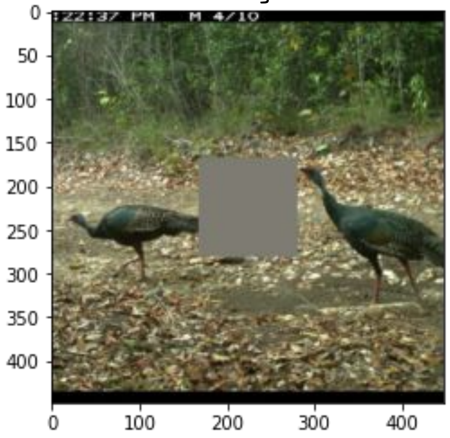
Image size for Training - (448 x 448)



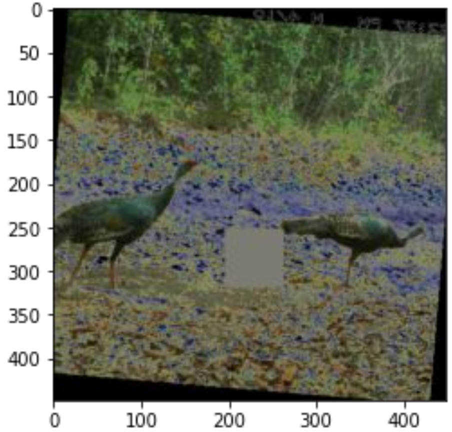
Weak Augment



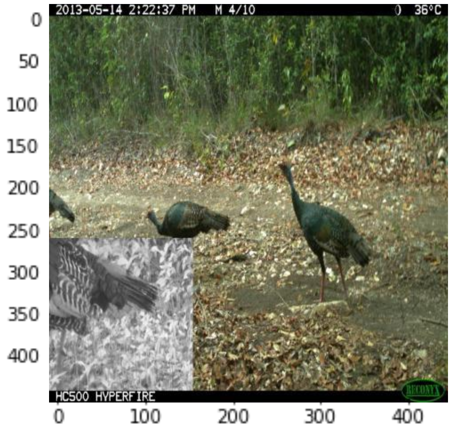
Cutout Augment



Strong Augment



Cut Mix Augment



Model Architecture :

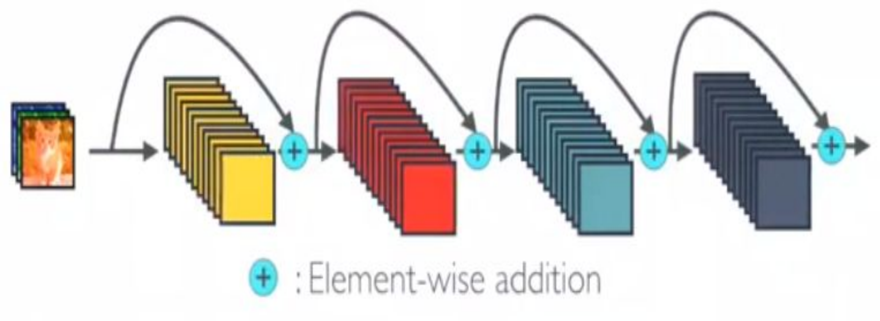
Provides the working parameters—such as the number, size, and type of layers in a neural network.

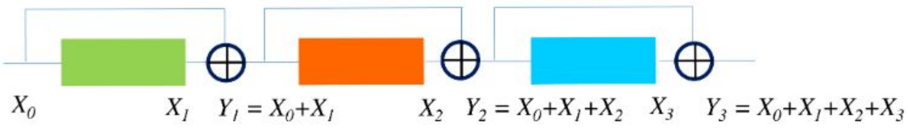
- ResNet
  - Convolutional neural network that utilises "identity shortcut connection" that skips one or more layers.
  - Tackles vanishing gradient issue arising with layer stacking.
  - Allows for very deep architectures, over a hundred layers.

- DenseNet
  - Convolutional neural network that utilises dense connections between layers, through Dense Blocks, where all layers (with matching feature-map sizes) are connected directly with each other
  - Concatenation is used. Each layer is receiving a "collective knowledge" from all preceding layers.
  - Since each layer receives feature maps from all preceding layers, network can be thinner and compact, i.e. number of channels can be fewer

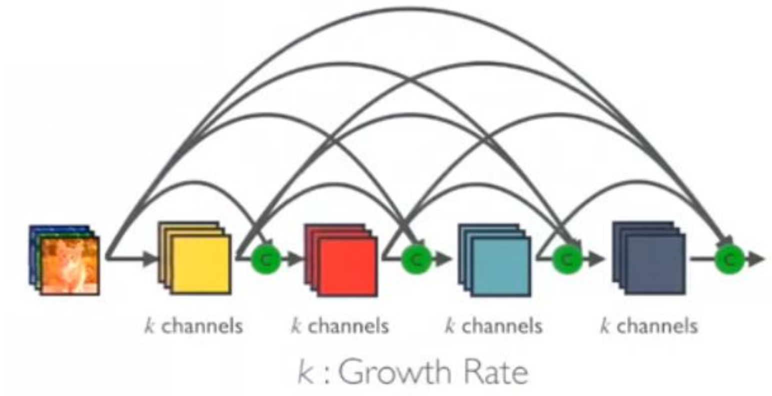
ResNet vs. DenseNet:

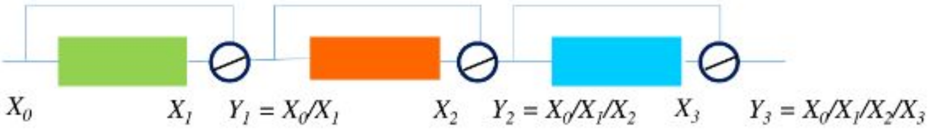
ResNet





DenseNet





Results

Baseline Experiment Results (Phase 1) :

Baseline Vs Reproduced Results

Model	Algorithm	Augmentation Method	Test ID Macro F1	Test OOD Macro F1
ResNet 50	ERM	Reproduced results	42.2	33.5
		Published Results	47.0 (1.4)	31.0 (1.3)
	IRM	Reproduced results	15.1	10.9
		Published Results	22.4 (7.7)	15.1(4.9)
	CORAL	Reproduced results	42.8	28.8
		Published Results	43.5 (3.5)	32.8 (0.1)

Phase 1 - Results

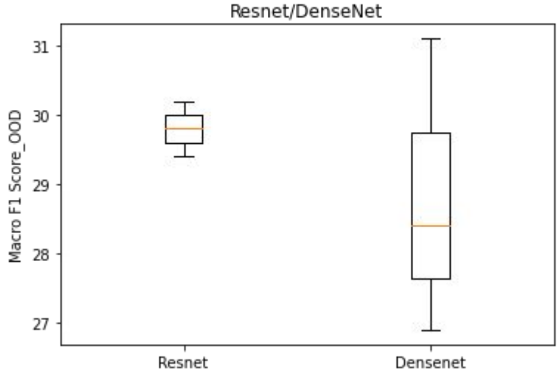
Model	Algorithm	Augmentation Method	Test ID Macro F1	Test ID Avg Acc	Test OOD Macro F1	Test OOD Avg Acc
ResNet 50	ERM	Strong	46.0	77.4	30.1	73.7
		Weak	**47.3**	**76.9**	32.6	74.1
		Cutout	41.5	74.8	29.9	68.2
		Cutmix	43.5	71.7	26.8	75.2
		No Augmentation	42.2	75.8	33.5	75.5
	IRM	Weak	20.3	59.0	12.5	53.3
		Cutout	22.9	59.1	16.6	63.6
		Cutmix	33.6	70.3	22.3	64.1
		No Augmentation	15.1	50.8	10.9	52.9
	CORAL	Strong	44.5	74.9	34.9	74.9
		Weak	44.6	75.0	33.2	73.3
		Cutout	46.6	74.7	**35.1**	**70.4**
		Cutmix	34.0	65.8	28.1	71.8
		No Augmentation	42.8	73.6	28.8	72.5

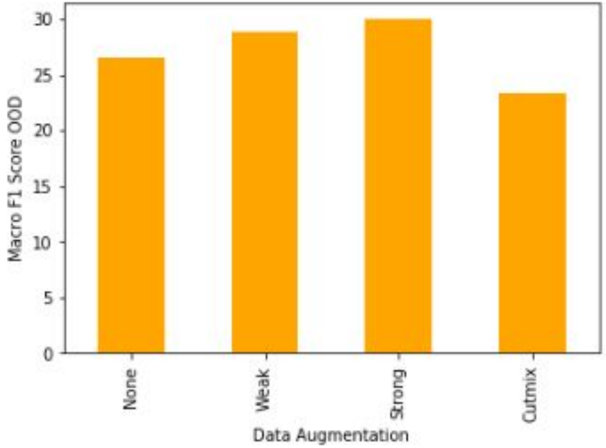
Selection criterion for further analysis :

Model: ResNet 50 Agorithm: ERM Augmentation: None							
Experiment	**Target Image Resolution**	**Dataset (%)**	**Run Time (Hrs)**	**Test ID Macro F1**	**Test ID Avg Acc**	**Test OOD Macro F1**	**Test OOD Avg Acc**
Baseline	(448, 448)	100	15.5	42.2	75.8	33.5	75.5
Exp 1	**(448, 448)**	**40**	**4.0**	**43.9**	**75.6**	**27.8**	**66.4**
Exp 2	(224, 224)	100	4.0	41.7	73.2	25.5	59.5
Exp 3	(224, 224)	40	2.25	36.0	73.0	22.1	67.6

Experiments with partial dataset (Phase 2) :

Model	Algorithm	Augmentation Method	Test ID Macro F1	Test ID Avg Acc	Test OOD Macro F1	Test OOD Avg Acc
ResNet	ERM	No Augmentation	44.94 (1.06)	76.18 (0.42)	26.63 (1.72)	70.23 (1.79)
		Weak	**47.50 (1.37)**	77.43 (1.40)	**29.80 (0.40)**	72.83 (5.06)
		Strong	43.60 (2.39)	76.93 (0.76)	**29.37 (0.50)**	69.93 (3.45)
		Cutmix	36.16 (1.43)	70.70 (0.92)	25.83 (0.83)	70.60 (2.83)
	CORAL	No Augmentation	38.21 (1.89)	70.76 (2.25)	28.04 (0.63)	73.72 (3.76)
		Weak	**39.10 (3.70)**	72.76 (1.89)	**29.01 (1.08)**	69.84 (3.67)
		Strong	37.67 (3.30)	71.21 (4.01)	26.10 (1.50)	70.18 (0.80)
		Cutmix	22.93 (1.03)	57.30 (1.74)	19.03 (0.59)	66.26 (3.07)
DenseNet	ERM	No Augmentation	42.90 (0.36)	75.70 (1.30)	26.58 (2.42)	72.97 (3.73)
		Weak	**42.73 (1.80)**	77.33 (0.55)	**28.80 (2.13)**	73.80 (2.11)
		Strong	39.73 (2.48)	77.50 (0.17)	**29.98 (1.67)**	75.87 (2.13)
		Cutmix	35.83 (1.46)	75.51 (4.52)	23.42 (0.61)	73.34 (1.56)
	CORAL	No Augmentation	37.03 (1.36)	71.24 (2.49)	25.78 (1.67)	71.63 (1.64)
		Weak	**37.16 (1.42)**	73.35 (1.39)	**26.73 (0.89)**	69.59 (0.55)
		Strong	32.53 (2.34)	68.00 (2.34)	24.72 (2.37)	74.44 (0.60)
		Cutmix	21.56 (0.30)	53.86 (4.27)	16.63 (0.16)	66.30 (0.92)





Results Interpretation

Observations on species classification with Data augmentation

Without Augmentation				
Class	precision	recall	f1-score	support
0	0.920	0.928	0.924	5589
2	0.617	0.550	0.581	433
6	0.200	0.125	0.154	24
8	0.544	0.690	0.608	464
12	0.000	0.000	0.000	7
17	0.032	0.096	0.048	52
20	1.000	0.075	0.140	53
23	0.000	0.000	0.000	52
27	0.667	0.154	0.250	13
33	0.833	0.672	0.744	67
36	0.907	0.271	0.418	2122

With Augmentation				
Class	precision	recall	f1-score	support
0	0.946	0.939	0.942	5589
2	0.708	0.633	0.668	433
6	0.286	0.583	0.384	24
8	0.571	0.744	0.646	464
12	0.182	0.286	0.222	7
17	0.643	0.173	0.273	52
20	0.600	0.170	0.265	53
23	0.556	0.096	0.164	52
27	0.500	0.385	0.435	13
33	0.808	0.881	0.843	67
36	0.873	0.806	0.838	2122

Observations on species classification with Data augmentation

Class 24 [bos taurus]

Class 36 [aepyceros melampus]

Image Class	Image Category	Without Augmentation	With Augmentation
24	bos taurus	80.3	36.3
36	aepyceros melampus	31.4	87.5

Misclassification scenarios

Existence of small animals in dark settings

Existence of birds in bushy background

Incorrect labelling

Species represented partially

Existence of a similar non rare class species in the dataset

Multiple species existing in same image

Scenarios of empty images misclassification

Barking Deer wrongly detected in Empty image

Common Warthog wrongly detected in Empty image

Great Tinamou wrongly detected in Empty image

Antelope wrongly detected in Empty image

Future work:

- Transformer based architecture (visual transformers)
- 2 step classification : Empty and Non empty images + Classification on Non empty Images
- Mega Detector - Pretrained model by Microsoft to find animals in picture that outputs bounding boxes over the animals.
  Crop these boxes and classify them