Detecting Wildlife in Uncontrolled Outdoor Video using Convolutional Neural Networks

Connor Bowley*, Alicia Andes+, Susan Ellis-Felege+, Travis Desell*
Department of Computer Science*, Department of Biology+
University of North Dakota

Introduction

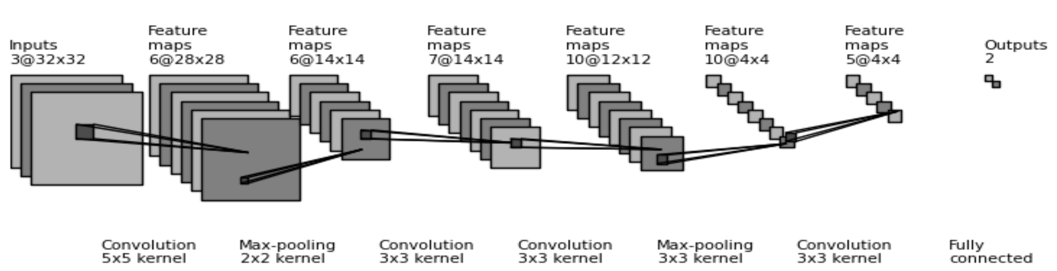
Wildlife@Home combines crowd sourcing and volunteer computing. There is over 100,000 hours of video to analyze, too much to feasibly go through manually. This is where computer vision techniques come into play.

Methods

Training Data was created by taking variable sized images and breaking them down into 32x32 sub-images by a striding technique. Careful cropping was needed to minimize misclassified training data.

An image is pulled in the top left-hand corner then a stride is taken to the right and another image is pulled. Once it reaches the end a row it will go back to the left side and move down by the stride amount.

Convolutional Neural Network (CNN) was trained to classify between two classes, tern not in frame and tern in frame.

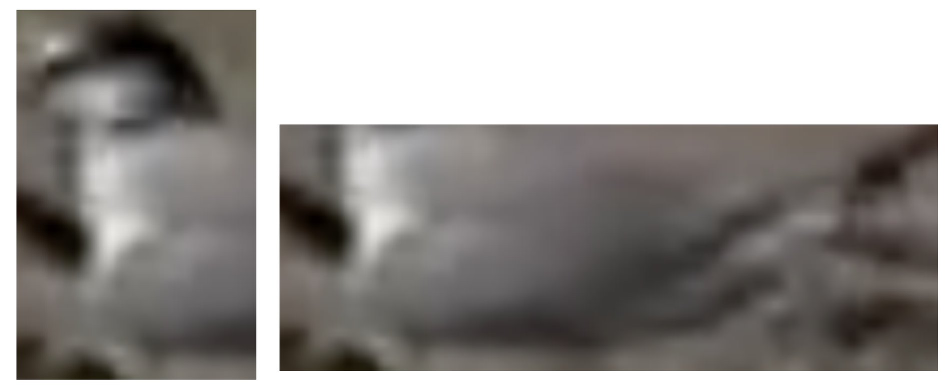


For increased runtime speed, a parallel algorithm was developed that utilized all available OpenCL devices

Computer	Devices	Time (h:mm:ss)	Seconds/Frame
Mac Pro	1 GPU	48:07	5.12
Mac Pro	CPU	32:01	3.41
Mac Pro	2 GPUs	27:34	2.93
Mac Pro	CPU and 1 GPU	20:45	2.21
Mac Pro	CPU and 2 GPUs	17:27	1.86
MacBook Pro	GPU	1:17:02	8.20
MacBook Pro	CPU	35:06	3.73
MacBook Pro	CPU and GPU	26:03	2.77

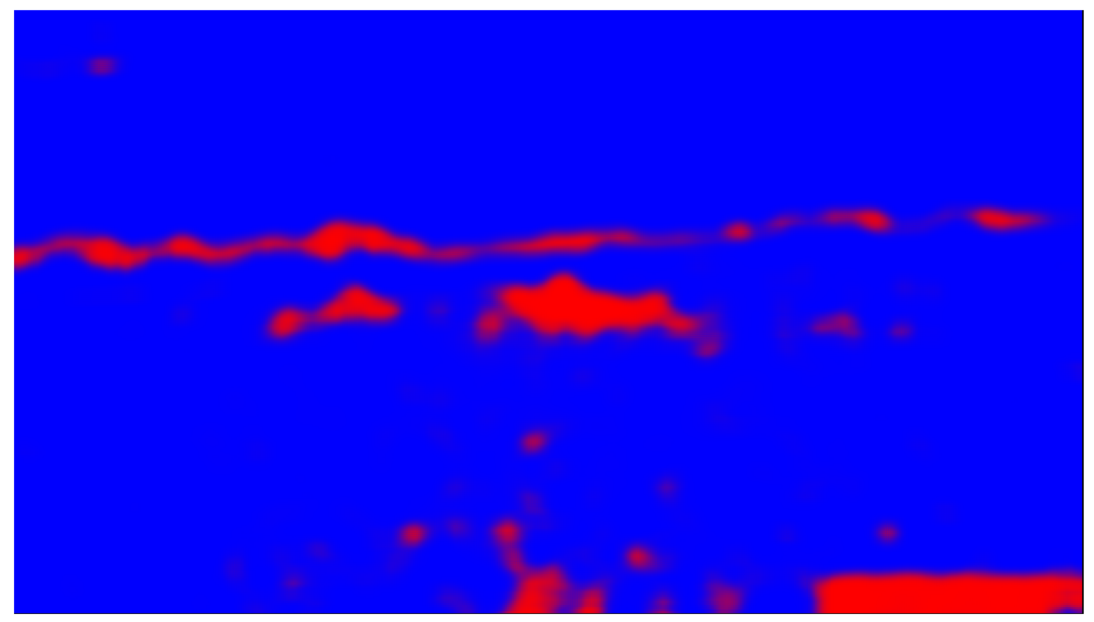
Runtime Performance Results
Results from running on 56 seconds of video (564 frames) with a stride of 15 in both directions.

An image of the positive class that contains a significant area that is of the negative class.

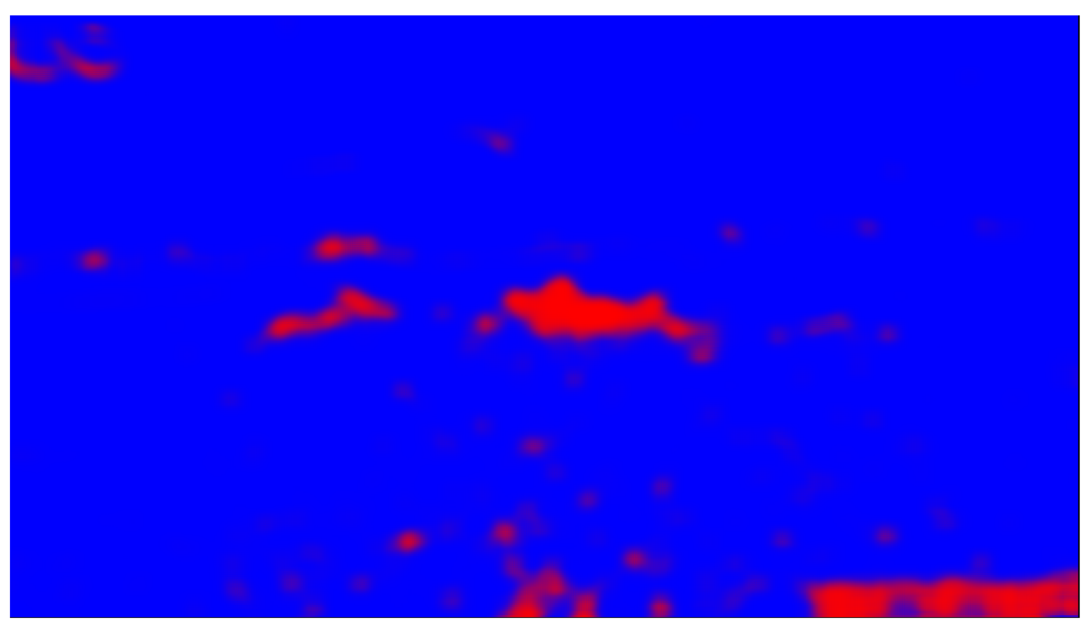


By cropping carefully we can remove the "bad" part of the image.

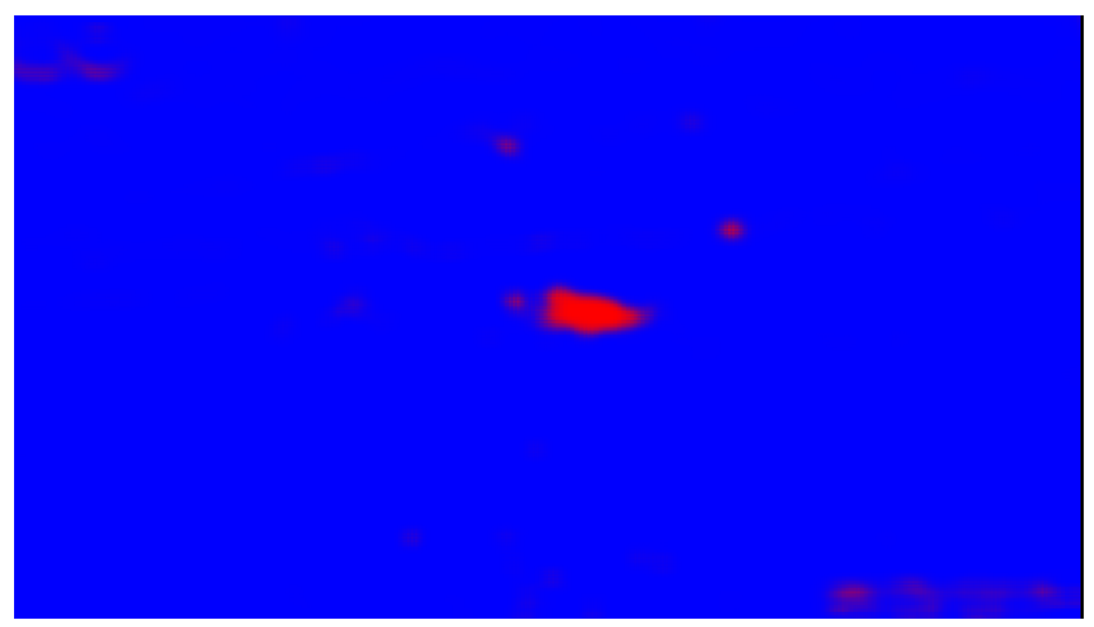
Original Image



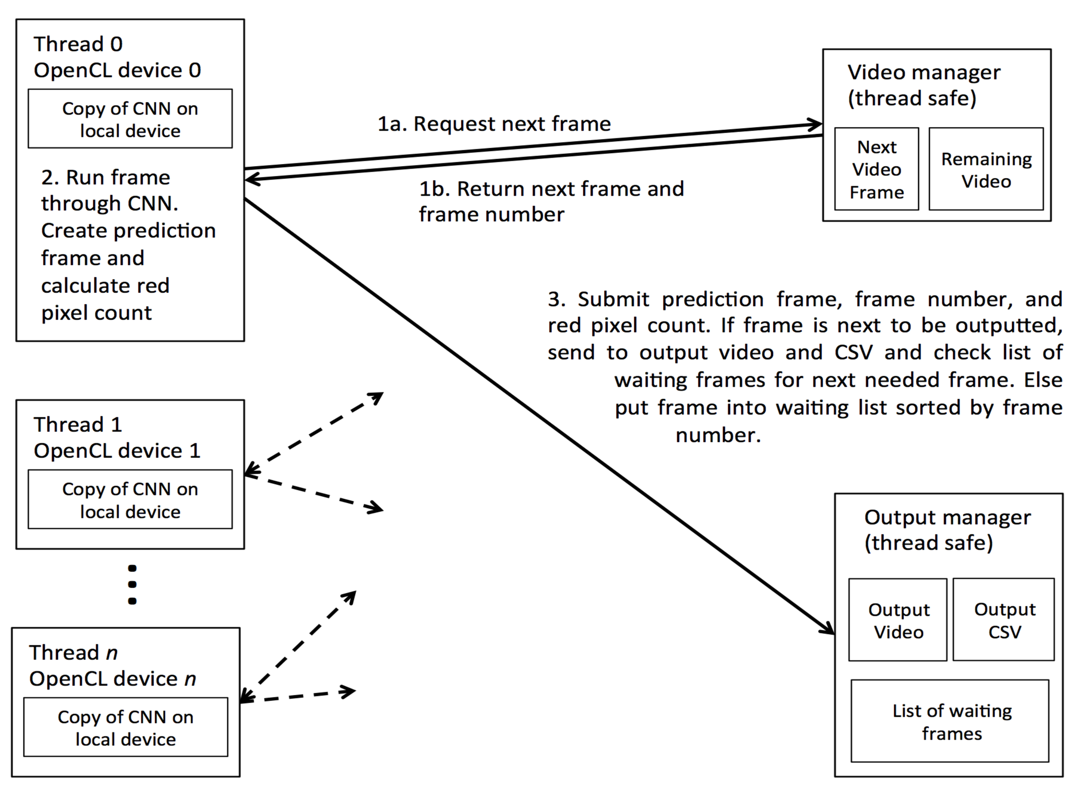
After Initial Training



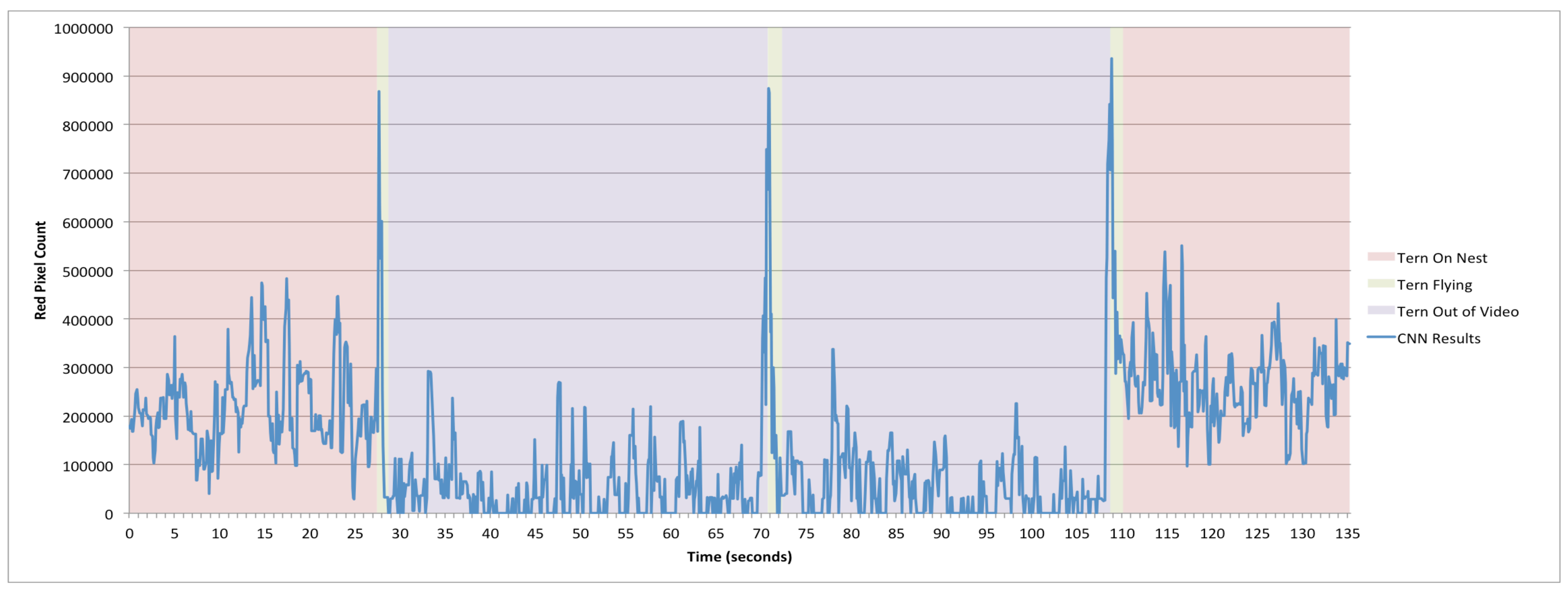
After 2 Extra Epochs



After 4 Extra Epochs



Parallel Runtime Approach



Presence of red pixels from pixel classifier

Results

After original training of the CNN we had an accuracy of

- 95% on the training data
- 82% on the test data
    - 77% of error on test set were false positives

Prediction images and videos are made by running the trained CNN over frames from the video. Initial results showed consistent misclassification of trees and ground stubble. This prompted a retraining on a specialized training set of mostly trees and ground stubble.

Acknowledgements

We appreciate the support and dedication of the Wildlife@Home citizen scientists who have spent significant amounts of time watching video. This work has been partially supported by the National Science Foundation under Grant Number 1319700. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Funds to collect data in the field were provided by the U.S. Geological Survey.