Using Pre-trained Convolutional Neural Networks to Classify Wildlife Animals

Jimmie Hagle, Gurman Gill
Department of Computer Science, Sonoma State University, Rohnert Park

Introduction:

In our research we implement a machine learning algorithm using pre-trained convolutional neural networks to classify wildlife animals. The term "trained" refers to the models algorithm weights that are adjusted to better classify a given set of training images. We compare classification accuracy using models trained on individual trail cam images versus images of all cameras. The hypothesis here is that individual cameras will perform better even though their datasets are smaller due to similarity in the data.

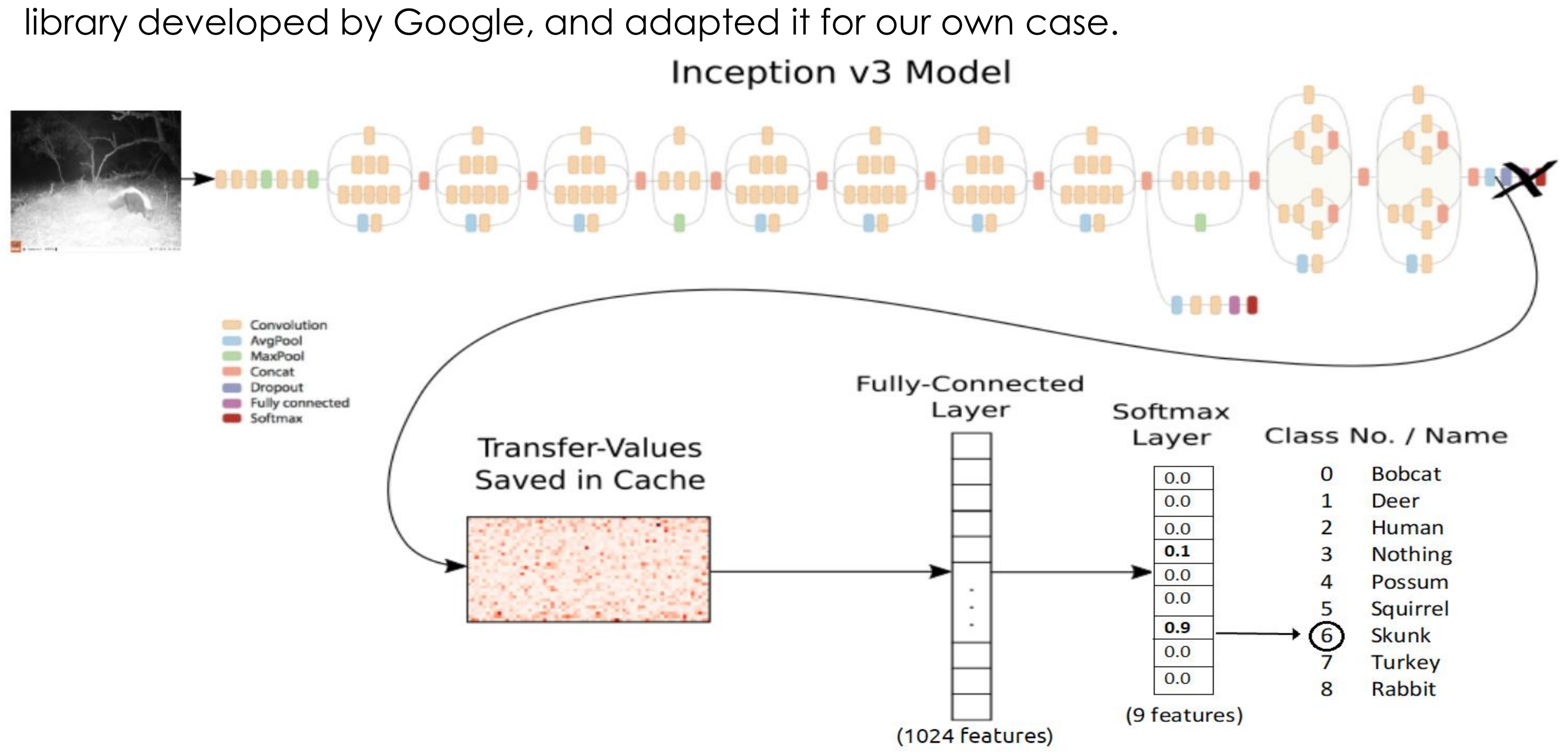
Dataset:

The dataset we used for the experiments is composed of 24973 images taken by trail cameras on SSU Preserves. The trail cameras take a series of three images every time they detect motion. The dataset contains twenty three classes of images across six camera groups. Classes with a minimum of 200 images were used in these experiments reducing the number of classes to nine. This is done to prevent a negative influence on the systems classification accuracy by excluding classes that have a low amount of images that can be trained on. All the images below are examples of the database and underneath the class name is the confidence the trained CNN had in its prediction

File: human/LowerTrailCamera_EK004212.JF True: human Pred: human (1.000)
File: squirrel/NorthernTowerMeadowCamera_EK000144.JP True: squirrel Pred: squirrel (0.962)
File: skunk/UpperTrailCamera_EK000221.JPG True: skunk Pred: skunk (0.962)
File: deer/UpperMostROWCamera_EK002600.JPG True: deer Pred: deer (0.999)

Deep Learning:

In this work, we explore using convolutional neural network (CNN) which are successive "layers" of neurons that learn to respond to image features such as color, texture, edges. These are combined in later layers to detect larger concepts such as faces, animals, etc. In particular we implement "transfer learning" a method where features pre-trained by the CNN are pulled for the new classification system to recognize images in your own database. In order to achieve this we used CNN model architecture called InceptionV3(Google Inc) which has been pre-trained on "imageNet" a database containing 1000 different categories. In order to complete the retraining process, we used an example script provided by TensorFlow, an open source machine learning software library developed by Google, and adapted it for our own case.



Experiment Setup:

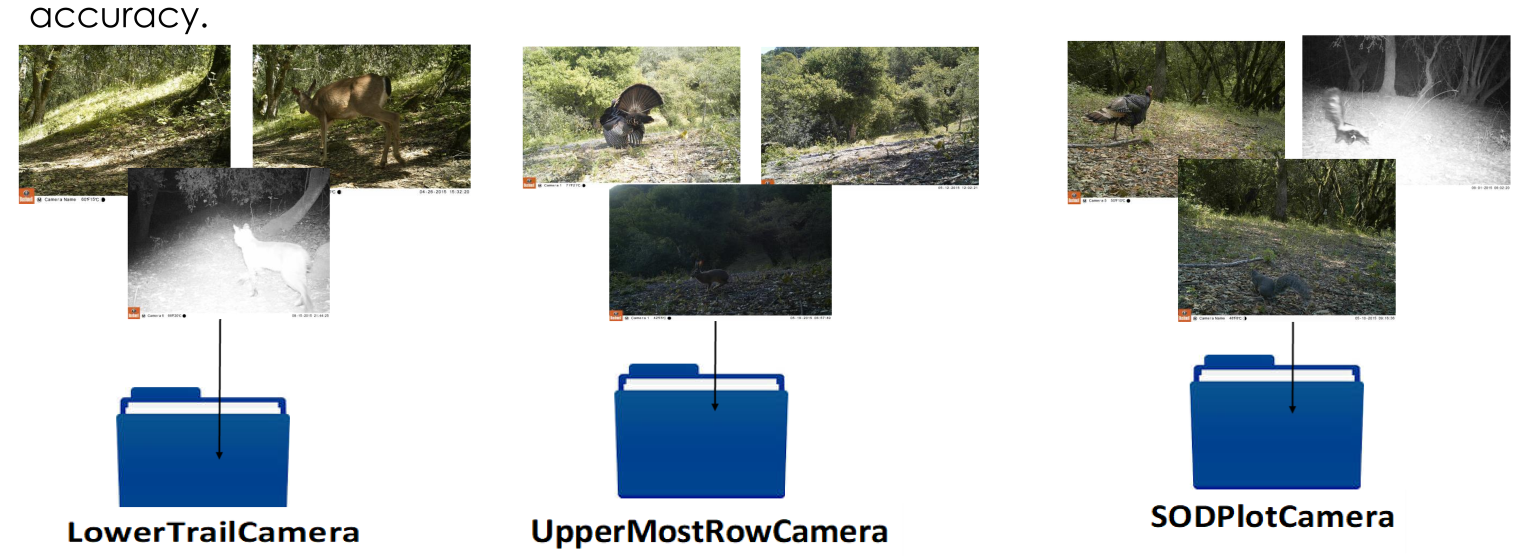
Base Experiment: In this experiment we are setting up results in which to compare our other two tests. This is run using all six camera groups with class that contain more than 200 images.

Experiment 1: In this experiment we split the images among their individual camera groups but continue to use the same nine classes.

Experiment 2: In this experiment we apply a 200 image minimum to each animal class for the individual cameras.

Evaluating on Individual Cameras:

In order to try to improve model accuracy, we try to obtain classifications on individual cameras. The reasoning behind this is that even though the size of individual datasets would decrease, the influence by the various backgrounds would decrease as well. What we consider as background is anything else in the image that is not the animal we are trying to classify. In some cases animals of a single class may more likely be seen in certain areas thus keeping the larger portion of images on a specific camera hopefully allowing for better accuracy.



Results:

Base Experiment

The following results are the model's accuracy on the test subset across all cameras.
5-fold Validation
Average Accuracy = correctly classified in a class/ total images of class
Average Accuracy overall: 0.8579 | Standard Deviation: 0.0063

Average F1 Score overall = (2 * precision * recall)/(precision + recall)
Average F1 Score overall: 0.8026 | Standard Deviation: 0.0887

Confusion Matrix for the 5th-fold Validation

Predicted Label (columns) / Actual Label (rows)

	bobcat	deer	human	nothing	possum	rabbit	skunk	squirrel	turkey
bobcat	59	21	0	4	1	15	1	1	0
deer	10	798	8	31	4	17	4	10	2
human	1	1	402	14	0	0	0	6	0
nothing	4	35	28	1741	27	50	33	120	26
possum	4	2	1	0	78	7	5	0	0
rabbit	8	21	0	25	6	287	4	14	5
skunk	0	4	0	1	3	3	105	0	0
squirrel	1	18	2	40	1	14	0	408	6
turkey	1	1	2	10	0	1	0	9	110

Number of Images Per class

bobcat	575
deer	4272
human	2200
nothing	10416
possum	515
rabbit	1806
skunk	590
squirrel	2397
turkey	643

Experiment 1:

Results of Individual Cameras:

LowerTrailCamera
Average Accuracy: 0.8884 | Standard Deviation: 0.0096
Average F1 Score overall: 0.8187 | Standard Deviation: 0.0789

	bobcat	deer	human	nothing	possum	rabbit	skunk	squirrel	turkey
bobcat	51	3	0	2	0	3	0	0	0
deer	0	59	3	9	0	13	1	1	0
human	2	6	333	13	0	0	1	8	0
nothing	1	3	4	862	6	22	1	24	1
possum	0	0	0	0	23	2	0	0	0
rabbit	5	0	0	1	0	77	1	4	1
skunk	0	0	0	1	2	4	37	0	0
squirrel	1	1	0	16	0	7	0	109	2
turkey	0	2	0	2	0	0	0	1	36

bobcat	235
deer	383
human	1772
nothing	4740
possum	135
rabbit	454
skunk	229
squirrel	692
turkey	195

UpperMostROWCamera
Average Accuracy: 0.7827 | Standard Deviation: 0.0237
Average F1 Score overall: 0.7001 | Standard Deviation: 0.1719

	bobcat	deer	human	nothing	possum	rabbit	skunk	squirrel	turkey
bobcat	13	2	0	0	0	2	0	0	0
deer	11	324	0	29	0	15	0	8	0
human	0	2	24	0	0	1	0	1	0
nothing	2	4	2	182	2	20	3	7	0
possum	0	1	0	2	5	4	0	0	0
rabbit	3	10	0	26	2	113	1	4	0
skunk	0	2	0	1	0	1	4	0	0
squirrel	0	0	0	10	0	2	0	40	0
turkey	0	0	0	1	0	0	0	6	9

bobcat	139
deer	1896
human	128
nothing	1163
possum	48
rabbit	767
skunk	65
squirrel	249
turkey	50

NorthernTowerMeadowCamera
Average Accuracy: 0.9064 | Standard Deviation: 0.0050
Average F1 Score overall: 0.7980 | Standard Deviation: 0.1402

	bobcat	deer	human	nothing	possum	rabbit	skunk	squirrel	turkey
bobcat	4	5	0	2	0	0	0	0	0
deer	0	123	0	11	0	0	0	0	0
human	0	0	2	0	0	0	0	1	0
nothing	0	10	0	401	2	0	0	21	1
possum	1	0	0	5	6	0	0	0	0
rabbit	0	3	0	0	0	3	0	0	0
skunk	0	1	0	0	0	0	7	0	0
squirrel	0	1	0	6	0	0	0	69	2
turkey	0	0	0	0	0	0	0	0	24

bobcat	26
deer	747
human	10
nothing	2109
possum	49
rabbit	18
skunk	45
squirrel	436
turkey	119

UperROWWoodChipFieldCamera
Average Accuracy: 0.8726 | Standard Deviation: 0.0075
Average F1 Score overall: 0.7204 | Standard Deviation: 0.1288

	bobcat	deer	human	nothing	possum	rabbit	skunk	squirrel	turkey
bobcat	3	1	0	0	0	1	1	0	0
deer	4	37	0	4	0	2	0	3	0
human	0	0	3	2	0	0	0	0	0
nothing	2	1	0	304	0	4	0	7	7
possum	0	0	0	0	0	2	0	1	0
rabbit	2	5	0	1	2	72	0	3	0
skunk	1	1	0	0	1	0	3	0	0
squirrel	0	0	0	1	0	0	0	22	0
turkey	0	0	0	3	0	0	0	0	12

bobcat	61
deer	218
human	9
nothing	1587
possum	14
rabbit	422
skunk	18
squirrel	104
turkey	158

UpperTrailCamera
Average Accuracy: 0.6053 | Standard Deviation: 0.0401
Average F1 Score overall: 0.6081 | Standard Deviation: 0.1950

	bobcat	deer	human	nothing	possum	rabbit	skunk	squirrel	turkey
bobcat	7	0	0	2	2	3	1	0	0
deer	0	29	0	5	1	0	0	0	0
human	2	0	7	0	0	0	0	0	0
nothing	0	5	0	16	5	15	0	0	0
possum	0	0	0	0	8	1	1	0	0
rabbit	0	0	0	3	1	25	0	0	0
skunk	0	0	0	1	0	3	9	0	0
squirrel	0	0	0	0	0	0	0	0	0
turkey	0	0	0	0	0	0	0	0	0

bobcat	83
deer	195
human	43
nothing	167
possum	69
rabbit	120
skunk	83
squirrel	0
turkey	0

SODPlotCamera
Average Accuracy: 0.8562 | Standard Deviation: 0.0088
Average F1 Score overall: 0.7578 | Standard Deviation: 0.2411

	bobcat	deer	human	nothing	possum	rabbit	skunk	squirrel	turkey
bobcat	1	0	0	3	0	0	1	1	0
deer	0	149	0	8	1	0	1	1	0
human	0	1	46	0	0	0	0	0	0
nothing	1	3	0	121	1	0	0	8	1
possum	0	0	0	0	33	1	0	2	0
rabbit	0	0	0	2	0	2	0	2	0
skunk	1	0	0	1	1	0	25	1	0
squirrel	0	3	0	34	1	0	0	154	0
turkey	0	0	0	0	0	0	0	0	21

bobcat	31
deer	833
human	238
nothing	650
possum	200
rabbit	25
skunk	150
squirrel	916
turkey	121

Experiment 2:

Due to some cameras capturing lower numbers of certain kinds of animals implementing a minimum image count prevents negative influences by these classes on the overall accuracy.

LowerTrailCamera
Average Accuracy: 0.8901 | Standard Deviation: 0.0090
Average F1 Score overall: 0.8224 | Standard Deviation: 0.0739

	bobcat	deer	human	nothing	rabbit	skunk	squirrel
bobcat	35	1	0	0	3	0	0
deer	1	56	0	3	4	3	0
human	0	2	332	11	1	1	5
nothing	0	6	8	855	23	8	48
rabbit	6	2	1	0	86	1	4
skunk	1	0	0	2	5	40	0
squirrel	1	0	2	14	3	0	127

bobcat	235
deer	383
human	1772
nothing	4740
rabbit	454
skunk	229
squirrel	692

UpperMostROWCamera
Average Accuracy: 0.7953 | Standard Deviation: 0.0167
Average F1 Score overall: 0.7537 | Standard Deviation: 0.0778

	deer	nothing	rabbit	squirrel
deer	282	61	21	9
nothing	10	202	15	5
rabbit	5	41	112	6
squirrel	1	16	0	29

deer	1896
nothing	1163
rabbit	767
squirrel	249

NorthernTowerMeadowCamera
Average Accuracy: 0.9064 | Standard Deviation: 0.0050
Average F1 Score overall: 0.8893 | Standard Deviation: 0.0635

	deer	nothing	squirrel
deer	128	10	4
nothing	3	417	13
squirrel	0	14	69

deer	747
nothing	2109
squirrel	436

UperROWWoodChipFieldCamera
Average Accuracy: 0.9560 | Standard Deviation: 0.0126
Average F1 Score overall: 0.9238 | Standard Deviation: 0.0454

	deer	nothing	rabbit
deer	32	6	7
nothing	5	297	12
rabbit	0	0	86

deer	218
nothing	1587
rabbit	422

SODPlotCamera
Average Accuracy: 0.8904 | Standard Deviation: 0.0160
Average F1 Score overall: 0.8926 | Standard Deviation: 0.0575

	deer	human	nothing	possum	squirrel
deer	156	0	2	0	4
human	0	50	0	0	1
nothing	0	0	81	11	30
possum	0	0	3	38	3
squirrel	2	1	3	0	182

deer	833
human	238
nothing	650
possum	200
squirrel	916

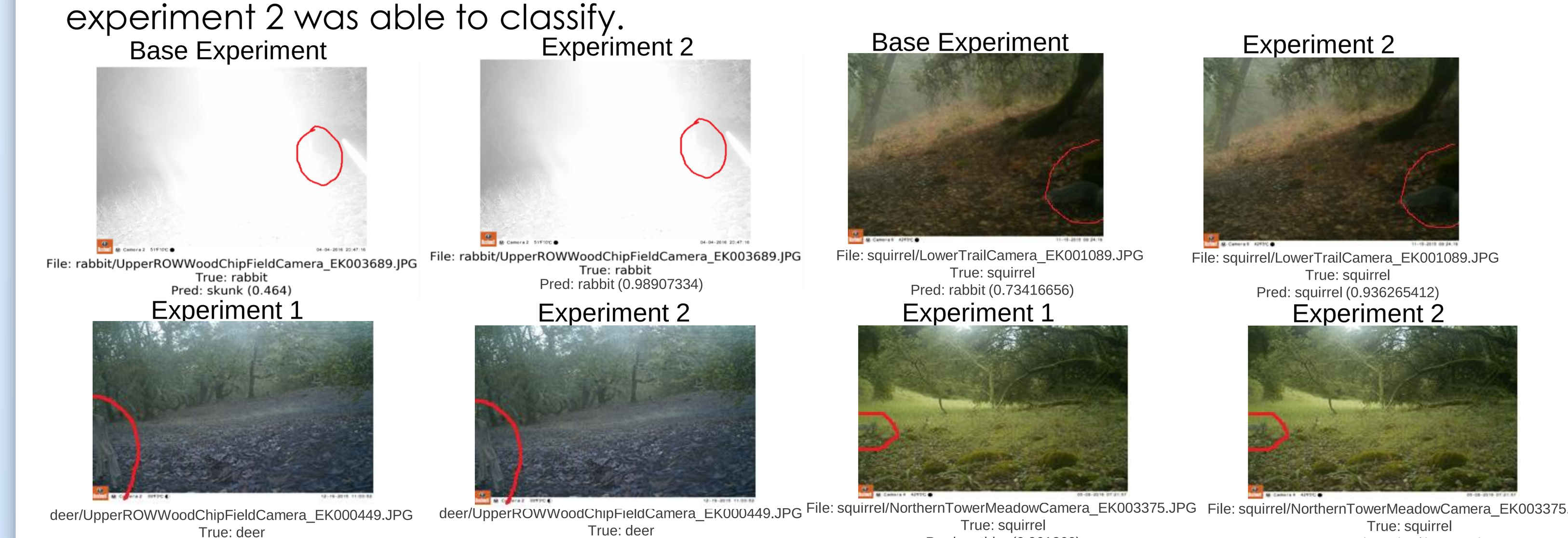
UpperTrailCamera No Classes Contain at Least 200 Images

bobcat 83
deer 195
human 43
nothing 167
possum 69
rabbit 120
skunk 83

Results Analysis:

Splitting the database up amongst the individual cameras as preformed in experiment 1 showed that certain cameras were better at classifying specific animal groups then others. UpperMostROWCamera(UMRC) is better at classifying bobcats than NorthernTowerMeadowCamera(NTMC), but NTMC is better at classifying turkeys than UMRC. This can be contributed by the differences in image quantities for these cameras. NTMC is also better at classifying skunks than UMRC even though their image quantities for these classes are similar. Meaning that the camera backgrounds are influencing these cameras differently.

Experiment 2 demonstrates an increase in classification accuracy due to the loss of influence by classes containing small image amounts. Individual camera groups were able to classify harder to classify images that the base experiment was not able to do. The image of the rabbit to the left is barely noticeable due to the glare. The base experiment classified the image as skunk while experiment 2 correctly classified the image with high confidence. To the right is a squirrel nearly out of the image yet again experiment 2 was able to classify.



Conclusion and Future Research:

- Each of these cameras receive different inputs so processing them as a group hurts accuracy.
- Camera backgrounds are not similar so data pulled from these will alter confidence levels in other cameras.
- Some cameras record more images of a specific animal class than others so treating the cameras separately can improve classification confidence for these animals.
- Splitting amongst individual cameras does hurt its ability to classify animal groups that have a low number of images.
- Future work would involve testing data augmentation for classes that comprise of small datasets as well as feature extraction from earlier layers in the CNN.

Acknowledgements:

The database used in these test were provided by Dr. Chris Halle. This project has been funded by the Norwick award and a special thanks goes out to the family.