Animal Motion Tracking in Forest: Using Machine Vision Technology

Bhaskar Borah[1], Ria Saikia[2], Piyali Das[3]

[1]B. Tech Final Year Student, [2-3]B.Sc Final Year Student

[1]Department of Electronics and Telecommunication Engineering, Assam Engineering College, India
[2-3]Department of Information Technology, The Assam Royal Global University, India

---

Abstract - The safety of wildlife is affected by various poaching done in the forests. Rhino poaching in Assam is one of the major environmental issues in India which continues in the region of Kaziranga National Park, Manas National Park, and some other grasslands of Assam. Indian rhinoceros inhabited most of the floodplain of the Indogangetic and Brahmaputra riverine tracts and the neighboring foothills. This also includes poaching of other animals like deer, tiger, etc. Despite taking preventive measures to curb the poaching of rhinos and protect all animals in the Unesco world heritage site Kaziranga National Park and Tiger Reserve (KNP and TR), the results have not shown much impact on the situation and manual observation of animal motion activity is time and cost intensive. Therefore automatic detection and monitoring of live captive animals is of major importance for assessing animal activity and, thereby, allowing for early recognition of changes indicative for threats, diseases, and animal welfare issues. To the end, this project gives a much more accurate technique to track animal motions and ensure the safety of animals. We use computer vision, being a non-invasive method for the automatic monitoring of animals. More specifically, we are using YOLO v4 (You Only Look Once version 4) model for detecting animals along with Deep SORT (Simple Online and Realtime Tracking with a Deep Association Metric) for animal motion tracking. Computer vision thereby outperforms manual and sensor-based exhaustive monitoring of the animals. This in turn can also be used for animal behavioral analysis and thus for real-time animal monitoring.

Key Words: Motion tracking, Animal detection, YOLO, Deep SORT

1. INTRODUCTION

The main objective of animal motion tracking is to prevent the poaching of animals, which in return ensures the safety of wildlife.The study has proposed that many animals have been poached in Assam, in the past few years. As per official data, 190 rhinos have been poached in Assam, since 2000. And a maximum of poaching incidents occur in Kaziranga National Park as Kaziranga National Park (NP) in Assam, India holds about 71% of the world's wild population of the greater one-horned rhino. The State Forest Department looks after the wildlife in Assam, and this paper praises their policies for animal protection along with NGOs, including the help of the local communities.

Thus, there is a need for a much more accurate technique to monitor the animals and ensure their safety of animals. To meet this criterion, one of the measures is to monitor the animal movements so as to detect any potential threat to their life.Observing, measuring, and evaluating animal behavior are important indicators to determine the safety status of animals. Moreover, humans are often not available all day for observations, therefore the time is limited, in which animals can be observed without gaps. Also, the animals may often behave differently in the presence of humans, which may also cause bias [12,15]. Thus monitoring methods that allow observing, evaluating, and evaluating the behavior of animals in the absence of humans are needed.

An automated motion tracking system for animals can be useful to continuously monitor specific or irregular events, which would ensure the safety of animals.Radio-frequency identification (RFID) technology can be used for automated animal monitoring [16,17] or specific space use but an RFID tag is implanted in an ear tag, collar, or leg band. One more method is the use of accelerometers. In [1], they used collar sensors with a 3-axis accelerometer and magnetometer for cattle. In [2], they utilized accelerometer data from leg sensors of cattle to classify activities like lying, standing, or walking. The RFID systems and sensors, like accelerometers, require certain interventions, like the implantation of an RFID chip or equipping the animal with an RFID tag or a sensor. These interventions and wearing these devices may cause stress for the animals and could have an effect on their behavior.

Thus for these reasons, analysis of animal movements using video material or images represents an effective tool to obtain information. Therefore the combination of digital video and computer vision techniques is a non-stressful, non-invasive, cost-effective, and easy

method for monitoring animal behavior that allows largely unbiased measurements and analyses of animal activities. Deep learning models, and, in particular, the use of convolutional neural networks (CNN), becomes increasingly important. In [3], used a pre-trained FasterRCNN+InceptionResNetV2 network for automated detection of European wild mammal species. Ratnayake et al. applied background subtraction together with deep learning-based detection to detect and track honeybees.

To monitor the behavior of animals automatically, we trained YOLOv4 for animal detection and combined the weights with Deep SORT for animal motion tracking.Simple Online and Realtime Tracking (SORT) is a pragmatic approach to multiple object tracking with a focus on simple, effective algorithms. It tracks objects through longer periods of occlusions, effectively reducing the number of identity switches.

The results of this system can be used to analyze the different movement patterns, which helps to differentiate between different activity levels and ensure their safety from any potential threat.

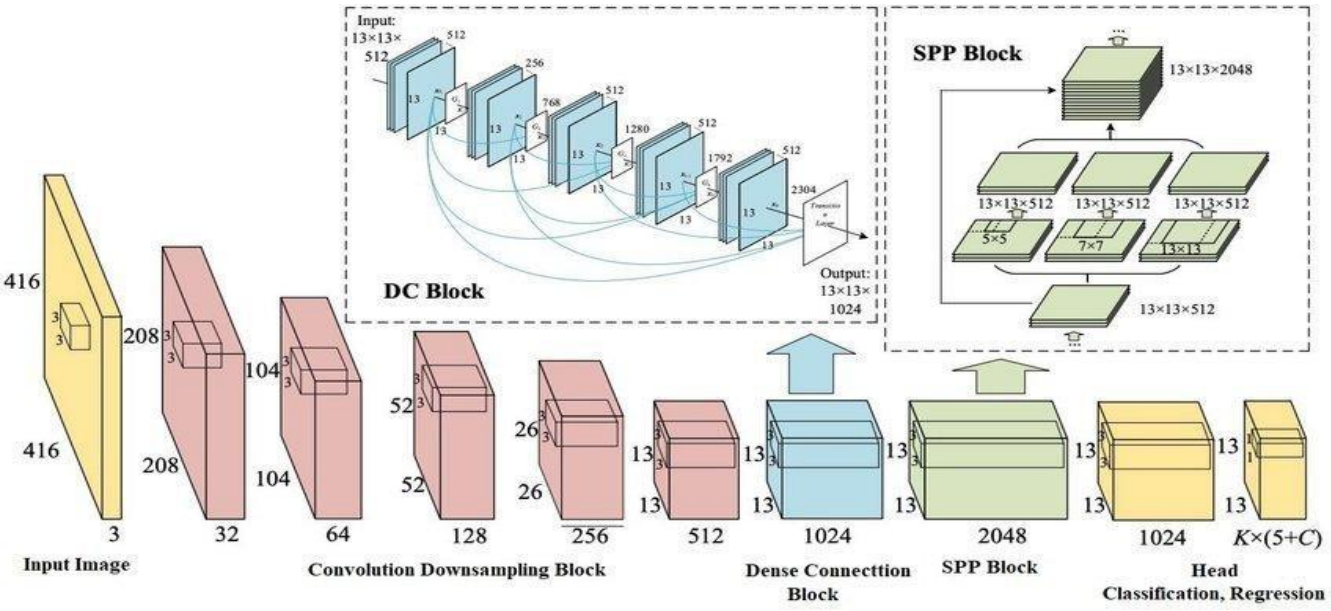

Figure 1. General Architecture of YOLO v4

2. METHODOLOGY

In this section we have explained our assessment's detailed methodology with appropriate block diagrams. The methodology consists of 5 steps: Data collection and pre-processing, yolo v4 modeling and implementation, deep sort implementation, and model validation.

2.1 Data Collection

Data collection is the procedure of collecting, measuring, and analyzing accurate insights for research using standard validated techniques. We have collected a total of 2000 images in four categories that is elephant, tiger, dog, and cat. The images are collected from various open sources available online. The images are chosen such that the model can perform well in adverse conditions. The dataset is then sent for pre-processing.

2.2 Data pre-processing

Data preprocessing and augmentation are integral parts of any computer vision system. Data augmentation is the most important and widely used regularization technique in object detection and instance segmentation. The object detection and segmentation problems are more challenging than simple image classifications because some transformations (like rotation or crop) need to be applied to the source image and the target (masks or polygon regions). The basic principles for the preprocessing step are disabling augmentation, avoiding destructive resizing, and visually inspecting the outputs. Also, before sending the pictures into the model, they must be converted into a specific format, such as binary form.

2.3 YOLOv4 Modelling and Implementation

Figure. 1 shows the general architecture of YOLOv4.You only look once (YOLO) [10] is a one-stage object detection algorithm for real-time object detection using convolutional neural networks (CNN) [4,5]. YOLOv4 consists of a 'backbone', a 'neck', and a 'head' [6]. The backbone is a CSPDarknet53, an open-source neural network framework, to train and extract features [5,6]. The neck is a path aggregation network (PAN) and spatial

pyramid pooling (SPP) used to collect feature maps from different stages [6]. The head, YOLOv3 [5], is used to implement object detection [6]. YOLOv4 is a state-of-the-art detector, which is faster and more accurate than other available detectors.The images which have been used as the input images are then used for feature extraction which is done by CSPDarknet53 which is a backbone network for YOLOv4. The backbone network then sends the extracted features of the input images to the neck of the YOLO architecture which collects all the extracted features of the input images. All these extracted features are then sent to the prediction layer which is the head of the YOLOv3 which helps to give us the required output.

We trained YOLOv4 with different configurations and observed the model's training loss and validation loss in the object detection process. We then went for model validation to check our model's efficiency and accuracy.

Table 1. YOLO v4 parameters for animal detection

Parameter	Value
Classes	4
Maxbatches	8000
Filters	27
Steps	6400,7200
Learning rate	0.001
Batch size	64

2.4 Deepsort Implementation

DeepSORT [11] is a computer vision tracking algorithm for tracking objects while assigning an ID to each object. DeepSORT is an extension of the SORT (Simple Online Realtime Tracking) algorithm. DeepSORT introduces deep learning into the SORT algorithm by adding an appearance descriptor to reduce identity switches, Hence making tracking more efficient

Simple Online Realtime Tracking (SORT)

SORT is an approach to Object tracking where rudimentary approaches like Kalman filters and Hungarian algorithms are used to track objects and claim to be better than many online trackers. SORT is made of 4 key components which are as follows:

A. Detection
This is the first step in the tracking module. In this step, an object detector detects the objects in the frame that are to be tracked. These detections are then passed on to the next step. Detectors like FrRCNN, YOLO, and more are most frequently used.

B. Estimation
In this step, we propagate the detections from the current frame to the next which is estimating the position of the target in the next frame using a constant velocity model. When detection is associated with a target, the detected bounding box is used to update the target state where the velocity components are optimally solved via the Kalman filter framework.

C. Data association
We now have the target bounding box and the detected bounding box. So, a cost matrix is computed as the intersection-over-union (IOU) distance between each detection and all predicted bounding boxes from the existing targets. The assignment is solved optimally using the Hungarian algorithm. If the IOU of detection and target is less than a certain threshold value called IOUmin then that assignment is rejected. This technique solves the occlusion problem and helps maintain the IDs.

D. Creation and Deletion of Track Identities
This module is responsible for the creation and deletion of IDs. Unique identities are created and destroyed according to the IOUmin. If the overlap of detection and target is less than IOUmin then it signifies the untracked object. Tracks are terminated if they are not detected for TLost frames, you can specify what the amount of frame should be for TLost. Should an object reappear, tracking will implicitly resume under a new identity.

The objects can be successfully tracked using SORT algorithms beating many State-of-the-art algorithms. The detector gives us detections, Kalman filters give us tracks and the Hungarian algorithm performs data association.

DeepSORT

SORT performs very well in terms of tracking precision and accuracy. But SORT returns tracks with

a high number of ID switches and fails in case of occlusion. This is because of the association matrix used. DeepSORT uses a better association metric that combines both motion and appearance descriptors. DeepSORT can be defined as the tracking algorithm which tracks objects not only based on the velocity and motion of the object but also on the appearance of the object.

For the above purposes, a well-discriminating feature embedding is trained offline just before implementing tracking. The network is trained on a large-scale person re-identification dataset making it suitable for tracking context. To train the deep association metric model in the DeepSORT cosine metric learning approach is used. According to DeepSORT's paper, "The cosine distance considers appearance information that is particularly useful to recover identities after long-term occlusions when motion is less discriminative." That means cosine distance is a metric that helps the model recover identities in case of long-term occlusion and motion estimation also fails. Using these simple things can make the tracker even more powerful and accurate. Figure 2 shows the architecture of Deep SORT.

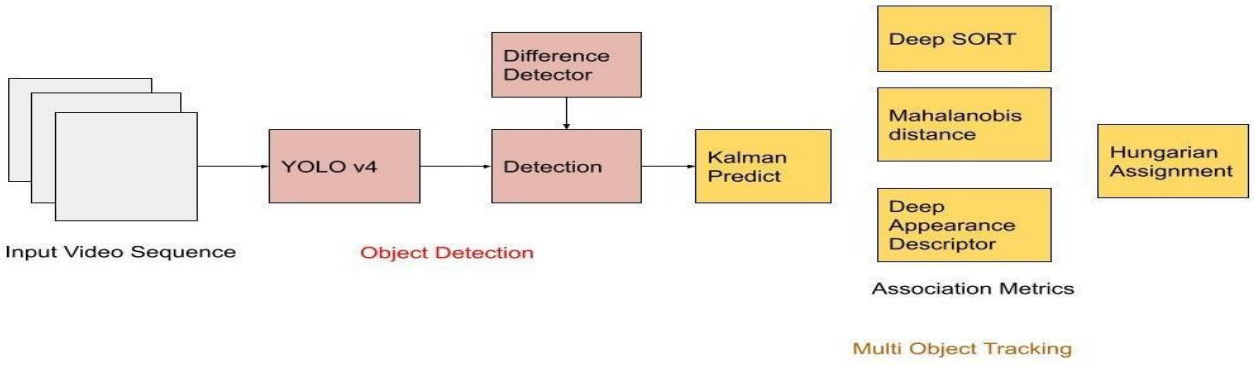

Figure 2. General architecture of Animal Motion Tracking using YOLOv4 and Deep SORT

2.5 Model validation

Evaluation always plays a big role when experimenting and testing out new things.

For the evaluation of the YOLOv4 model performance, the following values were determined: mean average precision recall, (mAP), precision (Equations (1)–(3), respectively), and detection speed.

$$\text{recall} = \frac{TP}{TP+FN} \quad (1)$$

$$\text{mAP} = \frac{\sum_{C=1}^{C} AP(c)}{C} \quad (2)$$

$$\text{precision} = \frac{TP}{TP+FP} \quad (3)$$

Where

AP = Average precision
C = number of classes
TP = number of true positive
FP = number of false positive
FN = number of false negative

The AP is determined using the interpolated average precision as described in Everingham et al. [11]:

$$AP = \frac{1}{11}\sum_{r \in \{0,0.1,\ldots.1\}} \text{Pinterp}(r) \quad (4)$$

$$\text{Pinterp}(r) = \max p(\text{rcap}).\ \text{rcap}: \text{rcap} \geq r \quad (5)$$

p(rcap) = precision at recall rcap. Equation (5) gives the desired smoothening of the precision-recall curve.

Intersection over Union (IoU) was used to determine the values TP and FP. A detection is true positive if

IoU 0.5 and false positive if IoU < 0.5. If an image is labeled and the model does not detect anything, it is a false negative.

$$IOU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (6)$$

For DeepSORT we will be judging its performance based on some standard metrics. As we know, DeepSORT is a multi-object tracking algorithm, so to judge its performance we need special metrics and benchmark datasets. We will be using CLEARMOT [8] metrics to judge the performance of our DeepSORT on the MOT17 [9] dataset.

MOT Challenge benchmark is a framework that provides a large collection of datasets with challenging real-world sequences, accurate annotations, and many metrics. MOT Challenge consists of various datasets like persons, objects, 2D, 3D, and many more. More specifically, there are several variants of the dataset released each year, such as MOT15, MOT17, and MOT20 introduced to measure the performance of multiple object trackers. MOT15, along with numerous state-of-the-art results that were submitted in the last years.MOT16, which contains new challenging videos.MOT17, which extends MOT16 sequences with more precise labels.MOT20, which contains videos from the top-down view. For our evaluation, we will be using a subset of the MOT17 dataset.

ClearMOT metrics

It is a Framework for evaluating the performance of the tracker over different parameters. A total of 8 different metrics are given to evaluate object detection, localization, and tracking performance. It also provides us with two novel metrics: (i) Multiple Object Tracking Precision (MOTP) and (ii) Multiple Object Tracking Accuracy (MOTA).

These metrics help evaluate the tracker's overall strengths and judge its general performance. Other measures are as follows:

MOTA

This measure combines three error sources: false positives, missed targets, and identity switches.

Avg Rank

This is the rank of each tracker averaged over all present evaluation measures.

MOTP

The misalignment between the annotated and the predicted bounding boxes.

IDF1

The ratio of correctly identified detections over the average number of ground-truth and computed detections.

FAF

The average number of false positives.

MT

The ratio of ground truth trajectories that are covered by a track hypothesis for at most 20% of their respective life span.

ML

The ratio of ground-truth trajectories that are covered by a track hypothesis for at most 20% of their respective life span.

FP

The total number of false positives.

FN

The total number of false negatives.

ID Sw

The total number of identity switches

Frag

The total number of times a trajectory is fragmented that is interrupted during tracking.

Hz

Processing speed on the benchmark.

Figure 3. Result samples from animal motion tracking using YOLOv4 and Deep SORT

Where

FN = number of false negatives

FP = number of false positives

IDS = number of identity switches at time t

GT = number of ground truth

For animal tracking, we will be evaluating our performance based on MOTA, which tells us about the performance of detection, misses, and ID switches. The accuracy of the tracker, MOTA (Multiple Object Tracking Accuracy) is calculated by:

$$\text{MOTA} = 1 - \frac{\sum_t FNt + FPt + IDSt}{\sum_t GTt} \quad (7)$$

3. RESULTS

3.1 Model validation assessment

Table 2 shows the YOLO v4 performance. From Table 3 we can infer that Deep SORT has performed well. The metrics show good results. DeepSORT implementation has good speed. The accuracy can be improved by using algorithms like FairMOT, and CentreTrack, which are very advanced and can reduce ID switches significantly and handle occlusions very well. It is seen that the YOLO v4 model has substantially detected animals considering the precision, recall, specificity, F1 measure, and overall accuracy. The TP, FN, FP, and TN were derived from the confusion matrix. Therefore the results obtained reveal that model has performed

well. Figure 3 shows some of the tracking results. It is seen that the model presents adequate detection and tracking of the classes.

Table 2. YOLO v4 Model Performance Evaluation

Classes	Recall [%]	Precision [%]	Average IoU [%]	mAP [%]
Dog	98.93	100	90.4	90
Cat	99.89	90	92.5	95
Elephant	93.65	95	92.7	97
Tiger	79.65	80	71.5	80
Horse	98.99	98	95	97
Giraffe	99.65	99	91.5	98

Table 3. Deep SORT Performance Evaluation

IDF1	IDP	IDR	Rcll	Prcn	FAIR	GT	MT	PT	ML	FP	FN	IDs	FM	MOTA	MOTP	MOTAL
77.8	89.2	78.6	80.1	89.4	8.90	89	9	80	78	1619	1899 4	98	190	87.4	89.8	87.6
78.6	89.5	75.7	81.4	89.6	8.58	90	9	86	86	1896	1900	95	196	89.6	80.6	88.6
77.8	89.6	78.2	85.8	90.6	9.56	89	8	86	90	1867	1899	97	197	89.5	86.6	87.6

4. CONCLUSION

This paper proposes a robust method for animal motion tracking using YOLO v4 and Deep SORT. Our system is robust to pose as the dataset contains images that are taken from different views for animal background verification. This paper gives a novel deep learning model for animal motion tracking. The model was trained by transfer learning from a pre-trained CSP-Darknet53 backbone with a COCO dataset. The model's testing was conducted using the least training and validation loss value at the 3000 epoch on the withheld 400 testing images. The testing evaluation confusion matric shows high performance, indicating its suitability for the detection of animals. Soon it will be worthwhile to deploy this animal motion tracking model on a system, which will be placed across animal corridors, and animal habitats. This system will continuously monitor animal movements 24 x 7 and will alert for any potential threat to the animals, also this system can also be used for early recognition of animal welfare issues.

ACKNOWLEDGEMENT

The authors acknowledge Dr Anupam Das, Associate Professor, Royal School of Information Technology/Engineering and Technology, The Assam Royal Global University for his valuable suggestions and advice in completing this work.

REFERENCES

[1] Dutta, R.; Smith, D.; Rawnsley, R.; Bishop-Hurley, G.; Hills, J.; Timms, G.; Henry, D. Dynamic cattle behavioural classification using supervised ensemble classifiers. Comput. Electron. Agric. 2015, 111, 18–28.

[2] Robert, B.; White, B.J.; Renter, D.G.; Larson, R.L. Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. Comput. Electron. Agric. 2009, 67, 80–84.

[3] Carl, C.; Schönfeld, F.; Profft, I.; Klamm, A.; Landgraf, D. Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. Eur. J. Wildl. Res. 2020, 66, 1–7.

[4] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection; Cornell University: Ithaca, NY, USA, 2016.

[5] Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement; Cornell University: Ithaca, NY, USA, 2018.

[6] Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection; Cornell University: Ithaca, NY, USA, 2020.

[7] Nicolai Wojke, Alex Bewley, Dietrich Paulus, Simple Online and Realtime Tracking with a Deep Association Metric; Cornell University: 2017.

[8] ClearMOT: Bernardin, K., Stiefelhagen, R. Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. J Image Video Proc 2008, 246309 (2008).

[9] ShiJie Sun, Naveed Akhtar, HuanSheng Song, Ajmal Mian, Mubarak Shah, Deep Affinity Network for Multiple Object Tracking; Cornell University:2018.

[10] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao, YOLOv4: Optimal Speed and Accuracy of Object Detection; Cornell University. 2020.

[11] Everingham, M.; van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. Int. J.Comput. Vis. 2010, 88, 303–338.

[12] White, B.J.; Coetzee, J.F.; Renter, D.G.; Babcock, A.H.; Thomson, D.U.; Andresen, D. Evaluation of two-dimensional accelerometers to monitor behavior of beef calves after castration.

[13] Hosey, G. Hediger revisited: How do zoo animals see us J. Appl. Anim. Welf. Sci. JAAWS 2013, 16, 338–359.

[14] Hemsworth, P.H.; Barnett, J.L.; Coleman, G.J. The Human-Animal Relationship in Agriculture and its Consequences for the Animal. Anim. Welf. 1993, 2, 33–51.

[15] Sorge, R.E.; Martin, L.J.; Isbester, K.A.; Sotocinal, S.G.; Rosen, S.; Tuttle, A.H.; Wieskopf, J.S.; Acland, E.L.; Dokova, A.; Kadoura, B.;et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. Nat. Methods 2014, 11, 629–632.

[16] Iserbyt, A.; Griffioen, M.; Borremans, B.; Eens, M.; Müller,W. How to quantify animal activity from radio-frequency identification (RFID) recordings. Ecol. Evol. 2018, 8, 10166–10174.

[17] Will, M.K.; Büttner, K.; Kaufholz, T.; Müller-Graf, C.; Selhorst, T.; Krieter, J. Accuracy of a real-time location system in static positions under practical conditions: Prospects to track group-housed sows. Comput. Electron. Agric. 2017, 142, 473–484.