Wildlife Identification using Object Detection using Computer Vision and YOLO

Prasham Mehta[1], Keval Shah[2], Rohit Raval[3], Manan Shah[4]

[1]Information Technology, Atharva College of Engineering
[2]Computer Science (A.I), NMIMS
[3]Computer Science (A.I), NMIMS
[4]Computer Science (A.I), NMIMS

Abstract

Wildlife monitoring plays a vital role in conservation efforts by providing insights into population dynamics, habitat utilization and species behavior. This research paper aims to explore the application of the YOLO (You Only Look Once) algorithm a cutting-edge object detection framework, in wildlife monitoring and conservation. The primary focus of this study is to evaluate the effectiveness of the YOLO algorithm in identifying and tracking wildlife species, analyzing their behavior, and supporting conservation efforts. By utilizing CCTV cameras placed in wildlife habitats the YOLO algorithm enables real-time detection, tracking, and classification of wildlife objects thereby providing researchers and conservationists with data. To conduct this research we employed a methodology that involved training the YOLO model using wildlife datasets[7]. This training enabled the model to recognize and classify species accurately. Furthermore, we optimized the model's performance before deploying it on CCTV camera feeds allowing for the monitoring of wildlife populations. The YOLO algorithms' efficient video frame processing capabilities ensure real-time object detection enabling access to insights about species presence, behavior patterns, and potential threats. The utilization of the YOLO algorithm has proven beneficial as it enables real-time identification of elusive species. This technological advancement plays a role, in providing information, for conservation efforts surrounding these species.

Keywords—

Object detection, Wildlife monitoring, Conservation, YOLO algorithm, Computer vision, Behavior analysis, Human-wildlife conflicts, Real-time monitoring.

INTRODUCTION

Wildlife monitoring is essential for understanding the dynamics of animal populations, assessing habitat utilization, and informing conservation strategies. Over the years, technological advancements in computer vision and object detection have provided powerful tools for automating the process of wildlife monitoring. These tools enable efficient and accurate identification, tracking, and analysis of wildlife species, contributing to conservation efforts worldwide

In recent years, the emergence of the YOLO (You Only Look Once) algorithm has revolutionized object detection in computer vision. YOLO stands out for its real-time capabilities, allowing for the rapid processing of video frames and providing immediate insights into the presence and behavior of wildlife species[8]. By leveraging this algorithm, researchers and conservationists can harness the power of object detection to monitor wildlife populations in real time, leading to timely decision-making and effective conservation strategies.

The primary objective of this research paper is to investigate and evaluate the application of the YOLO algorithm in wildlife monitoring and conservation. By deploying CCTV cameras strategically placed in wildlife habitats, the YOLO algorithm can be leveraged to analyze video feeds and detect wildlife objects efficiently. The algorithm's ability to identify and classify multiple objects simultaneously enables researchers to obtain accurate and reliable information on species presence and distribution.

The methodology employed in this study involves training the YOLO model on annotated wildlife datasets, ensuring the algorithm's ability to recognize and classify different species accurately[4]. The training process involves optimizing model parameters, selecting appropriate loss functions, and fine-tuning the network architecture. The trained model is then deployed on CCTV camera feeds, enabling continuous monitoring of wildlife populations.

One of the key advantages of the YOLO algorithm is its real-time performance, enabling immediate detection and tracking of wildlife objects as they appear in the camera's field of view[5]. This real-time capability facilitates the monitoring of fast-moving species, elusive animals, or those with limited visibility. By analyzing video frames at a high frame rate, the YOLO algorithm ensures that wildlife monitoring is not limited by temporal constraints, offering uninterrupted and comprehensive insights into species activities.

Furthermore, the YOLO algorithm facilitates the analysis of animal behavior, a crucial aspect of wildlife monitoring and conservation. By tracking the movements, interactions, and activity patterns of individual animals, researchers can gain insights into habitat preferences, social dynamics, and breeding behaviors. This behavioral analysis contributes to a deeper understanding of wildlife ecology and supports evidence-based conservation management.

Another important application of the YOLO-based wildlife monitoring system is the mitigation of human-wildlife conflicts[3]. By promptly detecting instances of wildlife encroaching into human territories, the system can generate immediate alerts, allowing for timely intervention and minimizing potential conflicts. This early warning system not only safeguards the safety of human populations but also ensures the protection of wildlife by preventing retaliatory actions.

In conclusion, the integration of the YOLO algorithm into wildlife monitoring practices holds significant promise for advancing conservation efforts. By leveraging the algorithm's real-time object detection capabilities, researchers and conservationists can monitor wildlife populations with greater accuracy, efficiency, and timeliness. The ability to analyze animal behavior and mitigate human-wildlife conflicts further highlights the practical implications and relevance of the YOLO-based wildlife monitoring system

.

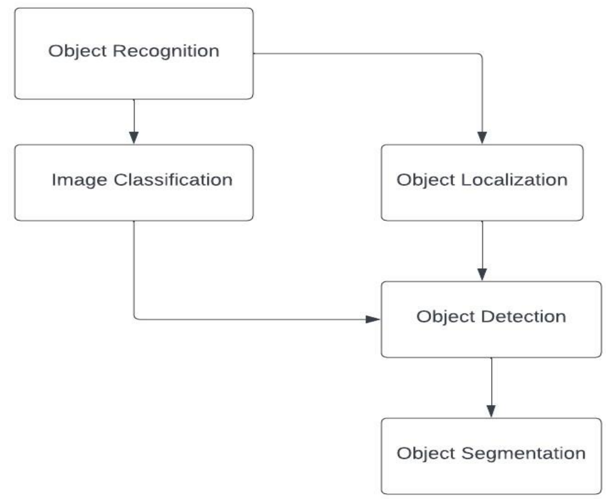


Fig 1. Image Recognition

Methodology

The importance of wildlife monitoring is of critical importance and this can be because of several reasons and aspects. It is always necessary to keep an eye on the movements of animals, whether they are safe or not also it will be helpful to many conservationists to identify threatened or endangered species and take appropriate measures to protect them. The various aspects of health and status can be understood by monitoring wildlife populations and their habitats.

We have often seen the engagement of humans in wildlife habitat which has also led to many conflicts which have lead animals to change their habitat forcefully and look for a new habitat. Through wildlife monitoring these conflicts can be controlled in a secure way and can assist in developing strategies that can mitigate these issues. Monitoring their population can help certain species which are particularly sensitive to environmental changes.

As we now know about the importance of wildlife monitoring, let's look at the importance of wildlife monitoring using object detection. Object detection basically works in two major terms as that is artificial intelligence and computer vision, without both the working of object detection is practically not possible. Through this, we can automatically identify and locate animals or specific objects of interest within images or video footage captured in wildlife environments[1]. Various wildlife researchers and conservationists can use this technology in order to keep track of animals and their habitats. This technology is also valuable for several reasons such as population monitoring, species identification, illegal poaching and logging detection, behavioral analysis, early disease detection, habitat monitoring, non- invasive monitoring, and real-time monitoring.

I. YOLO

You Only Look Once (YOLO) is a real-time object detection system that utilizes a single neural network. It boasts impressive capabilities, processing images at a rapid rate of 30 frames per second and achieving a mean average precision (mAP) of 57.9% on the COCO test-dev dataset. Unlike its predecessors, which often repurposed classifiers or localizers, YOLO directly focuses on detecting objects by prioritizing high-scoring regions within an image.

YOLO operates through a neural network that divides the input image into regions and subsequently predicts bounding boxes and probabilities for each region. Its performance surpasses classifier-based systems and even outperforms R-CNN, a method that relies on thousands of region proposals per image. However, YOLO may face challenges when detecting small objects within crowded scenes, lagging behind other state-of-the-art algorithms in such scenarios.

Undoubtedly, YOLO has had a significant impact on the field of object detection, being recognized as one of the most influential frameworks[2]. Its contributions have led to notable advancements in model efficiency and performance, as well as influencing the development of subsequent models[9]. This framework continues to enhance the overall effectiveness of object detection systems, combining speed and accuracy in a remarkable manner.

1. Dataset Description-

Every technology which operates with a large no of images forms a large no of datasets. These images can be in the format of JPEG or PNG. So when we mainly talk about object detection, it requires a large amount of dataset and every dataset has its own description. Images, bounding box coordinates, annotations, class labels, annotation formation, and data split all are the tools for the description of the dataset. In order to train a Yolo model a high-quality dataset with accurate annotations is crucial for training. Suppose we have 0 0.25 0.5 0.125 0.25 now 0.125 and 0.25 are the width and height of the bounding box belongs to class 0 (e.g., car), and its normalized coordinates are (0.25, 0.5) for the top-left corner which are relative to image dimensions.

2. Model Architecture-

YOLO consists of a single neural network and its architecture is based on 24 convolutional layers, four max-pooling layers, and two fully connected layers. The architecture work in a particular format where first before going through its convolutional network it resizes the input image into 448x448. To reduce the number of channels first the 1x1 convolution layer is applied followed by 3x3 convolutional layer to generate a cuboidal output. We know the importance of activation layer and in case for yolo it is Relu which is more optimized and efficient except for the final layer which uses linear activation function. In order to prevent the model from overfitting some additional techniques, such as batch normalization and dropout are being regularized.

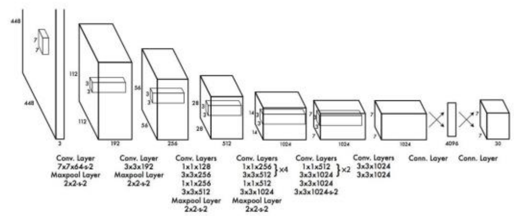


Figure 3: The Architecture. Our detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 × 1 convolutional layers reduce the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half the resolution (224 × 224 input image) and then double the resolution for detection.

3. Evaluation Metrics-

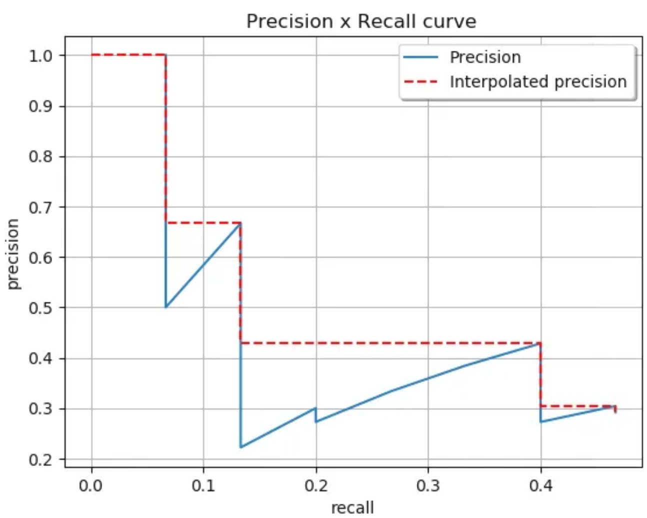
The evaluation metrics used in YOLO object detection model is the Mean Average Precision (MAP). Its value depends on several factors: IOU, Precision, Recall, Precision-Recall curve and the Average Precision. The average of all the AP values is the MAP of the model. The IOU is used to check whether the bounding boxes for each object are correctly predicted or not. We will discuss more about IOU in the further sections of this paper.

Precision is the property of the model to identify the appropriate and useful objects in the image. It is the ratio of the correct positive predictions to the total positive predictions. Recall is the property of the model to find all the real bounding boxes. It is the ratio of the correct positive predictions to the total actual positive predictions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{TP}}{\text{all detections}}$$

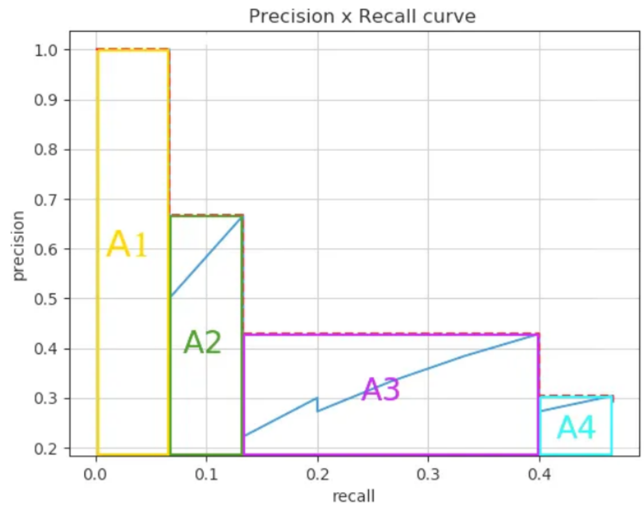
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{all ground truths}}$$

After finding all the precision and recall values, we will plot the precision-recall curve where the precision values are plotted with respect to the recall values, i.e., precision values are plotted on Y-axis and recall values on X-axis. Once the graph is plotted, a threshold value for the precision is set between 0 and 1. This value helps in eliminating the risk of over or under-detecting objects. It means that as we increase the threshold value, the risk of detecting objects more than required decreases but the risk of missing out on some objects increases. This can be interpreted as, if the threshold value is equal to 1, then none of the objects will be detected as the precision value will be 1 and the recall value will be 0. But similarly, if we take the threshold value as 0, then the precision value will also be,0 and the recall value will be 1. Hence, an infinite number of objects will be detected in this case. Hence, we need to set an appropriate threshold value that will suit our model.



As we can see in the graph above, the precision value decreases as we increase the recall value. For a machine learning model to be efficient enough, we can take a low threshold value, i.e., a high recall value as long as we can take a high precision value. For this YOLO model to be a better model, the points plotted should be in the higher right corner of the graph.

As we can see in the graph above, the line graph plotted is very mercurial in nature. Hence, we cannot easily or properly identify whether the model can be good or not. Hence to determine if the model is good enough, there is a more significant method known as Average Precision. This method involves finding the area under the curve of the graph. Hence, the more the graph is towards the top right corner, the more accurate is the model.



Here the graph is divided into 4 sections and the area under the graph is calculated for each section. This is called the average precision of each section. Finally, after finding all the average precisions, the mean average precision is calculated by taking the mean of all the values found. This value will determine the efficiency of the model used in this research.

4. Limitations and Future Directions-

No technology is a complete technology when it comes to performance and in the case of Yolo, we too have some limitations, as yolo is a single neural-based network it struggles with the detection of smaller objects which are located in crowded scenes or the objects found together and this also leads to lower accuracy. When we have a deeper architecture or a larger dataset training a Yolo can be memory intensive and it can also be very expensive. Yolo V3 which is the third version of Yolo can be the best suitable example for this. Occlusion handling is another such limitation for object detection distance and Yolo where if one object is occluded with another yolo then struggles to detect these accurately and precisely. Anchor boxes are always used by Yolo and in such cases, it uses only specific aspect ratios which in the end leads to ignorance of any other unseen aspect ratio of an image. Now with limitations the future directions are also very bright. The best part about Yolo is that it keeps on increasing or rather updating itself by upgrading the versions, the third version of Yolo is Yolo v3 and with this, we can say that all the limitations the Yolo is having can be resolved with these upgrades of Yolo. Efficient training can be a part of this future direction as Yolo in this phase sometimes struggles with the large dataset and thus it can now focus on optimizing training procedures to reduce training time and memory requirements while maintaining or improving performance. The ability of Yolo can be to understand complex scenes and improve detection accuracy can be done by including or rather incorporating multi-scale and contextual information. Real-time detection object detection is the main reason why Yolo is very practical and precise and thus in the future real-time improvements can be done when the accuracy and speed can be further increased and thus can accommodate more resource-constrained environments.

II. Object Detection

1. Real-Time Object Recognition in Videos

First, we need to understand that there is little difference between object recognition and object detection. Object recognition is a broad term under which object detection comes, i.e., Object detection is a subset of object recognition. Object detection is a process of detecting and locating objects in a real-time camera and adding a class label to the object to identify the class of the object and specifically what the object is. It is a combination of image classification and object localization. In this research paper, we have used the object recognition functionality of the YOLO algorithm in real-time CCTV cameras to identify any animal movements coming directly into villages or towns.

When a new object enters the frame of the camera, the camera captures it and tries to identify what the object is. If the object is detected as an animal, an alert is sent to the device connected to the camera. This whole procedure of animal detection includes the following three concepts using which an animal is detected: residual blocks, bounding box regression, and intersection over union(IOU). These three methodologies used collectively order-wise help in the efficient detection of animal movements [10].

1. Residual Blocks: Each frame of the live video is divided into a number of grid boxes called residual blocks. All the blocks in the grid are tantamount in shape. The grid is square in shape and has an NxN configuration. N can be any natural number depending upon how accurately you want to detect the object. An increase in N also increases the complexity of the detection process. Each and every block is used for detecting any new object entering the field of view of the live camera.

2. Bounding box regression: Each different object present in the image is outlined using a rectangular box[12]. This box is called a bounding box. The number of bounding boxes present in the image can have a maximum value of the total number of unique objects present in the image. A bounding box has a set of features that defines the object inside the bounding box such as the height and width of the object as well as the class of the object. In the bounding box regression method, a vector is formed for each bounding box which includes all these above features along with some more attributes such as the probability or confidence value of the object in the corresponding envelope grid and the coordinates of the center of the bounding box with regard to the envelope grid. An envelope grid is a collection of adjacent grids that together contain a particular object. For this application of our research, we only need a single class of objects, i.e., an animal. Whichever grids contain a probability value of zero are eliminated and the remaining grids are used for the calculation of the vector of each bounding box. These vectors are very essential for training the model on the dataset used in this research.

3. Intersections over unions (IOU): To further reduce the number of grids that are not relevant, we use the IOU method. In this method, if the object is in the grid but does not occupy a sufficient amount of grid area, that grid is also eliminated. For the determination of whether the object occupies sufficient area in the grid, a certain threshold is set. If the IOU value of the grid is greater than the threshold value then the grid is included otherwise it is eliminated. The IOU of a grid is calculated as the ratio of the intersection area to the union area of the object and the grid.

After all the grids are evaluated, we get a final picture of all the objects in the image. Now this can help in tracking the object's movement in a live video as the object has been captured using a bounding box. Hence, in this research, animals can easily be detected and their behavior can be monitored.

III. Computer Vision:-

Computer vision in artificial intelligence helps the computer to understand and compute the data such as video, images, and objects just like a human being. Computer vision involves developing algorithms through which it can extract meaningful details from visual representation and then process the data. Computer vision includes 3D analysis from 2D images. The visual data can be obtained through different sensors or either digital camera. Images can be there in different formats such as RGB or grayscale. Image processing techniques help to improve the overall quality of the image. Operations such as filtering the image, and reducing the noise help to improve the overall accuracy. There are several algorithms that can be used in computer vision such as convolutional neural networks (CNN), support vector systems (SVM), and YOLO. But for wildlife identification, we have used YOLO you only look once despite other available algorithms. The reason we have used YOLO is because Yolo can detect objects comparatively fast and accurately when compared to other algorithms available. Wild animals move fast in order to detect or track the animals Yolo can be proven more efficient and accurate. So using YOLO is a wise decision. Computer vision can be used in many places like in autonomous driving, and surveillance systems.

IV. Future Scope

There can be various improvements in wildlife protection. Firstly, various sensors can be placed at a border line so if any wild animal crosses that line it can start beeping the alarm or can notify the hub. So we can make sure that animals will not enter the human surroundings. With the help of the computer, we can analyze the movement of the animals and can predict that the animal is ill or suffering from disease or a problem. We have seen a lot of news in which it is proven that leather is made from the skin of wild animals such as crocodiles and snakes. So with the 24-hour movement of animals, we can make sure that animals are not killed and are taken care of. If a human being is in a forest where there are wild animals, with the help of computer vision it can be detected that humans are near wildlife and hence it can save life and prevent damage. Using computer vision also helps to analyze the behavior of the animals which can help to understand whether the animal is aggressive or not and can be treated accordingly.

V. Conclusion

In conclusion, the importance of wildlife monitoring cannot be overstated, given its multifaceted contributions to conservation and ecological understanding. By diligently observing animal movements and behavior, we not only ensure their safety but also provide critical data for the identification and protection of endangered species. Furthermore, the insights gained through monitoring wildlife populations and their habitats offer valuable information about ecosystem health and dynamics.

The integration of object detection technology, exemplified by the "You Only Look Once" (YOLO) system, has revolutionized real-time object detection in wildlife environments. YOLO's ability to swiftly process images while maintaining a commendable accuracy level underscores its significance in tracking animal movements and identifying objects of interest. Despite its strengths, challenges such as small object detection in crowded scenes remind us of the ongoing pursuit for improvement in detection algorithms.

As we continue to recognize the symbiotic relationship between technological innovation and wildlife conservation, it becomes evident that YOLO's influence extends beyond its immediate applications. Its impact resonates throughout the field, catalyzing advancements in model efficiency and performance, and inspiring the evolution of subsequent detection frameworks. With the ever-growing need to balance human activities and wildlife preservation, YOLO and analogous technologies stand as beacons of hope, offering the potential to bridge the gap between human progress and the protection of our natural world.

References

[1] Joseph Redmon; Santosh Divvala; Ross Girshick; Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection,"

[2] Chengji Liu; Yufan Tao; Jiawei Liang; Kai Li; Yihang Chen, "Object Detection Based on YOLO Network"

[3] Liang Tianjiao; Bao Hong, "A optimized YOLO method for object detection"

[4] Cong Tang; Yunsong Feng; Xing Yang; Chao Zheng; Yuanpu Zhou, "The Object Detection Based on Deep Learning"

[5] Reagan L. Galvez; Argel A. Bandala; Elmer P. Dadios; Ryan Rhay P. Vicerra; Jose Martin Z. Maningo, "Object Detection Using Convolutional Neural Networks"

[6] Ayushi Sharma; Jyotsna Pathak; Muskan Prakash; J N Singh, "Object Detection using OpenCV and Python"

[7] C. Chalmers; P. Fergus; S. Wich; S. N. Longmore, "Modelling Animal Biodiversity Using Acoustic Monitoring and Deep Learning"

[8] Xiaoyuan Yu, Jiangping Wang, Roland Kays, Patrick A Jansen, Tianjiang Wang & Thomas Huang, "Automated identification of animal species in camera trap images"

[9] Ivan Culjak; David Abram; Tomislav Pribanic; Hrvoje Dzapo; Mario Cifrek, "A brief introduction to OpenCV"

[10] Mauricio Marengoni; Denise Stringhini, "High Level Computer Vision Using OpenCV"

[11] M. Michalko; J. Onuška; A. Lavrín; D. Cymbalák; O. Kainz, "Tracking the object features in video based on OpenCV"

[12] Lijuan Liu; Yanping Wang; Wanle Chi, "Image Recognition Technology Based on Machine Learning"