Implementation of Counting Animals In the Forest using Deep Learning Model

Arihant M Sagare
Computer Science And Engineering
Alva's Institute Of Engineering And Technology
Moodbidri,Karnataka,India
sagarearihant2611@gmail.com

Gagandeep M D
Computer Science And Engineering
Alva's Institute Of Engineering And Technology
Moodbidri,Karnataka,India
, gm901996@gmail.com

Jagath S K
Computer Science And Engineering
Alva's Institute Of Engineering And Technology
Moodbidri,Karnataka,India
jagathskarunakara@gmail.com

Kartik Tomar
Computer Science And Engineering
Alva's Institute Of Engineering And Technology
Moodbidri,Karnataka,India
kstomar945@gmail.com

Senthil Kumar R
Computer Science And Engineering
Alva's Institute Of Engineering And Technology
Moodbidri,,Karnataka,India
senthil@aiet.org.in

***Abstract*— The number of cameras monitoring animals has surpassed one million, making human examination of camera-trap data unfeasible. While automatically extracting characteristics from photos has an advantage, deep learning (DL) methods are not fully employed in automated animal action recognition (AR). Different backdrops in wildlife situations present hurdles for DL, making it difficult for models trained on entire photos to perform well on fresh test data. In order to mitigate backdrop distortion, DL object detection (OD) techniques like YOLOv5 and Pose Estimation categorize based on animal localization as opposed to entire pictures. Their effectiveness in recognising wildlife action, however, has not been thoroughly studied. Expectations that favoured Pose Estimation because of its finer control and fewer training data required were not met by YOLOv5, which did better because of its greater accuracy (0.55 vs. 0.53), six times quicker, lower processing power requirements, and user-friendly inference. With higher dataset class sizes, YOLOv5's prediction ability should increase. Despite being hampered by animals obscuring crucial points for annotation and DLC requiring the entire image as input, Pose Estimation may still have potential. It is necessary to conduct more research to evaluate Pose Estimation's effectiveness while employing better labelling techniques and bounding boxes as DLC input. Although YOLOv5 and Pose Estimation are notable improvements, their usefulness for wildlife conservation might be improved by honing them using a bigger library of photos that includes many Red deer individuals and taking behavioural changes between subsequent photographs into account.**

Keywords: convolutional neural networks; deer; animal counting; deep learning; YOLOv5; SSD

Introduction:

Conservationists find it very important to study the behaviour of wild animals [1]. Understanding animal behaviour enables one to comprehend needs, habitat, preferences, dislikes, and requirements [2]. Accurate and comprehensive understanding of animal behaviour among ecologists can contribute to the management of conservation and ecology. The basis for comprehending an animal's behaviour is action recognition (AR), which is the process of determining what an animal is doing in a picture. The usage of ethograms to identify the movements of animals has quickly altered over the last several decades. Instead, animal electrical sensors that transmit signals that can be detected by a receiver are being used. The requirement that the observer be close to the animal both before and during the observation is a limitation on both of these techniques. Electronic sensors were seen as time-consuming, costly, dangerous, and useless for tracking animal behaviour when it comes to displacement data. Above all, tagging could not be scaled to large populations [3].

An other method to extract activities from the data is to use camera-traps. Ecologists employ camera-traps more often than other types of sensors because they offer a window into the world of animals at a cheaper cost and with less effort on the part of researchers [1]. Given how quickly photos are being acquired these days, automation of augmented reality is crucial to enabling large-scale analysis. Computer vision (CV) is the automated and standardized substitute for human image processing. In CV, computers are trained to perceive their environment visually. Due to their limited ability to identify individual components of an item and the need to manually calculate the relative distances between these parts in order to determine whether something was an object or not, early CV models were limited.

Including creatures in timberland conditions is a major errand essential for biodiversity evaluation, biological exploration, and untamed life preservation endeavors. Conventional techniques for creature counting frequently include manual perception, which is tedious, work serious, and inclined to human blunder. Nonetheless, ongoing advances in profound learning, especially using Locale based Convolutional Brain Organizations (R-CNN), offer a promising road for computerizing this cycle. R-CNN,

alongside its variations like Quicker R-CNN and Veil R-CNN, has exhibited surprising outcome in object location assignments by proficiently confining and arranging objects inside pictures. With regards to backwoods conditions, where creatures might display different appearances, cover inside foliage, and shifting sizes, R-CNN presents a convincing answer for precisely counting creatures while defeating difficulties like impediments and complex foundations. By utilizing huge datasets of untamed life camera-trap pictures and applying move learning strategies, R-CNN models can figure out how to distinguish and count creatures with high precision. This paper investigates the use of R-CNN for creature including in backwoods biological systems, meaning to show its capability to upset untamed life observing practices, give significant bits of knowledge into populace elements, and add to prove based protection methodologies. Through thorough assessment and similar examination, this exploration expects to feature the viability and dependability of profound learning draws near, especially R-CNN, in propelling comprehension we might interpret woodland environment and supporting economical administration rehearses.

Deep learning (DL), a branch of machine learning (ML), which is itself a subset of artificial intelligence (AI), has allowed CV to advance quickly in recent years. Machine learning (ML) focuses on giving computers the capacity to learn without explicit programming. Multi-layered neural networks (NN), which aimed to mimic the functioning of the human brain, allowed DL to diverge from ML. Due to the neural network's ability to estimate any function in order to fit the data, these DL-models proved to be incredibly strong [4]. The inclusion of automatic feature extraction in DL algorithms eliminated the need for manual involvement in the calculation of characteristics such as the distance between the snout and the toe. This can enhance objectivity and eliminate researcher bias, which arises from presumptions about which features are significant. The network can interpret automated feature extraction by identifying if there are big or little items, where they are in the image, or how the colors contrast with one another. DL models are sometimes referred to be "black boxes," despite the fact that automatic feature extraction sacrificed interpretability and human reasoning. However, DL algorithms frequently outperformed conventional ML techniques, particularly when working with big datasets.

Although deep learning (DL) has demonstrated impressive promise in many fields, applying DL to wildlife video trap data poses particular difficulties. The erratic and uneven backgrounds seen in these photos constitute one major challenge. The danger with DL algorithms is that they may unintentionally pick up background patterns, such animal movements, along with the intended visual notions. When applied to fresh data, this propensity to concentrate on background components might result in overfitting and decreased generalization performance. Conventional deep learning classification models, which are trained on whole pictures, sometimes have difficulties in real-world situations when lighting and backgrounds differ greatly. Even while they could function well in controlled laboratory settings, when confronted with the varied and unexpected situations seen in wildlife camera-trap data, their performance usually deteriorates.

YOLOv5[6] utilized in acknowledgment is right now a famous strategy as it can rapidly and precisely distinguish and characterize objects in light of a bouncing box. On the other hand present assessment technique Deep Lab Cut (DLC) extricates central issues from creature acts and is popular like it just results the applicable properties of the creature . The point of this study was to figure out which of the two calculation turns out best for natural life AR. As a contextual analysis, 506 single Red deer (Cervus elaphus) pictures, are utilized, gathered by camera-traps dissipated across Public Park Hoge Veluwe (NPHV), a 50-km2 game hold in the Netherlands. The objective of this study is to apply and look at two driving OD techniques (YOLOv5 and the posture assessment approach (PEA)) for Red deer AR and assess how well they adjust to new information, while thinking about the degree of human exertion in making the strategies.

Literature survey:

A thorough analysis of the literature on the use of deep learning to count animals in forest habitats finds an expanding corpus of work addressing the difficulties involved in deploying video traps for wildlife monitoring. Deep learning approaches have shown promise in automating the labor-intensive operation of animal counting, which has historically required human effort and specialized knowledge. Convolutional neural networks (CNNs) and other deep learning architectures have been investigated in a number of research for the purpose of detecting and counting animals in forest environments. Large datasets of camera-trap photos taken from different forest environments are frequently used in these methods to build reliable and accurate models.

One normal test tended to in the writing is the changeability in creature appearance and posture, as well as the presence of impediments and jumbled foundations common of woods conditions. To defeat these difficulties, scientists have proposed novel organization structures, information expansion methods, and preparing procedures customized to the qualities of untamed life camera-trap information. Moreover, the writing features the significance of move learning and area transformation strategies for working on model speculation to new areas and species. By tweaking pre-prepared profound learning models on track datasets, specialists have exhibited critical enhancements in counting exactness and strength across various backwoods living spaces and creature species. Additionally, studies have explored the coordination of relevant data, like spatial connections among creatures and their environmental elements, to improve counting execution. Methods, for example, object discovery and occurrence division have been utilized to limit and count individual creatures inside complex timberland scenes precisely.

Late headways in remote detecting advancements, especially LiDAR (Light Location and Running) and hyperspectral imaging, have fundamentally upgraded the capacities of profound learning approaches for distinguishing and including creatures in timberland conditions. LiDAR innovation, which estimates distances to objects by enlightening them with laser light and dissecting the reflected heartbeats, offers exact three-layered data about the woods shade design and territory. Hyperspectral imaging, then again, catches itemized ghostly data across many thin adjoining groups, considering segregation between various sorts of vegetation and land

cover.Similarly, hyperspectral imaging supplements LiDAR information by catching fine-grained unearthly marks of vegetation and land cover, which can help with recognizing various natural surroundings and vegetation types. By integrating hyperspectral data into profound learning systems, scientists can work on the separation of creatures from foundation mess and vegetation, consequently upgrading the exactness of creature counting calculations.
These multi-modular methodologies not just influence the integral data given by various remote detecting sensors yet in addition work with a more far reaching comprehension of backwoods biological systems and untamed life elements. By incorporating LiDAR, hyperspectral imaging, and profound learning procedures, analysts can conquer the restrictions of conventional camera-trap information, like impediments, lighting varieties, and foundation mess, to accomplish more exact and effective natural life observing results
In general, the reconciliation of remote detecting advancements with profound learning approaches holds extraordinary commitment for reforming untamed life observing and protection endeavors in backwoods conditions. Via mechanizing the work serious assignment of creature counting and giving nitty gritty experiences into populace elements and territory utilization, these high level procedures add to prove based protection techniques and the reasonable administration of timberland biological systems.

I. Proposed Work

The deep learning algorithms are used to assess high-resolution pictures taken by camera traps in the proposed method for automated animal counts in forest environments. By using this method, the drawbacks of hand counting—which is labor-intensive, time-consuming, and prone to human error—are to be avoided.

Profound learning calculations, especially those intended for object location and acknowledgment assignments, offer the possibility to recognize and count creatures inside complex woods conditions precisely. These calculations are equipped for gaining complex examples and highlights from huge datasets, empowering them to identify creatures even in testing conditions like differing foundations, lighting conditions, and impediments.

uring the assessment stage, the prepared profound learning models are tried on approval datasets to survey their exhibition with regards to exactness, accuracy, review, and computational proficiency. Similar examinations with manual counting strategies can additionally approve the unwavering quality and viability of the profound learning-based approach.

When conveyed in genuine woodland conditions, the profound gaining models independently break down approaching pictures from camera traps, recognizing and counting creatures with high accuracy and productivity. This robotized cycle essentially diminishes the time and exertion expected for natural life checking, empowering protectionists and scientists to get convenient and precise information for environment the executives and biodiversity preservation drives.

Moreover, constant checking and enhancement of the sent models consider continuous upgrades in execution and flexibility to advancing natural circumstances. Partner commitment and criticism assume a vital part in refining the models and tending to explicit client necessities, guaranteeing the effective execution and long haul supportability of the mechanized creature including arrangement in woods conditions.

The flow of the proposed model is presented in Fig 1.

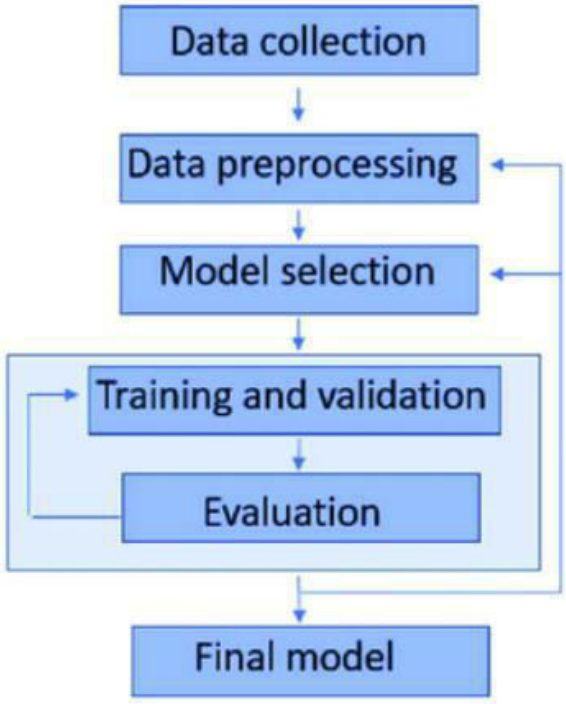

Fig 1.Flow Chart of the proposed

Model *A. Data collection*

It includes setting up camera traps in woodlands to take pictures of creatures. These snares have sensors to catch pictures when creatures cruise by. We recover these pictures consistently, it are functioning admirably to ensure the snares. Each picture is marked with subtleties like when and where it was taken. Specialists or devices distinguish the creatures in the photos and name them. We mark every creature in the pictures and coordinate the information cautiously. Once in a while, we change the pictures a piece to make our dataset greater. All through, we adhere to guidelines to consciously treat the creatures and their living spaces. This information assists us with preparing PCs to include creatures in the timberland.

B. Data Preprocessing

Information preprocessing includes preparing the picture information for investigation. We tidy up the pictures, eliminating any foggy or unimportant ones. Then, at that point, we put together the information flawlessly, adding valuable data like where and when each picture was taken. Then, we distinguish and name the creatures in the pictures, it is accurately perceived to ensure every one. Once in a

while, we change the pictures to make them overall a similar size or configuration. We likewise split the information into preparing and testing sets for the PC to gain from. At last, we actually take a look at all that to ensure it's precise and prepared for examination. In the wake of preprocessing, we might apply strategies to upgrade the pictures or equilibrium the dataset. This could include changing brilliance or difference, or creating extra manufactured information.

C. Model Selection

Model determination includes picking the best profound learning calculation for including creatures in backwoods pictures. We consider factors like model exactness, speed, and intricacy. Normal models incorporate convolutional brain organizations (CNNs) and their variations, like Consequences be damned (You Just Look Once) or Quicker R-CNN. We assess these models utilizing preparing information to see which one plays out the best. Furthermore, we might calibrate or tweak the picked model to more readily suit our particular errand and dataset. At last, we approve the chose model involving separate test information to guarantee its viability in true situations.

D. Training and Validation

Preparing and approval include showing the picked profound learning model to perceive creatures in timberland pictures and guaranteeing its precision. We feed the model with named preparing information, allowing it to learn examples and elements. During preparing, we change the model's boundaries to limit mistakes and further develop execution. Approval is finished utilizing separate information to check in the event that the model can precisely recognize creatures it hasn't seen previously. We screen measurements like exactness and misfortune to check the model's presentation. If essential, we adjust the model in view of approval results to upgrade its exactness further. At last, we guarantee moral lead all through the preparation and approval process, focusing on creature government assistance and ecological protection.

E. Evaluation

Assessment includes surveying the exhibition of the prepared profound learning model for creature including in timberland pictures. We utilize separate test information to assess how well the model performs on new, concealed pictures. Measurements like exactness, accuracy, review, and F1 score are determined to quantify the model's viability. We contrast the model's expectations and the ground truth marks to recognize any inconsistencies. Moreover, we might investigate blunder examples to comprehend where the model battles and how it tends to be gotten to the next level. Moral contemplations stay principal, guaranteeing that the assessment cycle regards natural life and their territories. At last, we record and report the assessment results to approve the model's dependability and illuminate future upgrades.

F. Final Model

The last model is the refined profound learning calculation picked for precisely including creatures in backwoods pictures. It goes through thorough preparation, approval, and assessment cycles to guarantee its viability. We calibrate the model in view of approval results, changing boundaries to enhance execution. Moral rules are maintained all through, focusing on untamed life government assistance and preservation. When approved, the last model is prepared for organization in true situations. It addresses the finish of endeavors to foster a solid device for untamed life checking and preservation. Progressing observing and updates might be important to keep up with the model's viability over the long haul.

II. Object Detection

Object identification is an undertaking inside PC vision, that includes distinguishing examples of items of a specific class (certain creatures, people or items in a picture). The principal way to deal with separate these classes from one another is frequently by utilization of profound learning [23]. Object discovery utilized through profound learning is chiefly through extraction highlights, which are extricated through a CNN and handled through different layers to recognize and separate applicable elements, for example, edges, surfaces and examples. This includes a few key stages, including input taking care of, feature extraction, highlight combination, and last forecasts. These means compare to various parts of the engineering, ordinarily known as info, spine, neck and head, see Fig. 2.

There are two sorts of streams in these finders: one-stage and two-stage. One-stage finders perform object recognition in a solitary step. Given an info picture, these models straightforwardly foresee the jumping box or centroid and their comparing class with no middle of the road steps. This sort of model is frequently quicker than two-stage and make it reasonable for ongoing complaint errands. The two-stage identifiers run two times, in the main stage it create proposition locales where there may be an item, in the second stage they arrange these locales and refine their jumping box arranges [24].

In this segment normal methodologies for object location will be covered. Initial two-stage District based CNNs (R-CNNs) [25]. Then, at that point, talking about one-stage models that are especially applicable to this theory including, SSD MobileNetV2, YOLOv5 [26] and FOMO MobileNetV2. These one-stage models are totally upheld by Edge Drive Studio for edge gadgets

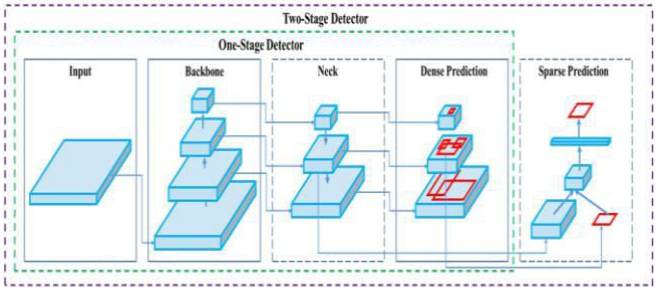

Fig 2.Object detector

• Input:
The input to the object detection system is an image or a video frame. – The image is typically represented as a grid of pixels with three color channels (RGB).

• Backbone: – The backbone network processes the input image and extracts high-level features. – It typically consists of multiple convolutional layers to capture hierarchical representations. – Popular backbone architectures include: CSPDarknet53, ResNet, and MobileNet.

• Neck: – Is commonly used in newer architecture. It fuses or combines the features ex tracted by the backbone.

• Heads: – Predicts the bounding box and class. Uses a dense predictor for one-stage and sparse predictor for two-stage models. – Dense predictors: SSD, YOLO, FOMO – Sparse predictors: R-CNN, Faster R-CNN

III. Model Training

Three particular models were prepared for the main job. The models chose involve MobileNetV2 SSD, a model recently shown to display better execution concurring than the discoveries of Olsson and Tydén [8]. Likewise, YOLOv5 and two variations of FOMO (Quicker Items, More Articles) MobileNetV2, including alpha boundaries of 0.1 and 0.35, were utilized. The alpha boundary impacts the width of the organization, with lower alpha qualities coming about in smaller models. These models were assessed in light of different measurements including F1 score, Slam use, streak memory utilization, and idleness. Furthermore, the Quicker R-CNN (Locale based Convolutional Brain Organization) model was additionally remembered for the examination for its viability in object location errands.

A. F-1 Score

The F1 score can be computed using precision and recall. They can be calculated using True Positives, False positive, True Negatives and False Negatives.

Precision = True Positives / True Positives + False Positives

Recall = True Positives / True Positives + False Negatives

F1 Score = 2 (Precision*Recall / Precision + Recall)

These formulas calculate the precision, recall, and F1 score, which are commonly used metrics for evaluating the performance of classification models.

- **Precision** estimates the model's capacity to distinguish just applicable occasions accurately. In different words, it measures the quantity of genuine up-sides out of the cases that the model anticipated to be positive (both genuine up-sides and misleading up-sides). A high accuracy indi cates a low bogus positive rate.
- **Recall** (otherwise called awareness or genuine positive rate) gauges the model's capacity to recognize every single significant occurrence, meaning it evaluates the quantity of genuine up-sides out of the real certain cases (both genuine up-sides and misleading negatives). A high review demonstrates a low bogus negative rate.
- The **F1 Score** is the consonant mean of accuracy and review. It gives a solitary score that adjusts both the worries of accuracy and review in one number. The consonant mean tends towards the more modest of the two components. Accordingly, the F1 score will be little if either accuracy or review is little. This makes it a more hearty measure than essentially taking the normal of accuracy and review, particularly in lopsided class dispersion circumstances.

IV. Implementation

A. System Overview

The proposed framework expands upon the current Ngulia framework, integrating new elements furthermore, improvements. The framework's backend is coordinated with the Edge Drive Programming interface, permitting for the preparation of new models and involving it as a C++ library. The backend runs a pipeline to incorporate firmware with the new models, OTA and general camera usefulness. The client interface incorporates a devoted tab for preparing new models and refreshing the edge gadget's running model. The edge gadget is updated with a LILYGO T-Camera from ESP32, empowering it to run further developed models and further develop camera quality. The firmware for the edge gadget can now be refreshed over the air, smoothing out the refreshing system

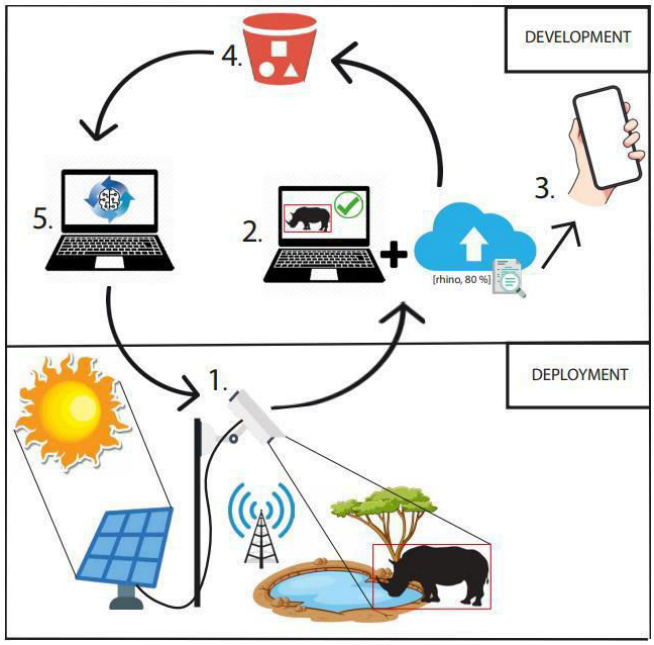

Fig 3. Workflow of the full system. 1. Picture is taken by the camera when object detection is notified. 2. Images are sent to an FTP-address together with the metadata handled by a server, ready to be annotated. 3. The metadata and images is sent park rangers to serve as information if detection is found. 4. Annotated pictures are sent to a bucket for researchers to access. 5. New models can be developed and sent to the camera to use. Images used were adapted from [33] and [34].

B. Training Pipeline

The most common way of building models was smoothed out utilizing Edge Motivation Studios, which offers critical benefits and enhancements for producing installed models. It was utilized for naming both preparation and testing datasets. The picked models were SSD MobileNetV2, FOMO MobileNetV2 and YOLOv5 [35], which all are accessible in Edge Drive Studio. To train models outside the studio, Edge Drive Programming interface were utilized, empowering preparing of new models without the need of manual studio logins. This Programming interface based approach upgrades adaptability and comfort all through the preparation stage

a) Data Collection: Pictures from past work were utilized to prepare an underlying model, and these models were then, at that point, improved with information from the Ngulia safe-haven. To gather explained information, JPG and PNG pictures with bouncing boxes in JSON design were transferred to an Amazon Web Administrations (AWS) Basic Capacity Administration (S3) pail [36]. The comment instrument utilized for the pictures can be found in Fig. 5.2. The can was then associated as an information source and connected to the Edge mpulse's Application Programming Point of interaction (Programming interface)
[35]. The information was separated into a 80/20 train/test split. The information was then trimmed and resized to guarantee all pictures were of equivalent size and aspects.

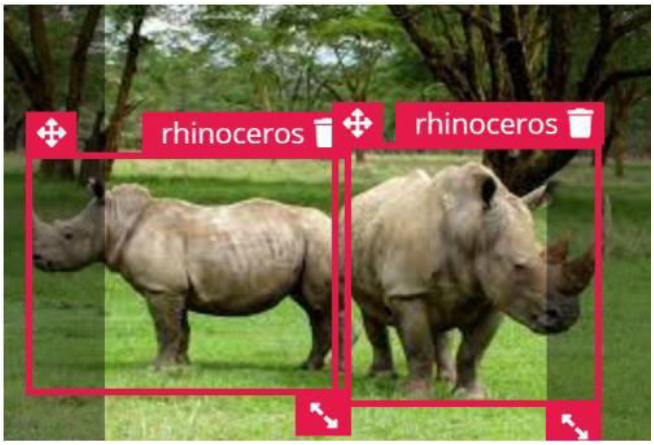

Fig 4. Annotation Tool Marking for Different Animal

b) ML - Design and Training: In this task, three unique models were utilized: MobileNetV2 SSD, FOMO MobileNetV2 0.1what's more, YOLOv5. The SSD is around 3.7MB in size and supports a RGB input at 320x320px.The Consequences be damned model is 1.8 MB in size and furthermore upholds a RGB input at 320x320px. The FOMO MobileNetV2 is intended to be <100KB in size and backing a grayscale input at any goal. The two models were used in this undertaking to survey their exhibition in object discovery errands tense gadgets, as well as to assess the relating streak memory necessities for their execution. After the models were prepared, they were approved to show a score of how the models perform.

V. Results

A. Model Performance

The models were prepared on a different dataset containing various natural life species ordinarily tracked down in woods conditions. The dataset comprised of a significant number of pictures, catching the presence of different creatures possessing the backwoods environment. Among the untamed life species remembered for the dataset were zebra, bison, rhinoceros, elephant, and numerous others usually experienced in timberland environments. With an enormous and various dataset within reach, the models were prepared to precisely identify and group these different creature species, adding to thorough natural life checking and preservation endeavors.

Animal Class	Appearances
Rhinoceros	547
Elephant	922
Zebra	241
Deer	550
Wild Boar	437

Table 1.Distribution of Training Data

Animal Class	Appearances
Rhinoceros	230
Elephant	344
Zebra	137
Deer	224
Wild Boar	216

Table 2.Distribution of Testing Data

B. Resultant Images

After the characterization of creatures utilizing profound learning models, the resultant pictures feature the identified creatures alongside jumping boxes or other visual markers featuring their areas inside the backwoods climate. These pictures give a visual portrayal of the model's presentation in precisely distinguishing and sorting different untamed life species.

Fig 5.After Successful classification of rhinoceros

By overlaying jumping boxes or names on the first pictures, the resultant pictures act as a significant instrument for untamed life observing and preservation endeavors. They

empower specialists, untamed life specialists, and moderates to picture and dissect the dispersion and conduct of creatures right at home, working with informed direction and the board methodologies. In addition, these resultant pictures add to the documentation and documentation of natural life populaces, assisting with following changes over the long run and evaluate the viability of preservation drives.

ACKNOWLEDGMENT

We would like to extend our sincere appreciation to the following individuals for their invaluable contributions to the implementation of counting animals in the forest using deep learning models, Arihant M Sagare, Karthik Tomar, Gagandeep M D, and Jagath S K, whose dedication and support were instrumental in the success of this project.

Special thanks to Senthil Kumar R for his guidance, mentorship, and expertise, which played a pivotal role in shaping our research direction and methodology. We are deeply grateful for the collaborative efforts and expertise provided by each individual involved. Additionally, we express our gratitude to Alva's Institute Of Engineering And Technology for providing the necessary resources and environment for conducting this research.

Thank you to everyone involved for their unwavering support and commitment.

REFERENCES

[1] J. Bergenas, "Project Ngulia: A national security think tank's unlikely journey," Stimson Center, Tech. Rep., 2017. [Online]. Available: http://www.jstor.org/stable/resrep10863

[2] TsavoTrust, "Ngulia rhino sanctuary – rhino viewing platform," 2023, [Accessed June 13, 2023]. [Online]. Available: https://tsavotrust.org/ngulia-rhino-sanctuary-rhino-viewing-platform/

[3] J. Linder and O. Olsson, "A smart surveillance system using edge devices for wildlife preservation in animal sanctuaries," Master of Science Thesis, Linkoping University, De partment of Electrical Engineering, June 2022.

[4] O. Wearn and P. Glover-Kapfer, "Camera-trapping for conservation: a guide to best practices from WWF," WWF, Tech. Rep., October 2017.

[5] D. L. Diefenbach, Camera Traps in Animal Ecology: Methods and Analyses. Journal of Wildlife Management, 2008, ch. A History of Camera Trapping.

[6] J. A. Ahumada, E. Fegraus, T. Birch, N. Flores, R. Kays, T. G. O'Brien, J. Palmer, S. Schut tler, J. Y. Zhao, W. Jetz, and et al., "Wildlife insights: A platform to maximize the poten tial of camera trap and other passive sensor wildlife data for the planet," Environmental Conservation, vol. 47, no. 1, 2020.

[7] M. Fennell, C. Beirne, and A. C. Burton, "Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology," Global Ecology and Conservation, vol. 35, p. e02104, 2022. [Online]. Available: https: //www.sciencedirect.com/science/article/pii/S2351989422001068

[8] A. Tyden and S. Olsson, "Edge machine learning for animal detection, classification, and tracking," Master of Science Thesis, Linkoping University, Department of Electrical Engineering, June 2020.

[9] J. Forslund and P. Arnesson, "Edge machine learning for wildlife conservation: Detec tion of poachers using camera traps," Master of Science Thesis, Linkoping University, Department of Electrical Engineering, June 2021.

[10] R. Mouha, "Internet of things (iot)," Journal of Data Analysis and Information Processing, vol. 09, pp. 77–101, 01 2021.

[11] I. A. Zualkernan, S. Dhou, J. Judas, A. R. Sajun, B. R. Gomez, L. A. Hussain, and D. Sakhnini, "Towards an iot-based deep learning architecture for camera trap image classification," in 2020 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), 2020.

[12] J. H. Davies, MSP430 microcontroller basics. Oxford: Newnes, 2013.

[13] C. R. Banbury, V. J. Reddi, M. Lam, W. Fu, A. Fazel, J. Holleman, X. Huang, R. Hur tado, D. Kanter, A. Lokhmotov, D. Patterson, D. Pau, J. sun Seo, J. Sieracki, U. Thakker, M. Verhelst, and P. Yadav, "Benchmarking TinyML systems: Challenges and direction," 2021.

14] M5Stack, "ESP32 M5Stack Timercam," M5Stack Documentation, 2023, [Accessed on 10 June 2023]. [Online]. Available: https://docs.m5stack.com/en/unit/timercam

[15] Sony, "Sony Spresence," Sony Developer World, 2023, [Accessed on 10 June 2023]. [Online]. Available: https://developer.sony.com/spresense/product-specifications#secondary-menu-desktop

[16] LILYGO, "LILYGO T-CAMERA," LILYGO Online Store, 2023. [Online]. Available: https://www.lilygo.cc/products/t-camera-s3

[17] Jean-Luc Aufranc, "LilyGO T-SIMCam ESP32-S3-CAM Development Board with 4G LTE," CNX Software, September 2022. [Online]. Available: https://www.cnx-software.com/2022/09/27/lilygo-t-simcam-esp32-s3-cam-development-board-4g-lte/

[18] M. Awad and R. Khanna, Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, 1st ed. USA: Apress, 2015.

[19] C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," Electron Markets, vol. 31, pp. 685–695, 2021. [Online]. Available: https://doi.org/10. 1007/s12525-021-00475-2

[20] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification," Procedia Computer Science, vol. 132, pp. 377–384, 2018.

[21] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, L. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," arXiv preprint arXiv:1512.07108, 2015, https://arxiv.org/pdf/1511.08458.pdf.

[22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," CVPR, 2018.

[23] Z. Diao and F. Sun, "Visual object tracking based on deep neural network," Mathematical Problems in Engineering, Jul 2022. [Online]. Available: https://doi.org/10.1155/2022/ 2154463

[24] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.

[25] P. Bharati and A. Pramanik, "Deep learning techniques—r-cnn to mask r-cnn: A survey," in Advances in Intelligent Systems and Computing, vol. 999, August 2019. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-13-9042-5_56

[26] W. Fang, L. Wang, and P. Ren, "Tinier-yolo: A real-time object detection method for constrained environments," IEEE Access, vol. 8, pp. 1935–1944, 2020.

[27] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A re view," arXiv preprint arXiv:1807.05511, 2019.

[28] Ultralytics, https://github.com/ultralytics/yolov5, 2020.

[29] D. Rivas, F. Guim, J. Polo, and D. Carrera, "An analysis of scale invariance in object detection," arXiv preprint arXiv:2111.15451, 2021.

[30] L. Moreau and M. Kelcey, "Announcing fomo (faster objects, more objects)," https:// www.edgeimpulse.com/blog/announcing-fomo-faster-objects-more-objects, 2022, ac cessed: 2023-05-30.

[31] G. Steiner, "Transfer of learning, cognitive psychology of," International Encyclopedia of the Social & Behavioral Sciences, 2001.

[32] S. T. Krishna and H. K. Kalluri, "Deep learning and transfer learning approaches for image classification," International Journal of Recent Technology and Engineering, vol. 7, no. 5S4, 2019.

[33] Freepik, "Freepik," 2023, [Accessed June 6, 2023]. [Online]. Available: https: //www.freepik.com/

[34] Pngegg, "Pngegg," 2023, [Accessed June 6, 2023]. [Online]. Available: https: //www.pngegg.com/

[35] S. Hymel, C. Banbury, D. Situnayake, A. Elium, C. Ward, M. Kelcey, M. Baaijens, M. Ma jchrzycki, J. Plunkett, D. Tischler, A. Grande, L. Moreau, D. Maslov, A. Beavis, J. Jong boom, and V. J. Reddi,

"Edge impulse: An MLOps platform for tiny machine learning," arXiv preprint arXiv:2212.03332, 2023.

[36] P. Boisrond, "A position paper on amazon web services (aws) simple storage service (s3) buckets," 08 2021.

[37] Hunter, "Hunter Solar Panel+ Manual," Hunter, June 2023.

[38] EdgeImpulse, "Example signal from rgb565 frame buffer," https://github.com/ edgeimpulse/example-signal-from-rgb565-frame-buffer, 2023, [Accessed on 10 June 2023].