Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

Study of Deep Learning Algorithms to Identify and Detect Endangered Species of Animals

Tanishka Badhe
Computer Engineering
Savitribai Phule Pune University
Pune, India

Bhagyashree Waghmare
Computer Engineering
Savitribai Phule Pune University
Pune, India

Janhavi Borde
Computer Engineering
Savitribai Phule Pune University
Pune, India

Vaishnavi Thakur
Computer Engineering
Savitribai Phule Pune University
Pune, India

Prof. Anagha Chaudhari
Computer Engineering
Savitribai Phule Pune University
Pune, India

***Abstract* - The world's biodiversity has been diminishing at an unparalleled rate in recent years. Many species are on the verge of extinction, and the remaining populations must be safeguarded. Animals in their native habitats can be reliably monitored. Because of their usefulness and reliability in gathering data on animals in big volumes, more efficiently, and without operator hindrance, the use of automatic hidden cameras for wildlife monitoring has skyrocketed in today's globe. However, manually analyzing and extracting information from such vast datasets gathered by camera traps can be time-consuming and tiresome. This is a significant barrier for biologists and ecologists who want to observe wildlife in a natural setting. The authors have surveyed all the recent papers related to animal recognition and identification in the wild using deep learning. By studying various papers authors have identified different algorithms and techniques for the identification of endangered animals and found out better approaches with results having higher accuracy and efficiency.**

Keywords - Endangered animal; Detection; Classification; YOLO; Convolutional neural networks (CNN); SSD; mask R-CNN; Wildlife; VGG-Net

I. INTRODUCTION

Biodiversity is the abundance and diversity of life on the earth. It is our planet's most complicated and vital feature. Biodiversity is important for both ecological and economic reasons. From genes to ecosystems, it comprises the biological, ecological, and sociocultural processes that keep life on Earth going. However, the world's biodiversity is threatened and is at high risk due to technological and economical advancements, water pollution and air pollution, and some other factors such as demographic changes. The number of threatened species is increasing at an alarming rate. As of 2019, the International Union for Conservation of Nature has classified 13,868 animal species and 14,360 plant species as endangered, up from 1,102 animal species and 1,197 plant species listed in 1996. According to the WWF's Living Planet Report 2020, global population proportions of amphibians, birds, fish mammals, and reptiles decreased by 68 percent between 1970 and 2016. Many organizations, however, are currently attempting to improve the situation.

Nowadays, technologies like machine learning (ML) and Deep Learning (DL) have the potential to teach computers to identify a wide range of items from images. Visual information, according to wildlife researchers, gives irrefutable evidence of an animal's presence by capturing image frames and recognizing them as endangered species without human intervention or support, which would aid researchers, ecologists, and scientists in classifying and preserving these animals. These animal detection systems can further help in preventing animal-vehicle collisions which result in death, injury, and also property damage.

This chapter provides a comparative analysis of different algorithms for the identification and detection of animals within their natural habitat, this would indeed help beginners in their research work.

II. BACKGROUND STUDY

[1] A paper published in 2018, employed a 1,600-image dataset from the actual world. They offer a strategy for improving species identification even in photographs collected in difficult conditions by merging standard RGB and thermal pictures. The algorithm used here was SLIC segmentation and VGGNET. The architecture they proposed is of three fully connected layers.The input to the VGG-based convNet is a 224*224 RGB picture. The preprocessing layer subtracts the RGB picture with pixel values in the range of 0–255 from the mean image values produced for the whole ImageNet training set.After preprocessing, the input photos are transmitted via these weight layers. The final completely connected layer was adjusted to meet the number of classes used in the trials, which were eight distinct species of wild animals. They chose VGGNet with 16 layers as the best version of the suggested architecture. They identified the regions of interest from the thermal pictures using the SLIC segmentation technique. These areas were sent into a neural network, which subsequently recognized a specific species in

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

the area of interest(ROI). This experiment increased the accuracy by 6 to 10% as compared to earlier R-CNN models.

In this paper [5] a deep learning-based model is developed for automatic identification, description, and counting of wild animals in camera-trap photos. They trained the DNN to detect, sum up the number of animals, and give details of the behavior of the animals using the Snapshot Serengeti dataset, which has 3.2 million photos of 48 different species. They included labels such as moving, standing, resting, interacting, feeding, and whether or not young are present in the frame in addition to counting the number of animals in the picture. They added one, two-neuron softmax output layer for each additional attribute to predict the probability of behavior in the image. The attributes were not mutually exclusive and are a multilabel classification problem that was solved by adding a two-neuron softmax output layer. They trained nine neural network models which are AlexNet, NiN, VGG, GoogLeNet, and Resnet with different numbers of layers and compared them based on accuracy. The VGG model had the best accuracy of the nine neural networks, with a score of 96.8%. Their approach eliminated 99.3 percent of manual effort in animal identification.

In October 2019, a research [7] was proposed that contrasted machine learning and deep learning methods for animal species recognition using camera trap photos. They looked at SVM, RF, and deep learning techniques AlexNet and inception v3 among other machine learning techniques. They employed the KTH dataset, which contains 19 different animal groups from which 12 classes were chosen to assess the performer of the models. The accuracy achieved for the random forest was 90.4% and for SVM for 84% whereas that for AlexNet was recorded as 94% and for Inception v3 was 96.5% which was high among all the above-considered algorithms. Hence they concluded that Deep Learning Algorithms work better than Machine Learning algorithms for animal species recognition.

In 2018, [9] put forward a framework for automatic detection of animals using camera-trapped images. It focused on identifying the species and the number of species captured in the image using Object detection algorithms such as Single Shot Multibox Detector (SSD) and You Only Look Once (YOLO). SSD (Single Shot Detector) computes a feature map by running a neural network on input images only once, whereas YOLO (You Only Look Once) is an open-source approach of object recognition that can distinguish objects in photos and videos quickly. They used a Serengeti dataset for classifying animals. For YOLO they used a pre-trained TensorFlow model and for SSD they used a VGG-16 based CNN network. The average normalized precision for YOLO was 0.55 and for SSD it was 0.67. This demonstrates that an SSD is more sensitive to the object's size. SSD was better at detecting larger things with precision, but YOLO was better at detecting smaller items. They combined both algorithms YOLO and SSD as both have their advantages and disadvantages and hence improve the detection performance. Similarly, this combined algorithm can be used for counting the animals in the image, and even though the image is crowded an accuracy of 88% is achieved.

In November 2020, proposed a model [14] on a Super-resolution Mask RCNN based transfer deep learning approach for the identification of bird species. For identifying very minute differences they used Mask RCNN which works pixel-by-pixel of the image and applies a mask on it, hence it gets simple to identify the shape and size of the object. For segmenting complex shapes of objects Mask RCNN is much better as compared to traditional algorithms. Mask RCNN improves accuracy by applying the technique of object localization. They used this technique to identify features of birds such as beaks, eyes, heads, etc and stored them as a new class. And whenever a new image was introduced they extracted the features and compared them with the stored classes. They improved the resolution of input images by using super-resolution which is a pixel independent model. For automatically applying super-resolution to images they developed software. The SRCNN with Mask RCNN and Inception V3 achieves 51.73% precision are the results obtained from their experiments. The below shows a summarization of the papers surveyed.

In 2021, A transfer learning strategy for training neural models is presented. The study [19] is aimed at recognizing bird species and safeguarding these bird species. They created a fully automated, resilient deep neural learning approach for identifying bird species from an image file, minimizing human work and saving time. Over 11,788 sounds of 200 distinct species were utilised in the collection. A pre-trained Mask RCNN is used to extract Bird ROIs from photos, that are then input into the neural network constructed using the transfer learning approach and fine-tuned using the available dataset. They employed Mask R-CNN to position birds in each picture throughout the training and inference phases. The transfer learning approach reduces the requirement for massive computer resources for processing while also speeding up the training process by reusing information. When several than just species of bird are featured in the input picture, processing gets tough. Further factors include tiny bird ROIs, poor illumination settings, resemblance in bird body components, and camouflage environments. In neural networks, such circumstances are tough to handle.

The research [2] emphasises the importance of wildlife animal monitoring in natural environments for conservation and management choices. Hidden cameras or camera traps can be used to do this. However, editing all of these photographs and movies may be time-consuming and challenging. As a result, they suggest a wildlife monitoring system that is automated. Convolutional neural networks are the algorithm employed. The South-central Victoria Wildlife Spotter dataset, which contains 125,621 single-labeled photos, was used. The photos were taken both during the day and at night, without the use of a flash. At 1920x1080 or 2048x1536 resolutions, with infrared flash in both colour and grayscale. The photographs are divided into three species groups: mammals, reptiles, and birds, with an image labelled "no animal" if there is no appearance of an animal..The suggested recognition system is made up of two CNN-based image classification models, one for each of the two tasks. The first CNN-based model is used to train a binary classifier called Wildlife detector, while the second CNN-based model

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

is used to build a multi-class classifier called Wildlife identifier. Rescaling of the shorter side of the image to a given length is done for data preparation, as is rescaling of both width and height. The pixel intensities are normalised in the [0,1] range. Techniques like zooming and shearing are employed to enhance the data. The model with the highest accuracy was VGG-16, which had a 96.6 percent accuracy, followed by RESNET-50, which had a 95.6 percent accuracy.

Using an object identification technique, the researcher [6] developed a unique animal detection and collision avoidance system. For animal detection, the suggested method uses neural network architecture such as SSD and faster R-CNN. In this paper, a new dataset with 31,774 photos is created by considering 25 groups of different animals. Then, using SSD and quicker R-CNN object detection, an animal detection model is created. The suggested and current methods are judged on their ability to achieve the criteria of mean average precision (mAP). The suggested method's mAP and detection speed are 80.5 percent at 100 frames per second and 82.11 percent at 10 frames per second for SSD and faster R-CNN, respectively.

This paper is published in 2018. [20] They use two separate datasets to track animal populations and manage ecosystems all around the world in this article. Because analysing camera trap images is costly and time-consuming, recent deep learning approaches have been utilised to overcome this. They evaluate two deep learning object detection classifiers, Faster R-CNN and YOLO v2.0, to identify, quantify, and localise animal species inside camera trap photos, using data from the Reconyx Camera Trap and the Serengeti dataset. Fast R-CNN's accuracy is 93.0 % and 76.7 %, respectively, while YOLO's accuracy is 73.0 % and 40.3 %, indicating that Faster R-CNN is more accurate than YOLO. R-CNN can correctly classify more than one species per image given insufficient data due to transfer learning.

The below table shows summary of literature study.

TABLE I. Comparison of papers studied for endangered animal classification and identification

Author	Dataset	Algorithm	Objective
[1] Mauro dos Santos de Arruda	ImageNet	SLIC and VGGNet	Combines RGB and thermal images to accurately identify animals even if images are taken in rough condition.
[5] Mohammad Sadegh Norouzzadeha	Snapshot Serengeti	AlexNet, NiN, VGG, GoogLeNet and Resnet	VGGNet has the highest accuracy for identification, counting, and description of wild animals
Rajasekaran Thangarasu [7]	KTH which has 19 different species	Inception v3	Inception v3 has the highest accuracy in animal classification.
Alexander Loos [9]	Snapshot Serengeti	Yolo and SSD	For animal detection, they combined YOLO and SSD to achieve higher precision.
Hung Nguyen [2]	Wildlife Spotter Project	CNN	Classified 3 common species from the set of animal images taken in South-central Victoria, Australia
Sazida B. Islam [4]	Camera trapped images from Texas	CNN	Detected snakes, lizards, frogs from camera trap images collected from Bastrop County, Texas
Ashvini V. Sayagavi [10]	UAV images Kuzikus Wildlife Reserve park	YOLO	Captured animal images tracked using RFID, classified and identified using YOLO.
Sofia K. Pillai [14]	Avibase: the world bird database	Mask RCNN	To identify various features of birds minutely and with high precision.

III. MAIN FOCUS OF THE CHAPTER

A. *Convolutional Neural Network (CNN)*

A convolutional neural network is a deep learning neural network used for processing structured arrays of data. The strength of a convolutional neural network comes from a layer known as the convolutional layer. Multiple convolutional layers are piled on top of each other in a CNN, each capable of recognizing increasingly complex structures. It has 3 main layers: The convolutional layer, polling layer, and fully connected layer.

1) Input Layer: The input layer does not need any parameter because it simply reads the images.

2) Convolutional Layer: The Convo layer is also known as the Feature Extractor Layer since it extracts features from the picture. The procedure yields a single integer representing the output volume.

3) Pooling Layer: This layer is used for dimension reduction.

4) Fully connected layer- "(n+1)*m" parameters.

The Fig. 1. shows the network architecture of CNN

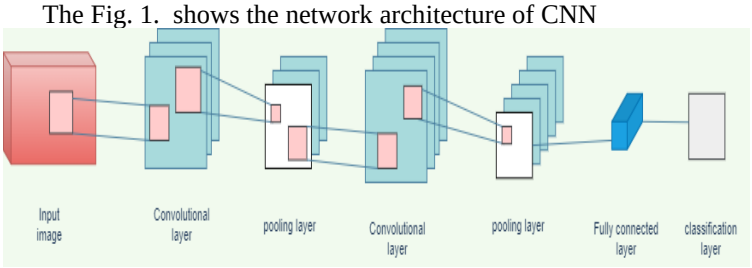

Fig. 1. Convolutional Neural Network

One of the most crucial elements of Neural Networks is the Loss Function. Loss is nothing more than a Neural Net prediction error. The Loss Function is the way for calculating the loss.

1) The mean of squared discrepancies between actual(target) and anticipated values is used to determine to Mean Squared Error (MSE).
2) The Binary Cross-Entropy (BCE)-BCE loss is used for binary classification jobs. The BCE loss function just needs one output node to categorize the data into two classes. An activation function called sigmoid should be used to process the result value, and the result range is (0-1).
3) The no. of result nodes must equal the no. of classes, according to Categorical Cross-Entropy (CC).
4) When utilizing the Sparse Categorical Cross-Entropy (SCC) loss function, the target vector does

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

not require one-hot encoding. Enter 0 if the target picture is a dog; otherwise, enter 1. Simply pass the index of the class which needs to be used.

B. You Only Look Once (YOLO)

The input image is divided into an SS grid by YOLO. The id cell for objects anticipates a set number of boundary boxes. On the other hand the object rule restricts how near it can get the objects that can be detected. YOLO has a solution for it. There are several restrictions on how close things can be. For each grid cell,

1) b boundary boxes are predicted, with a confidence score for each box.
2) By forecasting conditional class probabilities, only one item is recognized out of b boxes.

Calculate an element-wise product for each anchor box to get a probability that the box has a specific class. The confidence in anchor boxes is determined by two factors: pc and class. The likelihood of object x IoU is what it's called. IoU is an intersection over the union. It is a measure of the ground truth and anticipated bounding box overlap. Anchor boxes are filtered out in two phases.

Fig. 2. shows network design of YOLO.

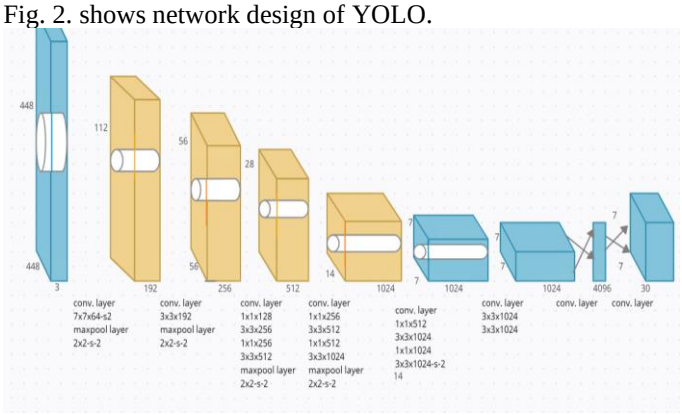

Fig. 2. You Only Look Once

The first level to filter out is to remove all those boxes whose scores are less than the threshold.

The second stage is to select the box having the maximum score and calculate its overlap with all other boxes, removing any boxes that overlap it more than the IoU threshold.

1) *Loss Function* - YOLO forecasts that each grid cell will have numerous bounding boxes. Only one of them should be in charge of the object while computing the loss for the true positive. The bounding box having the highest intersection over union is considered. As a result of this method, bounding box predictions become more specialized. Each prediction improves in its ability to forecast specific sizes and characteristics.
   - The misclassification.
   - The loss of localization.
   - The confidence loss
2) *Classification Loss* - The classification loss at each cell if an object is discovered is the squared error of the class conditional probability for each class.
3) *Localization Loss* - The mistake in anticipated anchor box locations and sizes is measured by the localization loss. Only the box responsible for verifying the object is counted.

C. Single Shot Detector (SSD)

The Single Shot Multibox Detector algorithm also localizes the object with the functioning of classifying it. The classification and localization tasks are completed in one pass, hence the name single shot. Szegedy came up with the term "multi-box" to describe this bounding box regression methodology.

SSD object detection consists of two-part:

1) A foundation concept for extracting feature maps.
2) An SSD head that detects the object using convolution filters.

Following Fig. 3. shows the architecture diagram of the SSD.

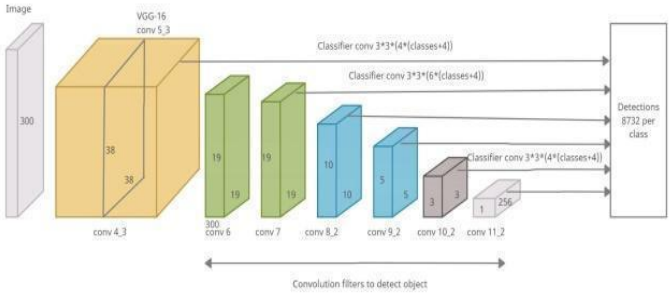

Fig. 3. Architecture Diagram of Single Shot Detector

The backbone is the VGG16 Architecture. The backbone network is a VGG-16 model that has been pre-trained on Image-Net for image categorization. Because of its strong performance in the categorization of images classification tasks, it is employed as the base network. When using the model for detection tasks, a few minor adjustments are made. The next step is to select the anchor boxes. Manual selection is needed for deciding the anchor boxes. SSD assigns a scale value to each feature map layer. The smallest scale of 0.2 is used at Convo4_3 on the leftmost side of the layers, it gradually increases to the rightmost layers up to the scale of 0.9. Aspect ratios of 1, 2, 3, 0.5, and 0.33 are good to aim for. The anchor box's width and height are calculated by multiplying the scale value by the specified aspect ratios. Following is the formulas to calculate height h and width w:

$w = \text{scale} \cdot \sqrt{aspect\ ratio}$

$h = \frac{scale}{\sqrt{aspect\ ratio}}$

Then SSD adds an extra default box with scale :

$\text{scale} = \text{scale} \cdot \sqrt{scale\ at\ next\ level}$

The two sorts of SSD predictions are positive and negative matches. If the related anchor box has an IoU greater than 0.5 with the ground truth box, the match is true. It's a no-no else. SSD only considers positive matches when calculating the cost of localization (boundary box distortion). The distinction between the anchor box and the ground truth box is the loss of localization. SSD penalizes predictions with a positive correlation. The failure to create a class forecast

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

causes a loss of confidence. Based on the confidence level of the associated class, it penalizes the loss for each positive match prediction. The total loss is the sum of the cumulative confidence loss and the localization loss.
loss_of_multibox=loss_of_confidence+alpha*loss_of_localiz ation
alpha is the weight for the loss of localization.

By removing the requirement for the region proposal network, SSD speeds up the procedure. SSD implements a number of enhancements, including anchor boxes and multi-scale features, to make up for the reduction inaccuracy. These improvements enable SSD to process images having lower resolution with good accuracy. The mAP is used to assess the accuracy of the forecasts.

D. Mask RCNN

The Mask R-CNN (regional convolutional neural network) is constructed on top of Faster R-CNN, a deep neural network framework. Faster R-CNN is a popular object detection system, and Mask R-CNN expands it with instance segmentation and other features. It can spot the difference between various objects in a picture or a video. For each object in a given image, Mask R-CNN gives the class label, object mask as well as bounding box coordinates. The system is divided into two stages: the first scans the image and creates a Region proposal network (RPN) to offer prospective item bounding boxes, and the second classifies the proposals and creates bounding boxes and masks for each class. The backbone structure is connected to both stages.

Fig. 4. shows the flow of Mask RCNN.

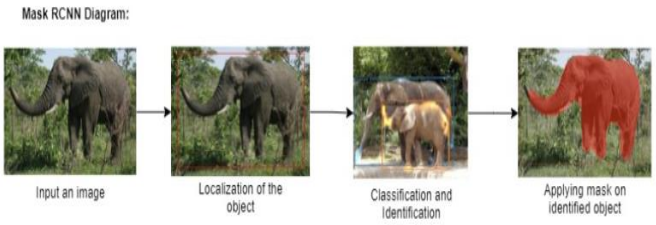

Fig. 4. Mask RCNN flow

1) *Backbone Model* - In Mask R-CNN, feature extracting is done from pictures using the ResNet 101 architecture. The next layer uses these features as an input.
2) *Region Proposal Network (RPN)* - The feature maps acquired in the preceding step are now subjected to a region proposal network (RPN). This essentially forecasts whether or not a thing is present in a particular area. At this step, the feature maps or regions that the framework forecasts would include an object.
3) *Region of Interest (RoI)* - RPN may be of varying size and shapes and uses a pooling layer to transform all of them to the same shape. The class label and anchor boxes are projected after transferring these areas across a fully linked network. It computes the region of interest in order to decrease overall the computation time. With the ground truth boxes, for each anticipated region, it calculates the Intersection over Union (IoU).

$$IoU = \frac{Area\ of\ the\ intersection}{Area\ of\ the\ Union}$$

4) Now, evaluate a region of interest if the IoU is larger than or equal to 0.5. Otherwise, overlook that section of the image.
5) *Segmentation Mask* - Having the RoIs based on the IoU values, add a mask branch to the proposed process. The segmentation mask for each object-containing region is returned by this method.

Following Fig. 5. shows the architecture of Mask RCNN.

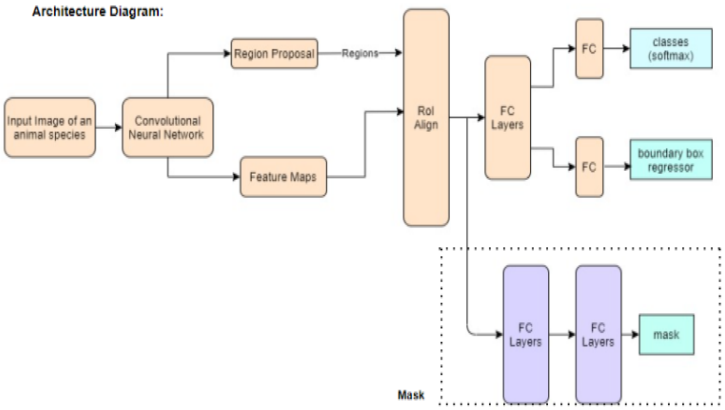

Fig. 5. Architecture of Mask RCNN

6) *Loss Function* - Mask R-CNN's multithread function of loss combines classification, localization, and segmentation mask reductions into a single function:

$$L = L_{cls} + L_{box} + L_{mask}$$

where,
L_{cls}: The loss of classification
L_{box}: The disappearance of the bounding box
L_{mask}: The mask branch creates a mask with dimensions of m × m for each RoI and class, totaling K classes. As a result, the total output is K*(m)2.

$L_{cls}$is the average binary cross-entropy loss, and for the region that corresponds to the ground truth class k, the kth mask is used.

$$L_{mask} = \frac{-1}{m^2}\sum_{1 \le i,j \le m} [y_{ij} log y_{ij}^{k} + (1 - y_{ij}) log(1 - y_{ij}^{k})]$$

The multithreaded function of loss, for losses in classification and bounding box regression, is as follows:

$$L = L_{cls} + L_{box}$$

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_{i} L_{cls}(p_i, p_i^*) + \frac{\lambda}{N_{box}}\sum_{i} p_i^* . (t_i - t_i^*)$$

where, L_{cls} can simply convert a multi-class classification into a binary categorization by forecasting whether a sample is an object to be targeted or not, is the log loss function over two classes.
L1 smooth represents a loss of L1 in a smooth manner.

$$L_{cls}(p_i, p_i^*) = -p_i^* \log p_i - (1 - p_i^*) log(1 - p_i)$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

IV. COMPARATIVE STUDY

1) *Mean Average Precision*

The area under the curve for a specific class is used to calculate Average Precision (AP). Depending on the different detection problems that exist, the mean Average Precision or mAP score is derived by taking the mean AP over all classes and/or overall IoU thresholds. The higher the MAP, the better is the model's performance. The Fig. 6. shows the comparison of mAP for various algorithms.

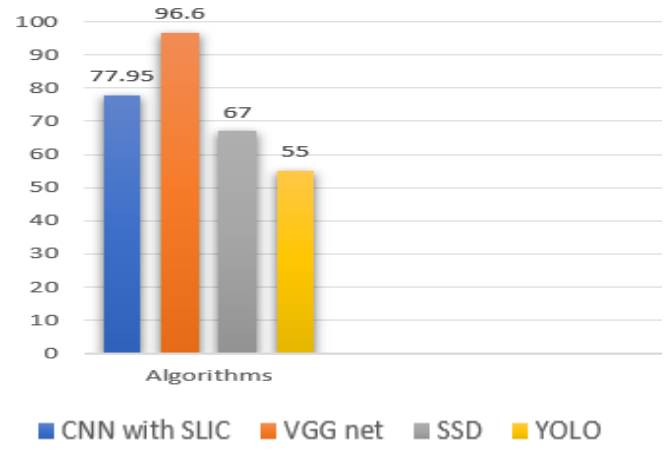

Fig. 6. Comparison of mAP

The authors studied various papers and algorithms such as CNN with SLIC segmentation which gave mean average precision (mAP) of 77.95%, VGG-Net gave mAP of 96.6 % (Hung Nguyen et al., 2017), SSD with 67% (Alexander Loos et al., 2018), and YOLO with 55 % mAP (Alexander Loos et al., 2018), and hence we can conclude that among all of these algorithms highest accuracy was achieved through VGG-Net for animal detection.

The Fig. 7. shows the dataset comparison using two parameters number of images and number of classes.

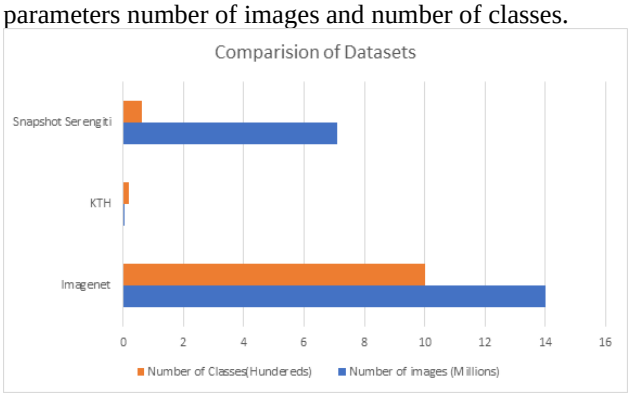

Fig. 7. Dataset Comparison

V. SOLUTION AND RECOMMENDATION

Common findings from the research are that convolutional neural networks can be used for object identification accurately, CNN is the most common algorithm used for this purpose which also gives good accuracy. A CNN is a deep neural network mapped out to analyze structured arrays of data like representations. It is a neural network that consists of multiple convolutional layers piled on top of each other each capable of detecting more complex shapes.

Depending on the multiple detecting obstacles that exist, the mean Average Precision or mAP score is calculated by averaging the AP over all classes and/or overall IoU thresholds. The mAP calculates a score by comparing the detected box to the ground-truth bounding box. The model's detection accuracy improves as the score rises. This parameter is widely used in many papers and is common among all of them.

The 2 most common datasets used were snapshot Serengeti which consists of almost 7.1 million images with 61 different categories of animals and imagenet dataset which also has a large number of images up to 14 million and 1000 classes.

VI. CONCLUSION

The authors reviewed animal detection and identification techniques and algorithms within the image. Detection of animals and extracting features from them would be of great importance to researchers for their research and in detail study of animal species. Hence, evolving technology of various ML and DL algorithms can be put into animal identification and detection. This chapter includes study of various animal detection algorithms such as CNN, YOLO, SSD, MASK R-CNN. These algorithms can be used for animal identification and localization and further these animals can be classified as endangered or not. After the survey, it is concluded that VGGNET was the best algorithm for animal classification with an accuracy of 96.6 %, similarly for object detection two algorithms namely SSD with APn of 0.67 and YOLO with APn of 0.55 are finalized, but at the end, it is concluded that a much higher APn of 0.73 is achieved by combining both these algorithms together. Mask RCNN can be explored more for animal detection and recognition as it works pixel by pixel and improves accuracy.

REFERENCES

[1] Mauro dos Santos de Arruda, Gabriel Spadon, Wesley Nunes Goncalves, & Bruno Brandoli Machado, "Recognition of Endangered Pantanal Animal Species using Deep Learning Methods," IJCNN, 2018.

[2] Hung Nguyen, Sarah J. Maclagan, Tu Dinh Nguyen, Thin Nguyen, Paul Flemons, Kylie Andrews, Euan G. Ritchie, and Dinh Phung, "Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring," Deakin University, Geelong, Australia, 2017.

[3] N. Banupriya, S. Saraya, Rashi Swaminathan, Sachinthaa Harikumar, Sukhita Palanisamya, "Animal Detection using Deep Learning Algorithm," Journal of Critical Reviews, 2019.

[4] Sazida B. Islam, Damian Valles, "Identification of Wild Species in Texas from Camera-trap Images using Deep Neural Network for Conservation Monitoring," CCWC, 2020.

[5] Mohammad Sadegh Norouzzadeha, Anh Nguyenb, Margaret Kosmalac, Alexandra Swansond, Meredith S. Palmere, Craig Packere, and Jeff Clunea, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," PNAS, 2018.

[6] Atri Saxena, Deepak Kumar Gupta, Samayveer Singh, "An Animal Detection and Collision Avoidance System Using Deep Learning," SpringerLink, 2020.

[7] Rajasekaran Thangarasu, Vishnu Kumar Kaliappan, Raguvaran Surendran, Kandasamy Sellamuthu, Jayasheelan Palanisamy, "Recognition Of Animal Species On Camera Trap Images Using Machine Learning And Deep Learning Models ," International Journal Of Scientific & Technology Research, 2019.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

[8] Zhongqi Miao, Kaitlyn M. Gaynor, Jiayun Wang, Ziwei Liu, Oliver Muellerklein, Mohammad Sadegh Norouzzadeh, Alex McInturff, Rauri C. K. Bowie, Ran Nathan, Stella X. Yu, Wayne M. Getz., et al. "Insights and approaches using deep learning to classify wildlife," Scientific Reports, 2019.

[9] Alexander Loos, Christian Weigel, Mona Koehler, "Towards Automatic Detection of Animals in Camera-Trap Images," European Signal Processing Conference (EUSIPCO), 2018.

[10] Ashvini V. Sayagavi, Sudarshan Tsb, Prasthanth C Ravoor, "Deep Learning Methods for Animal Recognition and Tracking to Detect Intrusions," ResearchGate, 2019.

[11] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," arXiv, 2016.

[12] MathieuBonneau, Jehan-AntoineVayssade, WillyTroupe, RémyArque, "Outdoor animal tracking combining neural network and time-lapse cameras," ScienceDirect, 2020.

[13] Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick,"Mask R-CNN," arXiv, 2018.

[14] Sofia K. Pillai, Dr. M. M. Raghuwanshi, Dr. P. Borkar, "Super-Resolution Mask-R CNN Based Transfer Deep Learning Approach For Identification Of Birds Species," International Journal of Advanced Research in Engineering and Technology (IJARET), 2020.

[15] Girish Pulinkala, Sai Sankar Sriram, Surya Walujkar, Pranjali Thakre, " Video Summarization Using Mask R-CNN," International Research Journal of Engineering and Technology (IRJET), 2020

[16] C.S.Arvind, R Prajwal1, Prithvi Narayana Bhat, A Sreedevi, K N Prabhudeva, "Fish Detection and Tracking in Pisciculture Environment using Deep Instance Segmentation," Sci-Hub, 2019.

[17] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, "SSD: Single Shot MultiBox Detector," arxiv, 2016.

[18] Jonathan Hui, " SSD object detection: Single Shot MultiBox Detector for real-time processing," Medium, 2018.

[19] Kamlesh Borana, Umesh More, Rajdeep Sodha, Prof. Vaishali Shirsath, "Bird Species Identifier using Convolutional Neural Network," International Journal of Engineering Research & Technology (IJERT), 2021.

[20] Stefan Schneider , Graham W. Taylor , Stefan C. Kremer, "Deep Learning Object Detection Methods for Ecological Camera Trap Data," Arxiv, 2018.

[21] Beibei Xu, Wensheng Wang, Greg Falzon, Paul Kwan, Leifeng Guo, Guipeng Chen, Amy Tait, Derek Schneider, "Automated cattle counting using Mask R-CNN in quadcopter vision system," Computers and Electronics in Agriculture, 2020.

[22] Aijiao Tan, Guoxiong Zhou, MingFang He, "Rapid Fine-Grained Classification of Butterflies Based on FCM-KM and Mask R-CNN Fusion," IEEE Access, 8, 124722–124733, 2020.

[23] Qiao, Y., Truman, M., & Sukkarieh, S., "Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming," Computers and Electronics in Agriculture, 165, 104958, 2019.