WILDLIFE ANIMAL DETECTION

A Project

Presented to the

Faculty of

California State Polytechnic University, Pomona

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

In

Computer Science

By

Stuti Jayeshkumar Trivedi

2024

SIGNATURE PAGE

PROJECT: **Wildlife Animal Detection**

AUTHOR: Stuti Jayeshkumar Trivedi

DATE SUBMITTED: Winter 2023
Computer Science

John Korah
Professor of Computer Science _____

Markus Edger
Professor of Computer Science _____

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

In 1956, the camera-trap was introduced and in 1995 using the formal mark and recapture model [1] [2] Karanth mentioned its usefulness for population ecology by identifying Panthera tigris in Nagarahole, India [3]. Camera traps, activated by heat or motion, have played a pivotal role in ecological research for over a century, undergoing a significant transformation in recent times. Such motion activated devices are helpful to observe wildlife activities without human interaction. With camera traps, the data collection is done by less disturbance to the animal. Motion triggered camera-traps are more useful due to its commercial availability and wide range of sectors aligned with it. Observing wildlife activities, animal habitat, lifestyle, and reducing over – exploitation of natural and environmental resources, camera traps have been used for decades, it is also helpful to count forest animal populations and provide valuable data for management decisions. To reduce and track wildlife poaching, animal detection helps law enforcement to detect and prevent wildlife trafficking and poaching activities. It also helps to identify and detect small and rare animals. For the research, there are a lot of references available which discuss wildlife animal monitoring, like radio tracking [4], wireless sensor network tracking, satellite and global positioning system [GPS] tracking [5] [6], and monitoring by motion sensitive camera traps. Also, it also helps confirm animal wellbeing in tourist areas of the forest and in the zoo and Animal detection in forest areas can help to see animal and environment relationships.

Object detection has been employed for identifying objects in an image or video frame for many years, where the image is not only limited to recognition of the object but also presents the boundaries with bounding boxes. Object detection detects and localizes the objects which belong to various classes. Object detection is useful in autonomous vehicles, surveillance systems, robotics, and image retrieval fields. Objects can be defined by their shape, texture, color, and other traits which are identical in the majority of cases. Once the camera-trap sensors identify the target object then it indicates its position in the given image, whereas for video sequence object detection tracks the position, size and pixel location for objects seen in various frames. Detection of an object could be seen as a classification problem, as it tells whether a specific object is present or not in the given image. The object detection pipeline is defined with three stages: (I) Region Selection the image with multiscale sliding window, which allows detection of all possible positions of objects with various sizes, (II) Feature Extraction: it is the process of transforming input image to features, which are informative and relevant to the problem statement such as, color texture, edges, which help to separate different objects and determine their relationship. These features are input for the ML and DL model to train on different patterns and relationships between objects, and (III) Classification: The ML and DL model is trained on various dataset with labeled examples. Because of variation in view point and lighting conditions it is challenging to perform object localization, so a lot more research is needed in this direction.

Animal detection and tracking is more in the center of research from the last few years as Computer Vision and Deep Learning offers more efficient techniques. Due to fast and accurate detection capacity YOLO (You Only Look Once) [7] became famous, whereas the other latest versions like YOLO [8], YOLOv3, YOLOv5[9] made sustainable enhancements to the original algorithm. YOLOv3 advanced the multi-scale prediction of classes, and improved detection performance. YOLOv5 boasts more simplified architecture that results in faster detection with improved accuracy.

include a paragraph talking about the difficulties of processing images from camera traps that may be in the thousands with triggers from wind etc.

MOTIVATION

For insights into wildlife activities, camera trap and object detection combination shows outstanding results in research where the main aim is to view the ecology and animal lifestyle, simultaneously protect the animal and wildlife resources from the illegal activities such as habitat destruction, illegal fishing, and illegal trade of endangered species. To reduce such activities and enhance ecological research many object detection systems were introduced. Traditionally Non-Neural network-based approaches were used for object detection in edge computing and Iot devices which rely on hand-engineered features and heuristic to detect objects, on the other side Neural network based approaches can automatically train itself for complex patterns and various features.

The current model easily identifies the large animals, human beings, vehicles and also birds, however a distant bird which is captured by the camera is not identified by the detection model. Moreover, sometimes the captured image color contrast makes it harder to identify the animal/bird whose color is the same as background or foreground. The small objects have ambiguous boundaries and low resolution, hence small object identification is challenging. Current models are based on high-level CNN features, which fail to capture fine grained object description due to increasing depth of network, hence leading to poor performance for small object detection.The current demand is to make an efficient model which can easily identify the small and distant bird. YOLO algorithms reach up to high standards to detect animals but for small animals we can increase the grid size and also utilize Non – Neural Network based MRF (Markov Random Field) technique for object detection.

RESEARCH OBJECTIVE

This research will focus on small bird detection with the help of non-neural network based and neural network-based approaches. For the traditional method, it will explore the scope of MRF technique to observe the pixel change and for the neural network-based technique it will perform a well known YOLO algorithm to the wildlife dataset and improve it for small bird detection. The objectives of this research are to:

1) Improve YOLO algorithm by changing grid size and detect the small bird.
2) Explore Change detection technique Markov Random Field (MRF) to identify pixel change over different timestamp and as per the pixel change, identify the change between two images.
3) Compare pros and cons of proposed traditional and neural network-based approaches for small bird detection.

LITERATURE SURVEY

Non-Neural Network-Based Approach for Object Detection: Histogram of Oriented Gradients (HOG) [10] and Support Vector Machines (SVM), this classical computer vision approach relies on extracting features based on gradients in image intensity. Based on the gradients in the image intensity, classical Computer Vision methods like Histogram of Oriented Gradients (HOG) and Support Vector Machines were introduced. Firstly, HOG captures the local object shape information and SVMs are responsible to classify those descriptors into classes. These methods are effective for some cases, however HOG + SVM struggles with diverse object variety. To detect small targets in infrared image sequences author Xinyu Wang introduced the gray-scale morphology for the removal of large image background regions which is working on fast top-hat transformation [11].

Neural Network-Based Approach for Object Detection are Region-Based Convolutional Neural Networks (R-CNN) [12] and Its Variants like R-CNN, Fast R-CNN [13], and Faster R-CNN [14], revolutionized object detection history. R-CNN generates a region proposal (object region) using a selective search algorithm, to extract features CNN is applied and then to classify the bounding boxes to objects, and each object requires separate training. Whereas Fast- RCNN works with a region of Interest (ROI) pooling layer, where it doesn't process a region separately, instead it passes entire image in a single forward pass through the CNN, then it generates feature vectors for a region by applying the feature maps which simplifies the training process compared to R-CNN. Instead of using selective search, Faster R-CNN works with Region Proposal Network (RPN) which is a fully convolutional network that uses the convolution feature with CNN. The end-to-end object detection without explicitly specifically defining the features, which make it best fit for the different object types and scenes.

The availability of open-source detection models offers an opportunity for individuals to independently assess and apply pre-trained tools. A notable example is MegaDetector, a system-agnostic object detection model developed by Microsoft explicitly for processing

camera trap data, the neural network-based object detection model [15] . This freely accessible model, trained on a vast dataset comprising millions of global images, is proficient in identifying three object classes in images: humans, animals, and vehicles. Consequently, it implicitly recognizes images without any objects from these classes. The automated categorization of images into these classes holds the potential for significantly faster processing compared to manual human efforts, with speed limitations primarily dictated by computer processing capabilities [16]. MegaDetector is an effective tool for accelerating data processing however it is lacking in performance.

YOLO algorithm pseudocode

1. image = readImage() # reize image into 140 * 140
2. NoOfCells = 7 #Grid = 7 => 7 * 7 = 49
3. NoOfClasses = 4 # predict = dog, cat, cow or rabbit
4. threshold = 0.7
5. step = height(image)/NoOfCells # 140 / 7 = 20 steps

   #stores the class for each of the 49 cells, each cell will have 4 values which correspond to the probability of a cell being 1 of the 4 classes

   #prediction_class_array[i,j] is a vector of size 4 which would look like [0.5 #cat, 0.3 #dog, 0.1 #wolf, 0.2, #cow]

6. prediction_class_array = new_array(size(NoOfCells,NoOfCells,NoOfClasses))

   #stores 2 bounding box suggestions for each of the 49 cells, each cell will have 2 bounding boxes, x y w h

7. predictions_bounding_box_array=new_array(size(NoOfCells,NoOfCells,NoOfCells,NoOfCells))
8. final_predictions = []
9. for (i<0; i<NoOfCells; i=i+1):
10. for (j<0; j<NoOfCells;j=j+1):
11. cell = image(i:i+step,j:j+step)
12. prediction_class_array[i,j] = class_predictor(cell)
    #we will first make a prediction on each cell as to what is the probability of it being one of cat, dog, cow, wolf #prediction_class_array[i,j] is a vector of size 4 which would look like [0.5 #cat, 0.3 #dog, 0.1 #wolf, 0.2 #cow]

    #sum(prediction_class_array[i,j]) = 1 #this gives us our prediction as to what each of the different 49 cells are #class predictor is a neural network that has 9 convolutional layers that make a final prediction

13. predictions_bounding_box_array[i,j] = bounding_box_predictor(cell)
14. best_bounding_box = [0 if predictions_bounding_box_array[i,j,0, 4] >
15. predictions_bounding_box_array[i,j,1, 4] else 1]
16. predicted_class = index_of_max_value(prediction_class_array[i,j])
17. if predictions_bounding_box_array[i,j,best_bounding_box, 4] *

18. max_value(prediction_class_array[i,j]) > threshold:
19. final_predictions.append([predictions_bounding_box_array[i,j,best_bounding_box, 0:4], predicted_class])
20. print final_predictions

METHODOLOGY

Improve YOLO algorithm:

To increases recall and reduce localization errors in YOLO, this research proposed to modify the YOLO by changing grid size more then 64 * 64 in above pseudocode. This adjustment will enhance the model's ability by accurately localizing the object. The large grid size will give more detailed information of object boundaries, which reduced the localization error and improve detection accuracy, on the contrary it may affect the processing time, however this research mainly focuses to solve small and distant bird detection.

Figure 1: Input and output of the model

Proposed MRF technique

Once the image is passed from the modified YOLO algorithm, then will take three consecutive images, and find the pixel changes over time. Below is proposed Mathematical formation of MRF technique.

Input:

1. Consecutive image$I_{t,}\ I_{t-1}, I_{t-2}$s
2. Weighting parameter λ
3. Smoothness parameter β
4. Training data for estimating P (pixel change at i | small bird present)

Step:1 Compute Pixel changes

$$\Delta I_i = I_t(i) - I_{t-2}\ (i)$$

Observe pixel changes between consecutive images, and analyze the pixel cluster.

Step:2 Data Term Calculation

$\Phi d(x_i)$ = log(P(ΔI_i |small bird present))

Estimate log (P(ΔI_i |small bird present)) using training data. This presents relatability of pixel change associated with the presence of small and distant birds. Where the logarithm is used as the energy cost. The negative sign will assign lower cost to pixels, where the formation of pixel changes happens due to small birds.

Step 3: Smoothness Term Calculation

$\Phi s(x_i, xj)$= 0, if $x_i = xj$

β, otherwise

Here, the smoothness $\Phi_s(x_i, xj)$ is spatial coherence, which present the degree of similarity between neighboring pixel , when pixels doesn't have similarity it will apply penalty β

Step 4: Energy Function

E(X) = $\Sigma_i \Phi d(x_i) + \lambda \Sigma_{(i,j)\varepsilon N} \Phi_s(x_i, xj)$

Combine the data term and smoothness term and lambda will control the trade-off between these two.

Step 5: Post – processing

Convert continuous values to binary prediction and determine which pixel belongs to the small bird, by introducing a threshold.

Once we have an idea of the group of pixels which are responsible for a bird formation, then highlight the Bounding box over that group of pixels and identify it as a region of small or distant bird. Additionally, this time change pixel comparison provides a dynamic approach which monitors the model by identifying object movements, appearances over time. This will improve the recall and localization accuracy. Nevertheless, this approach will face challenges like variations in lighting, shadows, and various environmental conditions like leaf movement. Those lead to more False Positives, to reduce false positives, perform Spatial Consistency Check over the pixel group given by the proposed MRF method. For that first, analyze the connected component and label it, then explore properties for each label, such as area and centroid and remove those regions that lead to False Positive. Finally, at the end, we will have those regions where there are high chances of presents of pixels which create small or distant birds.

EXPERIMENTAL SETUP AND DATASET

Cal Poly Pomona's Biological Science Department is actively engaged in wildlife animal observation over more than 24 regions in southern California. As part of this research project, cameras have been set up in various locations, resulting in a dataset over 90,000 images, where 70% camera-trap images are false positive.

Although this may seem trivial, establishing a confidence threshold is critical for assessing model performance, as adjustments can notably influence metrics like precision, recall, and F-score. Beyond impacting performance evaluation, altering thresholds can also shift the interpretation of model-derived results, potentially leading to disparate conclusions from the same set of data.

EVALUATION OF RESULT

The expected outcome for this research is a well trained YOLO algorithm with addition of the proposed state-of-art which can detect the small and distant bird with 10% accuracy. This result will identify it by analyzing TP (True Positive), TN(True Negative), FP(False Positive) and FN(False Negative), which will lead to increased high precision and low recall, as False Positives are costly in this research. Moreover, analyze the F1 score to balance Precision and Recall trad- of.

REFERENCE

[1] L. W. Gysel and E. M. Davis, "A simple automatic photographic unit for wildlife research," The Journal of Wildlife Management, vol. 20, no. 4, pp. 451–453, 1956.

[2] K. U. Karanth, "Estimating tiger Panthera tigris populations from camera-trap data using capture recapture models," Biological conservation, vol. 71, no. 3, pp. 333–338, 1995.

[3] S. Schneider, G. W. Taylor and S. Kremer, "Deep Learning Object Detection Methods for Ecological Camera Trap Data," 2018 15th Conference on Computer and Robot Vision (CRV), Toronto, ON, Canada, 2018, pp. 321-328, doi: 10.1109/CRV.2018.00052.

[4] G. C. White and R. A. Garrott, Analysis of wildlife radio-tracking data. Elsevier, 2012

[5] B. J. Godley, J. Blumenthal, A. Broderick, M. Coyne, M. Godfrey, L. Hawkes, and M. Witt, "Satellite tracking of sea turtles: Where have we been and where do we go next?" Endangered Species Research, vol. 4, no. 1-2, pp. 3–22, 2008.

[6] I. A. Hulbert and J. French, "The accuracy of GPS for wildlife telemetry and habitat mapping," Journal of Applied Ecology, vol. 38, no. 4, pp. 869–878, 2001

[7] R. Huang, J. Pedoeem and C. Chen, "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers," 2018 IEEE International Conference on

Big Data (Big Data), Seattle, WA, USA, 2018, pp. 2503-2510, doi: 10.1109/BigData.2018.8621865.

[8] C. Liu, Y. Tao, J. Liang, K. Li and Y. Chen, "Object Detection Based on YOLO Network," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2018, pp. 799-803, doi: 10.1109/ITOEC.2018.8740604.

[9] S. Li, Y. Li, Y. Li, M. Li and X. Xu, "YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection," in IEEE Access, vol. 9, pp. 141861

[10] N. -D. Nguyen, D. -H. Bui, and X. -T. Tran, "A Novel Hardware Architecture for Human Detection using HOGSVM Co-Optimization," in 2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Bangkok, Thailand, 2019, pp. 33-36. https://doi.org/10.1109/APCCAS47518.2019.8953123

[11] X. Wang, Jingdong Chen, Huosheng Xu and Xi Chen, "Adaptive method for infrared small target detection based on gray-scale morphology and backward cumulative histogram analysis," 2009 International Conference on Information and Automation, Zhuhai, Macau, 2009, pp. 173-177, doi: 10.1109/ICINFA.2009.5204915.

[12] S. Christin, E. Hervet, and N. Lecomte, "Going further with model verification and deep learning," Methods Ecol. Evol., vol. 12, no. 1, pp. 130–134, 2021. https://doi.org/10.1111/2041-210X.13494

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, pp. 580-587. https://doi.org/10.1109/CVPR.2014.81.

[14] R. Girshick, "Fast R-CNN," in 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448. https://doi.org/10.1109/ICCV.2015.169.

[15]S. Beery, D. Morris, and S. Yang, "Efficient pipeline for camera trap image review," ArXiv:1907.06772 [Cs], 2019. http://arxiv.org/abs/1907.06772.

[16] S. Beery, G. Wu, V. Rathod, R. Votel, and J. Huang, "Context R-CNN: Long Term Temporal Context for PerCamera Object Detection," ArXiv:1912.03538 [Cs, Eess, q-Bio], 2020. http://arxiv.org/abs/1912.03538.