Ecological Informatics 79 (2024) 102398

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

An automated approach for human-animal conflict minimisation in Assam and protection of wildlife around the Kaziranga National Park using YOLO and SENet Attention Framework

Bijuphukan Bhagabati [a,*], Kandarpa Kumar Sarma [b], Kanak Chandra Bora [c]

[a] Research Scholar, Assam Science and Technology University, Guwahati 781013, Assam, India
[b] Department of Electronics and Communication Engineering, Gauhati University, Gopinath Bordoloi Nagar, Guwahati-781014, Assam, India
[c] Dept. of Computer Science, University of Science and Technology Meghalaya, Ri-Bhoi, 793101, Meghalaya, India

ARTICLE INFO

Keywords:
Computer vision
Object detection
Animal detection
Human-animal conflict
Kaziranga
Deep learning
Yolo

ABSTRACT

Human-animal conflict in Assam, India's north-eastern state, is rising continuously. Because it occurs year-round, it damages agricultural productivity and kills people and animals, including elephants. When a herd of wild elephants emerges from a deep forest and enters human-inhabited territory around the Kaziranga National Park (KNP) in Assam, an alert must be sounded for the neighbourhood residents and forest workers to prevent conflicts. Another concern is that many wild animals die near the KNP while crossing the national highway NH-37 which traverses the area. During floods, animals flee to the highlands for food and shelter. An automated animal identification and warning system near the KNP may reduce human-animal confrontations. This paper reports the design of a system that attempts to address the above concerns. Artificial Intelligence (AI)-based strategies are utilized to recognize wild animals from live video sequences, provide warnings to avoid encounters, and protect humans and animals. Deep learning models and YoloV5 with the SENet attention layer are used to recognize wild animals in real-time. This model is trained using a public and customized dataset of animal species. Cameras attached to the cloud-based AI system take photographs from several KNP locations to confirm the model. The model's 96% accuracy in animal photographs and videos taken day and night and in feed from contemporaneous location has shown its utility. The model also improves reliability by 1–13% over previous methods.

1. Introduction

Assam, a state in northeastern India, is a biological hotspot, where human-animal conflicts, especially human-elephant standoffs, are on the rise. According to a March 2023 report by the Govt. of Assam's Minister of Forests, these conflicts claim the lives of over 70 people and 80 elephants per year in the state. The 2009 handbook of the Assam *Haathi* Project (Project, 2009) states that there are about 5000 elephants in Assam and that their numbers are increasing. Five distinct elephant reserves (ER) are located in the state, according to information on the Assam Government webpage (Forest Department Website, Government of Assam, n.d). The districts of Kokrajhar, Chirang, Baksa, and Udalguri are home to the Chirang Ripu ER (2600 sq. km.), Sonitpur ER (14,200 sq. km.), Dibrugarh and Tinsukia districts to the Dihing Patkai ER (937 sq. km.), Sonitpur, Nagaon, Golaghat, and Karbi Anglong districts to the Kaziranga-Karbi Anglong ER (3270 sq. km.), and Nagaon, Karbi-Anglong, and N.C. Hills districts to the Dhansiri-Lungding ER (2740 sq. km.). In Assam, the elephants' sole access to food, water, and shelter is via the forest. Herds of them gather together. There are now varying degrees of deforestation in various parts of Assam. Some individuals have settled there and are cultivating the forest area as well. Elephants emerge from the forest in quest of food in the nearby agricultural regions when there is a scarcity of food in the forest. This causes confrontations between people and elephants, which affects agriculture and results in human casualties. On rare occasions, incidents of elephants being murdered by humans have been reported including the use of electric shocks or poisoning in their food (The Deccan Herald, n.d).

According to The Deccan Herald on June 9, 2022 (The Deccan Herald, n.d), another issue that is causing worry in the state of Assam is the fact that speeding automobiles utilize the national highway 37 (NH-

* Corresponding author.
E-mail address: bbhagabati@gmail.com (B. Bhagabati).

https://doi.org/10.1016/j.ecoinf.2023.102398
Received 22 August 2023; Received in revised form 26 November 2023; Accepted 26 November 2023
Available online 30 November 2023

37) that traverses through restricted woods, which includes the Kaziranga National Park (KNP). There have been several wild creatures that have perished as a direct result of these vehicles. With its single-horned rhinoceroses, the park has gained a lot of attention. The protected area that is a component of the KNP has an area of about 430 km^2 and is situated on the southern bank of the Brahmaputra river. It is situated on the boundary of the Golaghat and Nagaon districts of Assam, which are two of the most ecologically gifted regions in the Eastern Himalayas. The park was listed on the list of places that were to be maintained as part of the World Heritage site by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) (UNESCO World Heritage Convention List, n.d) https://whc.unesco.org/en/list/ in the year 1985. The park is home to a vast range of wild animals, birds and creatures, including the one-horned rhinoceros, elephants, wild water buffalo, swamp deer, tigers, and monkeys, according to the information that is shown on the website of the Assam Forest Department. A Tiger Reserve designation was bestowed for the KNP in the year 2006. It was determined to be a significant place for birds by the Birdlife International, an organization that is committed to the preservation of the world's birds and other types of avifauna (Birdlife International Organization Portal, n.d). The park is separated into four distinct ranges for your enjoyment. The Kohora range is located in the centre of the park, the Agoratuli range is located in the eastern zone, the Bagori range is located in the western zone, and the Burhapahar range is located in the middle of the park, which is mostly mountainous. Several different species of animals use these regions as routes to go through. However, this well-known park is today facing a number of challenges that endanger its long-term existence. These challenges include the steady loss of marsh and grassland, encroachment, poaching, human-animal conflict, soil erosion, and the commercialization of neighbouring regions, among many other causes. As part of their discussion, Gogoi and Hira (2020) identified a number of challenges and roadblocks that must be surmounted in order to increase tourism in the KNP, including the generation of revenues. In the KNP, Medhi (2020) investigated the interactions that occurred between people and wild animals, as well as the compensation that resulted. The attitude of the local people in and around the KNP when interacting with forest authorities was shown in a compelling scenario that was described in the work mentioned above.

The entire length of the KNP is traversed by the NH-37, and various animal corridors are found surrounding the highway. Most human-wild animal encounters in the park occur along the NH-37. During the flood times in the KNP, many of the wild animals make their way out into the open area across NH-37 in search of food and safety in the park's highland areas. Most often, wild animals of this kind are struck and killed by cars and other vehicles while crossing the NH-37 on their way to the southernmost parts of the park. Particularly, during floods, vast numbers of such incidents occur. Even in the surrounding parts of the park, where elephants are known to go out into human-lived areas and agricultural areas in search of food, reports of human-elephant conflict are frequently seen. It is possible that a sophisticated automatic alert system could avoid the death of wild animals and lessen the amount of conflict between humans and elephants.

Sharma et al. (2021) have studied and reviewed work on the human-wildlife conflict problem and highlighted the causes and effects of the problem. The report reflects that traditional protection techniques like guarding and fencing are at the forefront of managing the problem. A more serious issue related to human-elephant conflicts is observed not only around the KNP, but also all over Assam, and many parts of India.

Computer vision and object detection methods are key to an automated approach to wild animal detection and alarm generation systems. Some recent works have reported the application of a class of machine learning (ML) and deep learning (DL) tools for these purposes. DL tools like the Convolutional Neural Network (CNN), due to its mammalian cortex-inspired processing, are effective in vision-based automatic discrimination and decision support for a range of applications (Yuvaraj et al., 2022; Tuia et al., 2022; Ghosh et al., 2022). As highlighted in the next section, there are proven approaches based on CNN models for controlling human-elephant conflicts and saving other wild animals' lives from fatal road accidents. To train a CNN-based model, large datasets are required. Various groups of people are maintaining large datasets of wild animals that can be used for training a specially configured CNN model to obtain better performance (Zhou et al., 2022). However, computation time and local resources required to train such approaches are high, which appears to be a major limitation of such approaches. Hence, ample opportunities exist to design and configure fast cloud resident-modified CNN-based DL models with multiple camera inputs for deployment around the critical infrastructure, including highways passing through important biological hotspots (like the NH-37 passing through the KNP).

In this work, the application of such a CNN-based model to prevent human-animal conflicts in Assam and save the lives of wild animals in the KNP, including deer and tigers, is proposed. The work centers on the YOLO model, which is a reliable, simple, and fast DL tool for training, testing, and deployment with a range of multi-dimensional data (Redmon et al., 2016). The current work highlights the effective use of the YOLOv5 network configured for wild animal detection from video sequences. To improve the model's performance, the SENet attention layer (Hu et al., 2019) is added to YOLOv5 (Zhu et al., 2021). The model is tested in a cloud-resident form with cameras deployed at four different locations, representing four different zones of the park where human-animal conflict rates are high. It is also expanded for performing a range of experiments to explore the possibilities of detecting a diverse class of animals captured with varied illumination, distance, and backgrounds with the provision of auto alarm generation. The main contribution of this paper is the enhancement of the detection accuracy using the YOLOv5 model with the SENet attention mechanism with the following advantages.

- This mechanism can enhance detection accuracy using fine-grain features of the object, enabling the detection of a diverse class of animals captured with varied illumination, distance, and background variations with the provision of auto alarm generation.
- The model gives an enhanced detection rate by weighing noisy channels and focusing on specific feature sets using the SENet attention layer.
- It is compatible with different CNN and other DL models.

This paper is divided into five sections. Section 1 is the introduction of the background concept and explanation of the human-animal conflict problems in Assam. Various existing systems are elaborated, and the need for an effective automated system is also summarized in this section. Section 2 is a review of some related works and a summary of some recent works. The methodology used in the proposed work is discussed in Section 3. A pseudo algorithm for the proposed model, dataset used, model architecture, and training environments with parameters are included in this section. The metrics used for performance measurement are also explained in this section. The results and discussion are included in the section 4. Experimental results are shown here, along with some analysis. The model deployment, comparison with previous works, and impacts of the work are also discussed in this section. Finally, section 5 concludes the work, followed by a future direction and references used in the work.

2. Related work

Some traditional techniques used for handling human-elephant conflicts in Assam, as reported in the handbook of the Assam Haathi Project (Project, 2009), are given below.

(i) Trip Wire Alarms - It is essential to have knowledge about the path that elephants take to get into the agricultural area. As a result, certain poles are installed in the entry point, and then the

Table 1
Summary of some recent works. The model used in the respective works, objective, dataset used, size of the dataset in terms of images and accuracy of the model is shown.

Reference	Model	Aim/objective	Dataset	Size	Accuracy
Banupriya et al. (2020)	CNN	To detect wild animal	Elephant train dataset, Cheetah train dataset	55 images	86.79%
Ghosh et al. (2022)	A series of CNN model, N1, N2, N3, N4, N5 & N6	To predict human-animal conflict in an unprotected area	An anonymized dataset comprising images of cattle-animal and human-animal conflicts	2628 satellite imaginary images covering area 10 km × 10 km, 8 km × 8 km, 4 km × 4 km	82.2%
Yuvaraj et al. (2022)	HOG + CNN	To recognize animals at night time	Thermal images. Contains deer with different pauses.	Out of 1500 images, only 1068 are applied	91%
Benjumea et al. (2021)	YOLO-Z	To recognize small objects	Car rally images	About 4000 images	96%
Zhang et al. (2023)	Improved YOLOv5	To detect human-animal conflict	Wild animal dataset containing tiger, panda, elephant, squirrel, giraffe, butterfly. The dataset has labelling annotations in XML format.	6050 images	95.6%

wires connecting these poles are strung together. The wire is damaged whenever elephants come close, and as a result, an alarm is triggered.

(ii) Watchtower- In the forests bordering areas, elephants suddenly enter into agricultural areas, frequently during the crop seasons. The villagers can better respond to the threat posed by elephants if they can access information via a watchtower. A watchtower is a simple building that consists of a platform, attached either to a tree or a freestanding building, elevated a few meters above the ground or may accommodate one or two people.

(iii) Techniques for sending elephants back to the forest: As soon as the villagers are alerted, many different strategies are used in Assam to send back the elephants. Common strategies are the use of fire, very loud noise created by using firecrackers, vehicle horns, drumming on drums, and rifle shots.

By utilizing technology, we may work towards developing solutions that are both effective and environmentally friendly to reduce the severity of these disputes and promote the harmonious coexistence of elephants and humans (Yuvaraj et al., 2022) (Tuia et al., 2022) (Ghosh et al., 2022). An automated method of monitoring human-animal conflict is required for many reasons, including the following:

(i) *Early detection*: An automated system can use a variety of sensors and technologies, such as drones, satellites, and ground-based sensors, to detect and identify elephant movements and behaviours. This allows for early detection through which fast responses and interventions can be taken to prevent or limit the escalation of conflicts before they even begin.

(ii) *Timely Response*- A prompt response is provided by automated monitoring systems, which, once a conflict has been identified, can immediately warn the appropriate authorities or local populations. This enables prompt reaction tactics to be undertaken, such as deploying trained staff to the impacted areas or utilizing deterrent measures to steer elephant herds away from human settlements. In addition, this makes it possible to implement initiatives in a timely manner.

(iii) *Accuracy and dependability*: When compared to human observation alone, automated systems can provide more accurate and reliable data (Yuvaraj et al., 2022). They can constantly watch enormous swaths of land, maintain a consistent data record, and examine patterns and trends over time. This information may help understand the dynamics and causes of disputes, which may then be applied to developing effective solutions for conflict mitigation.

(iv) *Cost-effectiveness*- The monitoring of human-elephant conflict by humans can be difficult and resource-intensive, as it requires continual surveillance and committed staff. However, this type of monitoring can be cost-effective. Once installed, automated monitoring systems can function at all hours of the day and night. This eliminates the requirement for continuous human presence and can potentially reduce the overall expenses connected with monitoring and response operations (Tuia et al., 2022).

(v) *Decision-making*: Automated monitoring systems can offer useful insights by collecting and analysing data on elephant movements, behaviour, and occurrences of conflict. These realisations can serve as a roadmap for evidence-based decision-making, such as locating high-conflict regions, formulating land-use strategies, or developing efficient elephant conservation initiatives (Tuia et al., 2022).

Safety for humans and elephants: Human-elephant conflicts can be hazardous for both parties involved, leading to injuries or even deaths for both elephants and humans. Keeping both elephants and humans safe is important. An automated monitoring system can assist in protecting the lives of people and their property by providing early warnings and making it possible for rapid intervention (Gogoi and Hira, 2020). This can also protect the well-being of elephants by lowering the number of times they retaliate against humans.

Bhagabati and Sarma (2016) highlighted various object detection techniques using video input. Arshad (2021) has reported the use of various CNN models in object detection. The use of deep neural networks (DNN) for object recognition is reported by Taskiran et al. (2020). Various AI-based techniques for animal detection are described in the literature. A CNN model was developed by Loos et al., 2011 for re-identifying great apes and later modified by Loos and Ernst (2013) to recognize chimpanzees. Recently, Tuia et al. (2022) have studied the problems of conventional animal monitoring and the perspective of ML approaches for wildlife conservation. A number of performance improvements and adopting solutions based on ML and sensors are highlighted in their work. The method of elephant recognition with classical computer vision methods has been discussed in the work of Ardovini et al. (2008). For the detection of elephants, they used shapes as well as nicks in the ears. Premarathna and Rathnayaka (2020) has studied the problem of human-elephant conflict, where 90% of elephant recognition was reported using CNN. The application of the YOLOv5 model for object detection and face recognition from images and videos has been demonstrated by Bhagabati and Sarma (2022b). In another recent work (Bhagabati and Sarma, 2022a), the parameters for performance improvement of CNN are studied. A model with a fine-grained dataset for detecting elephants using the YOLO model is reported by Körschens and Denzler (2019). Various challenges, like strong colour variations of animals, and a small number of images for each class of image, have been considered in the dataset.

Khalajzadeh et al. (2014) depicted the use of CNN with simple logistics classifiers for face recognition problems. The problems of non-

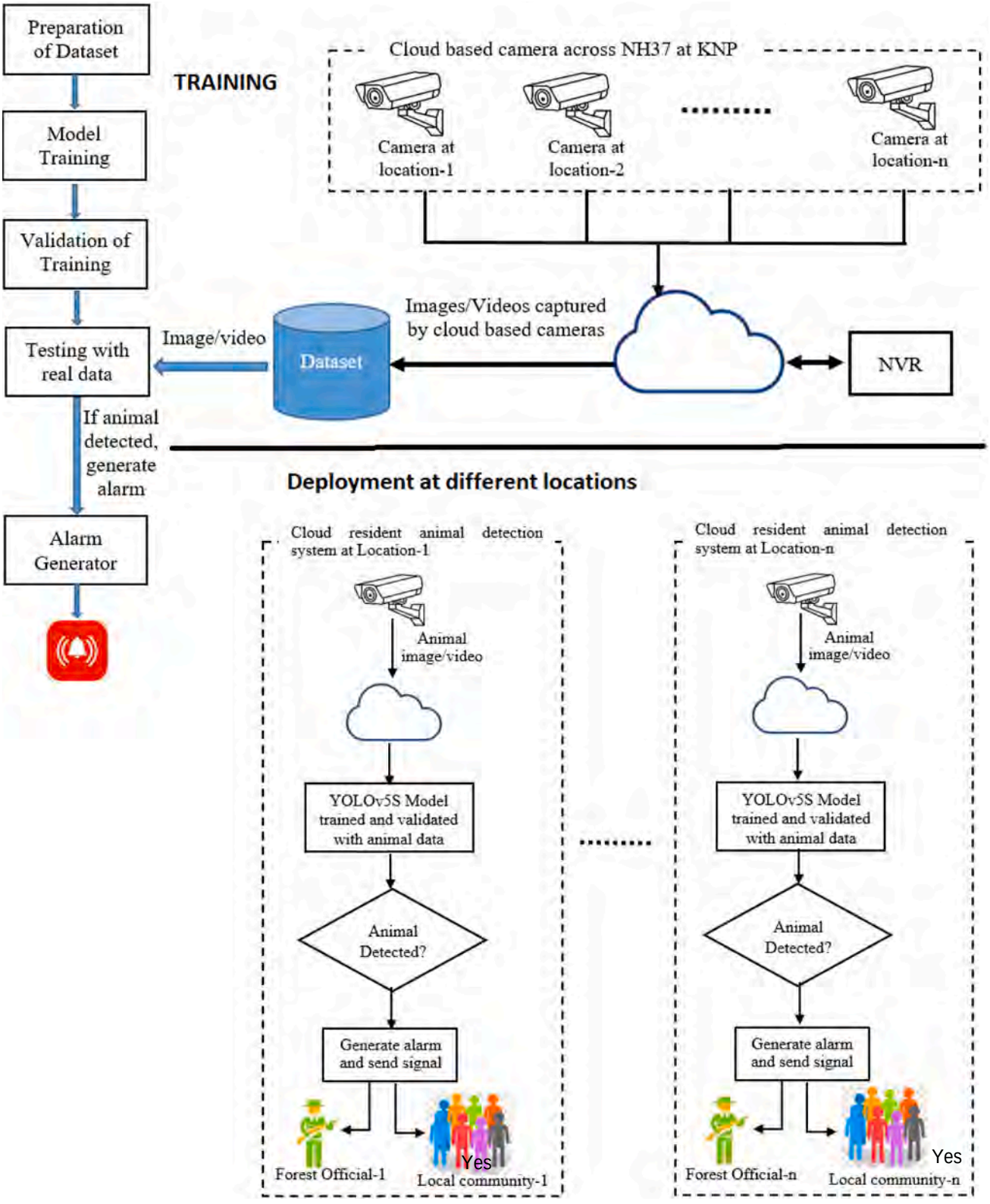

Fig. 1. Block diagram of the proposed approach showing the model training steps, the cameras connected to capture the images and videos of wild animals, and the scheme for deployment of the trained model.

uniform illumination in face recognition using the CNN are discussed by Schroff et al. (2015). Error occurrences for various number of epochs for different CNN system was addressed by some researchers (Nakada et al., 2017; Ramaiah et al., 2015).

Recently, the problem of animal-vehicle collisions at the roadside has been addressed by Yuvaraj et al. (2022). This work has applied a CNN for the detection of wild animals in the nocturnal period for plying of vehicles safely using a set of real-world data acquired with a thermal camera on the move in the city of San Antonio, USA. About 91% highest accuracy has been reported in detecting wild deer. In another recent work, Ghosh et al. (2022) have considered the human-wildlife conflicts in the Bramhapuri Forest Division in Maharashtra. They have claimed

that their work is the first effort to predict human-wildlife conflict in an unprotected area. This work has considered the data set in which 0.38% conflict per 100 km^2 as well as different datasets covering area by k km $\times$ k km in square, where $k = 4$ km, 8 km, 10 km. The main drawback of this approach is that if land use and land cover are considered, then only this model is applicable. For other cases, this model is not applicable. CNN has been used for detecting wild animals with an accuracy of about 82.2%. Banupriya et al. (2020) reported on the use of the DL algorithm for identifying wild animals to lessen animal-vehicle accidents. In this work, a feature-based matching technique has been used for detecting animals. As illumination of each image varies, which is why images are required to be normalized for detection. Considering the small object detection method, the YOLOv5 model has been modified for better recognition of small objects in a special application of autonomous racing (Benjumea et al., 2021). The effect on performance and inference time has been investigated by replacing certain structural elements of the model. The improved version of the YOLOv5 model is called YOLO-Z. In another recent work (Zhang et al., 2023), the YOLOv5s model has been studied for animal detection, and the model has provided improved performance for real-time target detection for animals. They used the RepVGG model (Ding et al., 2021) for simplification of network structure, the sliding window method for increasing convergence speed and real-time performance, and finally, they used C3TR for enhancement of feature extraction and feature fusion ability of the model. The dataset used in the work contained wild animals such as pandas, squirrels, tigers, elephants, giraffes, and butterflies. They improved the YOLOv5 model for detecting wild animals, and the highest accuracy obtained is 95.6%. The summary of the recent developments in the area of animal object detection is shown in Table 1.

From the literature review, it is observed that different CNNs have been applied on different datasets with gradual improvement of the accuracy of detection over time. Emphasis is given to accuracy as response time varies depending upon the processor used and codes. Different datasets for animal detection are used, some consisting of one type of animal having different poses and some consisting of multiple types of animals with different poses. Though the YOLOv5 model (Benjumea et al., 2021) has been improved to detect small objects only, and even though the modified version YOLO-Z has improved accuracy, this model may not be applicable in human-animal conflicts as all wild animals are not of small size. Again the YOLOv5 model has been modified to improve wild animal detection (Yuvaraj et al., 2022). The authors have claimed that the model can be extended for detecting other animals. However, it is not mentioned whether the model will work for other animals in both day and night vision.

The use of thermal cameras for animal detection is also seen in the literature. However, the use of CNN for detecting wild animals at night is a challenging task, and it needs to be addressed properly. In the process of detecting wild animals, further improvement in accuracy is needed. Further, there is an opportunity to develop approaches that are proficient in working well in a generalized approach under both day and night conditions with background variations for detecting human-animal conflict. The YOLOv5 model, with certain modifications and additions, is found to be suitable for developing a generalized framework for the detection of human-animal conflict under both day and night conditions with background variations. Especially, the addition of attention layers as part of the primary detection network helps not only to focus on key areas of the scene under study but also provides optimization in the training and enhanced accuracy. In view of the above, in this work, a SENet attention layer (Hu et al., 2019) is added to YOLOv5 for detecting human-animal conflict under both day and night conditions with background variations. The proposed network is extensively trained with samples of public databases and video streams capturing the scenes under study. The combination produces appreciably better outcomes. The details of the design have been included in the subsequent sections.

3. Methodology

In the present work, a model for real-time animal detection and alarm generation for the protection of the lives of human beings and wild animals is proposed. Beyond elephants, the proposed model can detect other wild animals, such as deer, tigers, etc. specially configured for conditions around the KNP. The model architecture is shown in Fig. 1.

Cloud service is used for connecting the cameras deployed in various locations in the KNP, and the images captured by these cameras are stored in a database for testing the model. For the actual deployment of the trained model for animal detection and alarm generation, a local cloud with protection against cyber threats is used to host the model and connect the camera deployed in that area. Subsequently, the local cloud facility is integrated into a global cloud which forms the crucial element of the work.

There are two phases in the model. In the first phase, the YOLOv5 model with the SENet attention layer is trained with variable parameters using open-source database samples. These trained models are tested with the data captured by cloud resident AI-system connected cameras deployed in different ranges of the KNP and along the NH-37 around the KNP. Four different datasets, by capturing images by cameras installed at four different locations in KNP, are generated for testing. Performance variation of the trained models is observed. In the second phase, the trained model is deployed to work in concert with a local cloud server for the detection of wild animals in that locality. Once wild animals like elephants, tigers, deer, etc., are detected, the information is passed to the local forest officials and to the public as an alarm. This second phase deployment is replicated in each of the target locations, as depicted in Fig. 1. The pseudo algorithm for the scheme is shown in Pseudo Algorithm 1.

Pseudo Algorithm 1. A method for automated approach for human-animal conflict minimisation using YOLO and SENet Attention Framework.

Input: Number of Classes, Class names, images, videos
Output: Alarm generated due to detection of animal
1. Load image dataset
2. Define model architecture as follows
2a. Backbone network (YOLOv5sBackbone with SENet)
2b. Neck Network (YOLOv5sNeck)
2c. Detection head (YOLOv5sHead)
3. Train the model:
3a. Compute loss on a batch of images
3b. Compute gradients and update weights using the optimiser
4. Prediction:
4a. Remove overlapping prediction
4b. Output final detection results (as bounding boxes, class probabilities, confidence score)
5. Detection:
5a. Use the model weight and detect objects from input images or video captured by cameras installed at different locations.
5b. If the animal is detected at location L, send a signal to ForestGuard-L and LocalPublic-L
5c. Raise alarm at location L

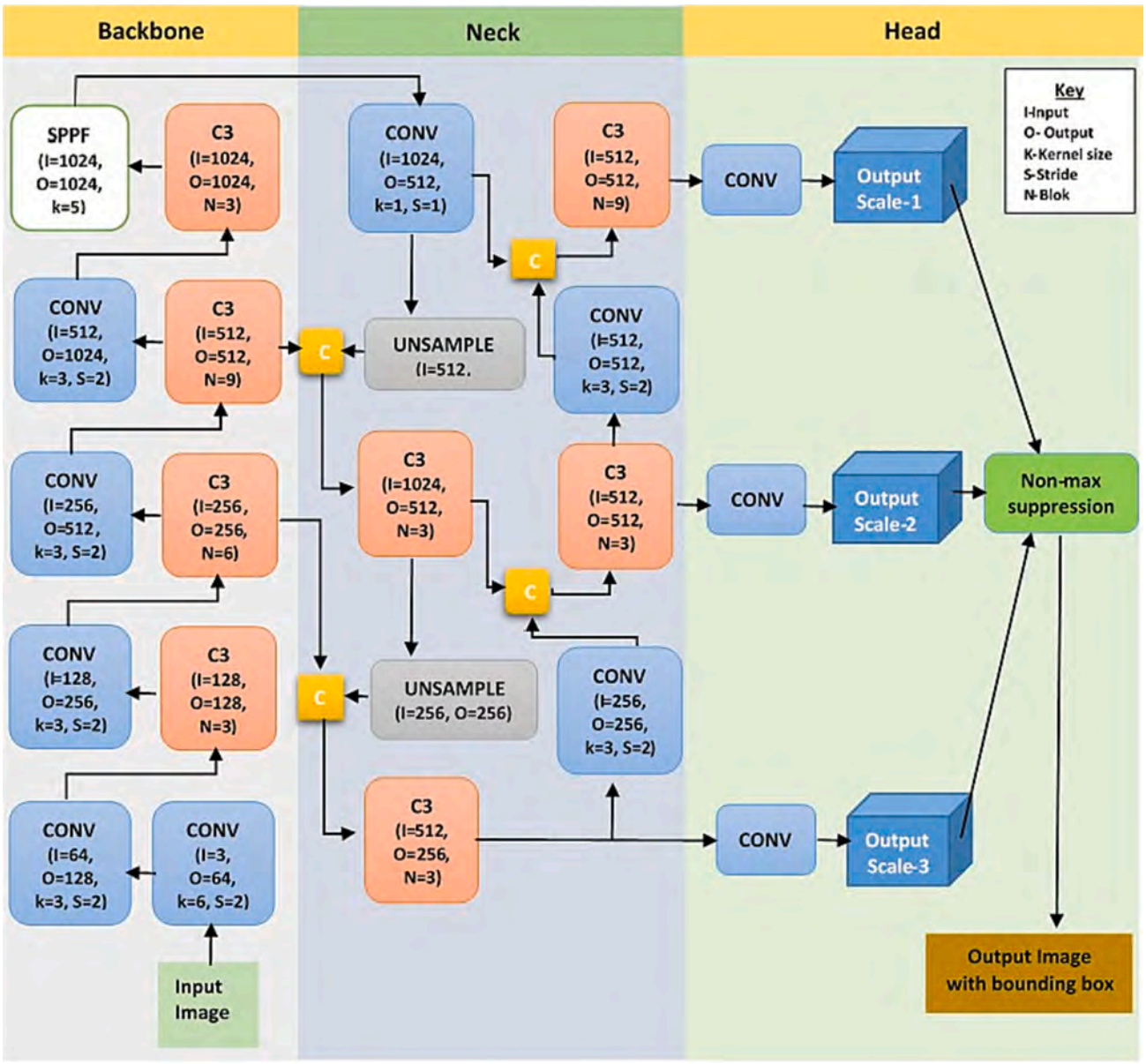

Fig. 2. YOLOv5 network architecture showing the three components of the model, namely Backbone, Neck and Head. Nos. of input layers, output layers, kernel size, strides and blocks are also shown (Xu et al., 2021).

3.1. Dataset

A publicly available dataset in Roboflow (Roboflow Dataset, n.d) named Animal2–v1 comprising 9952 images belonging to classes Bear (1530 images), Deer (966 images), Elephant (1684 images), Leopard (1888 images), Monkey (1214 images), Tiger (1388 images) and Wildboar (1282 images) is used. All the images are annotated in YOLOv5 PyTorch format. The preprocessing applied for the dataset is auto-orientation of pixel data with EXIF-format stripping, and the images are resized to 416 × 416 (stretch). Here, 70% of images are used for training, 20% images for validation, and 10% images are used for testing. No image augmentation techniques are applied.

3.2. Proposed YOLOv5- SENet Attention Network Architecture

The work is based on the popular YOLOv5 architecture integrated with a SENet attention mechanism (Zhu et al., 2021).

The YOLO (You Only Look Once) is a simple and extremely fast object detection algorithm that avoids a complex pipeline and considers frame detection as a regression problem (Redmon et al., 2016). YOLO processes real-time video with a latency of less than 25 milliseconds (ms), and the mean average precision (mAP) achieved is more than twice that of other such systems. An entire image sequence is fed as input to YOLO during training and testing; thereby, the contextual information about classes and their appearance is implicitly encoded. The input image is divided into grid cells of size $S \times S$. A grid cell is responsible for predicting B bounding boxes with five predicted values x, y, w, h, and c in each bounding box. (x, y) is the coordinate of the center point of a bounding box relative to a grid cell, w and h are the width and height of the bounding box, and c is the confidence score of an object being present in a bounding box. With C class probabilities, the prediction is encoded as ($S \times S \times (B \times 5 + C)$) tensor. YOLO has several versions. YOLOv5 is considered here (Fig. 2). The architecture of YOLOv5 consists of the following 3 components, as depicted in Fig. 2 (Xu et al., 2021).

- Backbone: A CNN, which acts as the main body of the network, is designed using the New CSP-Darknet53 (Cao et al., 2023) structure. It extracts key features from the input image.

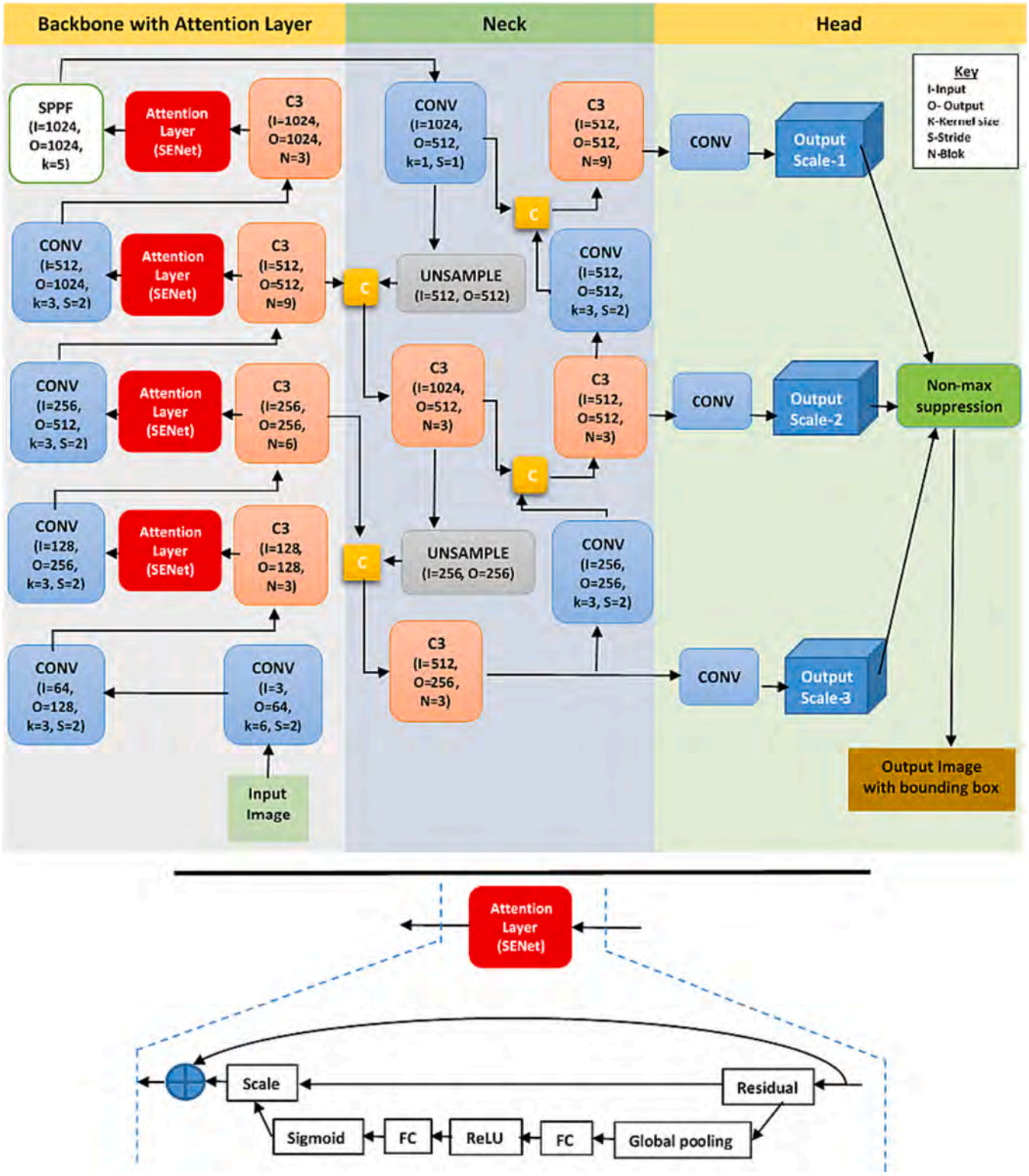

Fig. 3. YOLOv5 architecture with proposed attention layer. YOLOv5 blocks, SENet attention layer blocks and the embedded attention layers in the Backbone of the YOLOv5 model are shown.

- Neck: It connects the backbone and head part utilizing Spatial Pyramid Pooling Fast (SPPF) (He et al., 2015) and Cross-Stage-Partial (CSP)- Path Aggregation Network (PAN) structure (Wang et al., 2019). The CSP-PAN creates feature pyramids which help the model to generalize well for object scaling. It identifies the same object in various sizes and scales. C3 blocks concatenate the features together and create a richer representation capturing both low-level and high-level information.
- Head: It is the final detection step and generates final output vectors, class probabilities, objectness scores, and bounding boxes. Head uses Non-Maximum Suppression (NMS) (Hosang et al., 2017) for screening the multi-object box and output the predicted image.

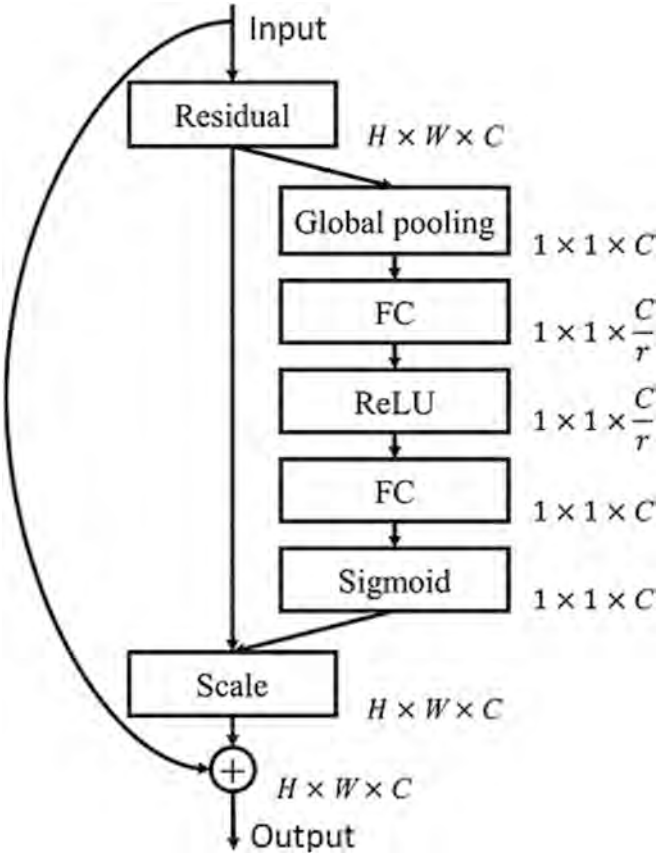

Fig. 4. SENet attention mechanism. Different blocks of the attention mechanism with sizes are shown.

Table 2
Hyper parameters and their values used in training the proposed model. The hyper parameters are based on the optimized values for YOLOv5 COCO training from scratch.

Parameter Name	Value	Parameter Name	Value
lr0	0.01	*fl_gamma*	0.0
Lrf	0.01	*hsv_h*	0.015
momentum	0.937	*hsv_s*	0.7
weight_decay	0.0005	*hsv_v*	0.4
warmup_epochs	3.0	*degrees*	0.0
warmup_momentum	0.8	*translate*	0.1
warmup_bias_lr	0.1	*scale*	0.5
box	0.05	*shear*	0.0
cls	0.5	*Perspective*	0.0
cls_pw	1.0	*flipud*	0.0
obj	1.0	*fliplr*	0.5
obj_pw	1.0	*mosaic*	1.0
iou_t	0.2	*mixup*	0.0
anchor_t	4.0	*copy_paste*	0.0

Though the YOLOv5 model has been used for object detection, it evolved with four major versions - YOLOv5s, YOLOv5l, YOLOv5m, and YOLOv5x (Bhagabati and Sarma, 2022a), the YOLOv5s version is considered for the current model (Dlužnevskij et al., 2021).

3.2.1. Attention mechanism

DL models have attention techniques applied to them in order to improve performance (Guo et al., 2022; Caron et al., 2021). This is done mostly for the purpose of concentrating on particular parts of the inputs. The method that lies behind the attention mechanism has been created with the goal of enhancing the capability of feature extraction, which in turn makes the overall effort of object detection even more accurate and

Table 3
Training summary for YOLOv5s model at epoch 150. The first column shows the animal classes. The second and third columns are the precision (P) and recall (R) of the trained classes, respectively. The fourth column is the mean average precision (mAP) calculated at a 50% threshold value for intersection over union (IoU). The fifth column shows mAP over multiple thresholds from 0.5 to 0.95 with steps of 0.05. The last column shows the model training time in hours. The first row shows the average values for all classes.

Class	P	R	mAP@0.5	mAP@0.5:0.95	Model Training Time
All	0.797	0.778	0.812	0.480	3.331 h
Bear	0.663	0.688	0.815	0.541	
Deer	0.741	0.762	0.733	0.581	
Elephant	0.658	0.812	0.774	0.602	
Leopard	0.917	0.812	0.914	0.485	
Monkey	0.861	0.829	0.673	0.356	
Tiger	0.915	0.859	0.921	0.453	
Wildboar	0.824	0.75	0.854	0.345	

Table 4
Training Summary for YOLOv5s with SENet Attention layer at epoch-150. The first column shows the animal classes. The second and third columns are the precision and recall of the trained classes, respectively. The fourth column is the mean average precision (mAP) calculated at a 50% threshold value for intersection over union (IoU). The fifth column shows mAP over multiple thresholds from 0.5 to 0.95 with steps of 0.05. The last column shows the model training time in hours. The first row shows the average values for all classes.

Class	P	R	mAP@0.5	mAP@0.5:0.95	Model Training Time
All	0.897	0.866	0.912	0.597	3.401 h
Bear	0.868	0.849	0.925	0.61	
Deer	0.936	0.887	0.923	0.721	
Elephant	0.828	0.849	0.897	0.671	
Leopard	0.937	0.912	0.954	0.565	
Monkey	0.891	0.829	0.873	0.506	
Tiger	0.955	0.889	0.941	0.663	
Wildboar	0.864	0.85	0.874	0.445	

optimized. The introduction of an attention mechanism into the YOLO block enables the model to selectively attend to important features while it is being trained. This results in an increase in both the training efficiency and the classification precision of the model. The YOLO is useful in its application, but it may have limitations in terms of adaptively focusing on specific areas or locations, which may not be static in nature in all of the available cases. This could potentially lead to the loss of information or contribute towards a rise in redundancy during the learning process. As a consequence of this, the SENet attention mechanism has been included to provide the YOLO with assistance in concentrating its attention on the prominent characteristics that play an important role in making recognition effective despite a range of variations. The SENet attention mechanism is used to apply weights to the features that are retrieved by the network that comes before it (Zhu et al., 2021). This makes it possible to differentiate key features despite variations in animal kind, size, distance, illumination, and background.

There are numerous attention processes, each of which places a premium on a certain aspect of attention. The squeeze-and-excitation network, also known as SENet, is a specialized attention mechanism that utilizes a distinct channel network for the purpose of feature extraction. According to Hu et al. (2019) research, SENet is able to recalibrate channel-wise feature responses by explicitly modelling channels' interdependencies. In the current work, the SENet attention mechanism is utilized in an effort to increase the overall performance of the YOLOv5s model when it comes to the detection of objects. In this study, the SENet attention mechanism is embedded in the backbone of the YOLOv5 model, as shown in Fig. 3. Despite the fact that the SENet attention mechanism can be applied to all three sections of the YOLOv5 model, namely the backbone, the neck, and the head, the SENet

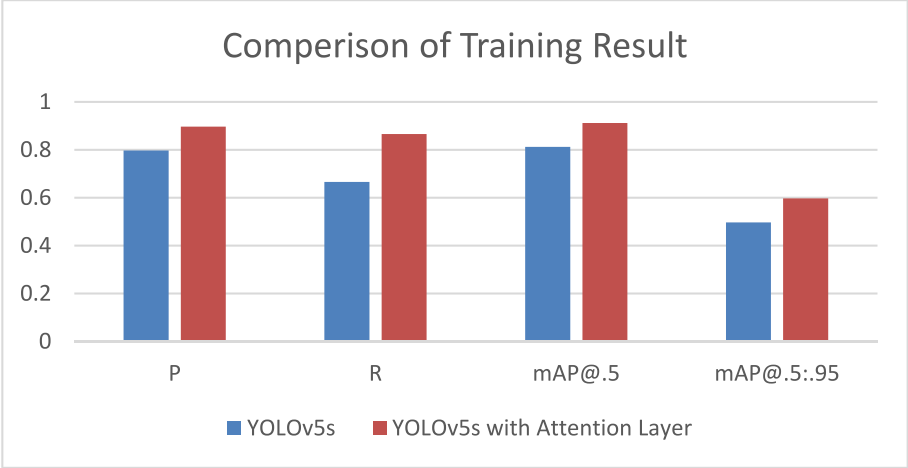

Fig. 5. Comparison of training results of YOLOv5s and YOLOv5s with attention layer. Precision (P), Recall (R), mean average precision (mAP) calculated at a 50% threshold value for intersection over union (IoU), and the mAP over multiple thresholds from (0.5 to 0.95 with steps of 0.05) are compared for training results for both the cases.

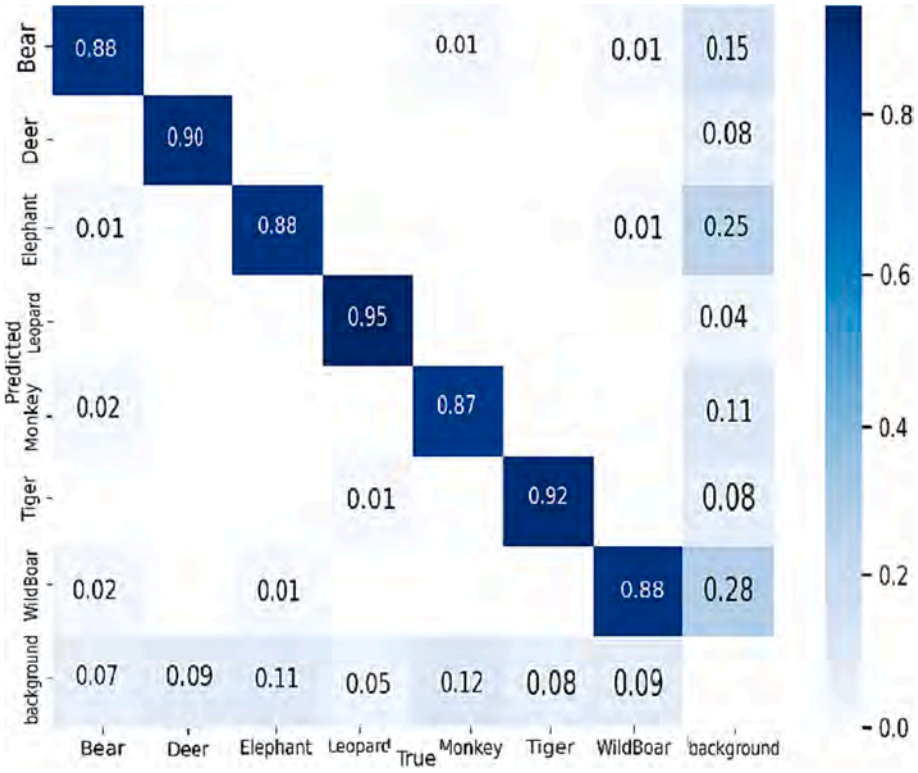

Fig. 6. Confusion Matrix of the trained model representing the performance of the proposed model on the validation set. The rows correspond to the ground truth classes, and the columns correspond to the predicted classes.

attention mechanism is embedded in the backbone of the YOLOv5 model because it is more convenient and drives the overall performance of the system. Moreover, the main feature extraction is done in the backbone. Therefore, if SENet is embedded in the backbone, it can extract more features and enhance the performance of the model. This is the key aspect of the present work.

The SENet attention mechanism comprises squeeze and excitation operations. The architecture of the SENet attention mechanism (Hu et al., 2019) is shown in Fig. 4. In the mechanism, the residual block is a stack of layers that takes a convolutional block as input and adds the output of the convolutional block to a deeper layer (He et al., 2016). The global pooling block squeezes each channel into a single numeric value using average pooling. The fully connected (FC) block is a dense layer followed by a ReLU layer which adds non-linearity and reduces the output channel complexity by a ratio. Another FC, a dense layer, followed by a sigmoid, provides each channel with a smooth gating function. The final and last process in the architecture, the scale block, weighs each feature map of the convolutional block based on the excitation network. For a given input x, the squeeze step for the C-th channel in the global average pulling (GPA) process is given by,

$$Z_C = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j) \tag{1}$$

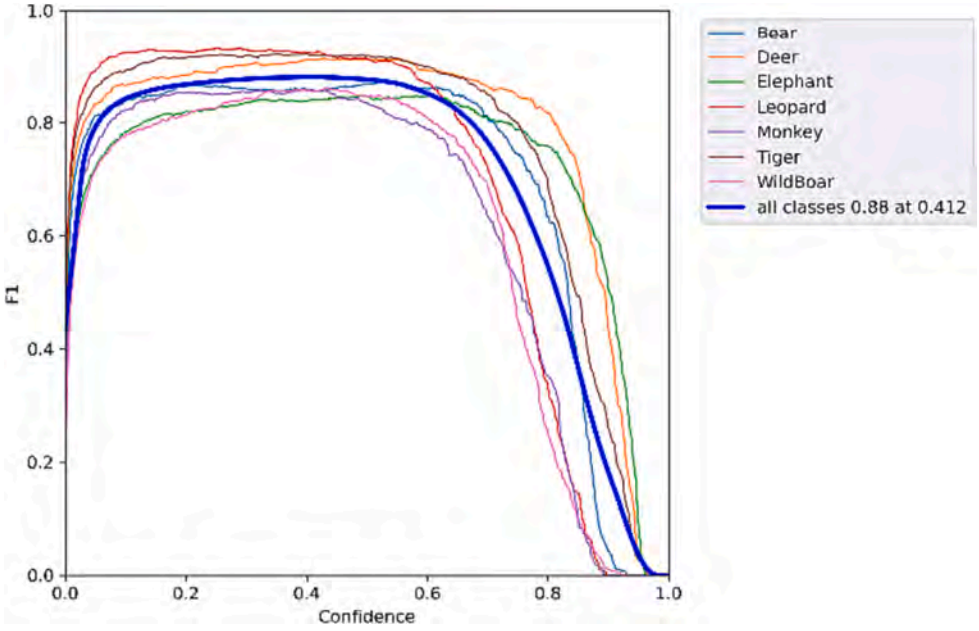

Fig. 7. F1 Curve.

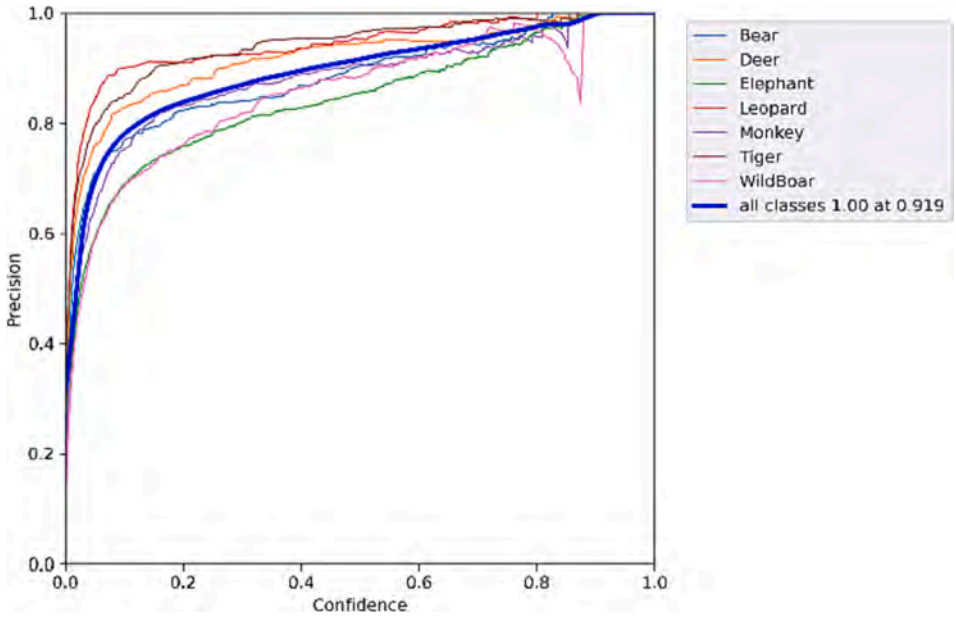

Fig. 8. P Curve.

where Z_C is the output of the C-th channel. The main aim of the excitation step is to fully capture the channel-wise dependencies.

$$\widehat{x} = x^{*}\sigma\left(\widehat{Z}\right) \tag{2}$$

where *is the channel-wise multiplication, σ is the sigmoid function, and $\widehat{Z}$ is the result generated by a transformation.

$$\widehat{Z} = T_2\left(\mathrm{ReLU}(T_1(Z))\right) \tag{3}$$

T_1 and T_2 are the two linear transformations that can be learned to capture the importance of each channel, with ReLU used in between.

A DL network can learn more relevant features after the addition of the SENet block, which models the feature map weight information of the input channel dimension. This contributes to an overall improvement in the model's capacity to recognize patterns (Niu et al., 2021). The global average pooling is used in the squeeze step to lower the spatial dimension of the input feature maps, which results in a compressed representation of those maps. Once again, during the time of excitation, channel-wise dependencies are represented by modelling the interdependencies among the compressed feature maps that were obtained from the squeezing step. The term "excitation layer" refers to the two layers that are completely coupled to one another. The channel-wise attention weights are produced once the excitation block has been processed. When everything is done, scaling and rescaling are accomplished by multiplying the attention weight by the initial feature maps. This reduces the weight of the channels that are not important while increasing the weight of the channels that are significant. As a result of this, the model's performance can be improved by increasing the channel weight of the important channel.

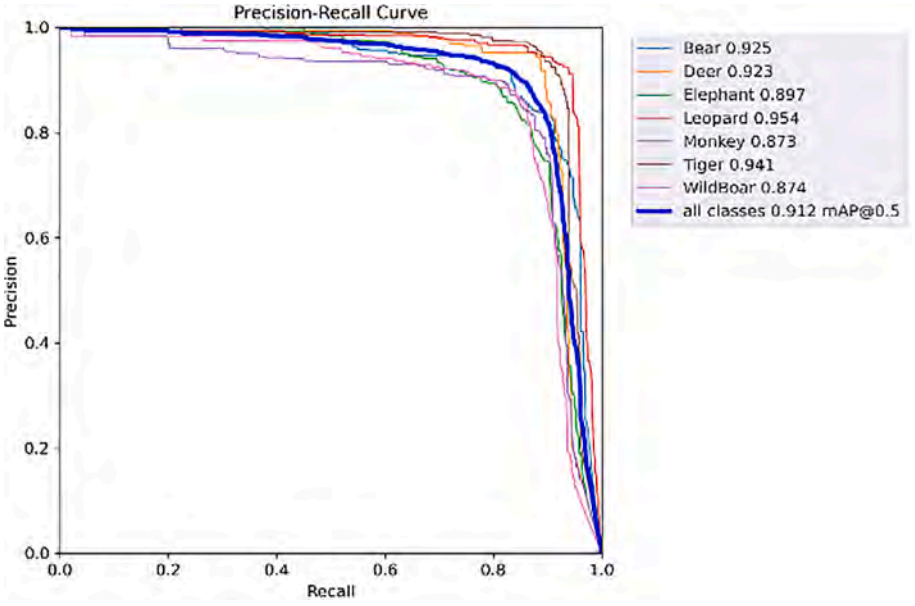

Fig. 9. PR Curve.

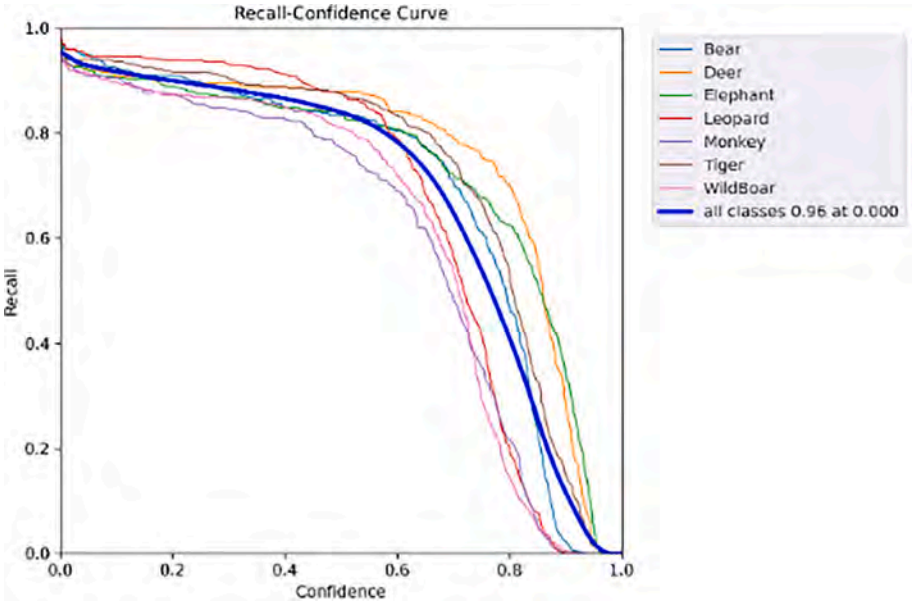

Fig. 10. R Curve.

4. Training

The previous version of YOLO was in DarkNet, but YOLOv5 has been shifted into the PyTorch framework (Bhagabati and Sarma, 2022a). So the model in this work is trained in the PyTorch framework with the following parameters. The default values of the hyper parameters optimized for YOLOv5 COCO training from scratch are used here, whose values are shown in Table 2.

The various functions used in the training of the model are as stated below.

(i) Activation Function - The Leaky ReLU activation function is used in the middle/ hidden layers, and the sigmoid activation function (Pretorius et al., 2019) is used in the final detection layer.

(ii) Optimization Function- Stochastic Gradient Descent (SGD) (Bottou, 2012) is used for training.

(iii) Loss Function Calculation- The compound loss calculator of the YOLO family based on objectness score, class probability score, and bounding box regression score is used (Bhagabati and Sarma, 2022a).

The state-of-the-art YOLOv5s model is used in this experiment. The network training is done by using a publicly available dataset in Roboflow which was annotated and exported into Yolo PyTorch format. The dataset was labelled with seven classes- (i) Bear (0), (ii) Deer (1), (iii) Elephant (2), (iv) Leopard (3), (v) Monkey (4), (vi) Tiger (5) and (vii) Wildboar (6).

The training is done from scratch using GPU (Tesla T4) and CUDA Version-12.0 in Google Colab with the parameters (i) Input image size = 416, (ii) Batch size = 16, (iii) Epochs = 100, 150, 200 and 250, (iv) Weights = Default to yolov5.pt., (v) Data = custom data set and (vi) Cache = True.

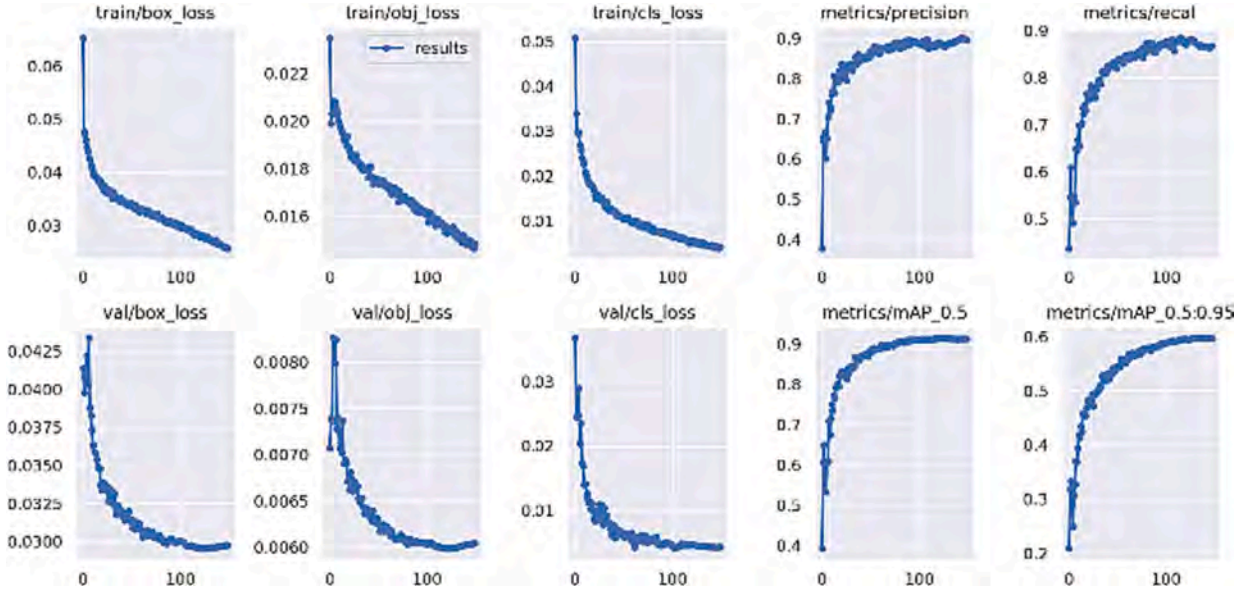

Fig. 11. Training Plots for the proposed model.

In recent years, it has been observed that deep learning methods can be trained with close to zero training error efficiently. The number of convolutional and dense layers directly affects the runtime of the model (Bienstock et al., 2023). The running time of a DL model increases in polynomial-terms with the increased number of associated layers. But with the effort of achieving zero error, there are possibilities of over-training and biased training. To avoid these situations, a restrictive and gradually increasing training cycle and accuracy calculation approach has been adopted. Based on mean square error (MSE), the optimal training state for the model has been obtained.

4.1. Performance metrics used to evaluate the model

The tools used to measure the performance of the trained model are the precision (P), recall (R), mean average precision (mAP), and the F1-score metrics (Nepal and Eslamiat, 2022). The precision gives the ratio of the true predictions to the total number of predictions, whereas recall is the ratio of true predictions to the total number of objects present in the image. The F1 score, which is the harmonic mean of the precision and recall, gives the model's test accuracy. Since mAP is also considered as a measure of accuracy for a machine learning algorithm, mAP is used for measuring the performance of the model in this work. These metrics are calculated using the following equations (Maxwell et al., 2021).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$\text{F1 Score} = 2 \times \frac{P \times R}{P + R} \tag{8}$$

$$AP = \sum_{k=0}^{k=n-1} [R(k) - R(k+1)] \times P(k) \Big] \tag{9}$$

$$mAP = \frac{1}{N} \sum_{k=1}^{k=N} AP_k \tag{10}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, FN is the number of false negatives, AP_K is the average precision (AP) of class k, n is the number of Recall Thresholds, and N is the number of IoU thresholds.

The area under the Precision-Recall (P-R) curve is referred to as average precision (AP) or mean average precision (mAP) (Maxwell et al., 2021). The method for calculating AP and mAP for object classification is shown in expression (10). A minimum intersection over union (IoU) defines whether the prediction of an individual object is correct or incorrect. The mAP values are calculated at a 50% threshold value for intersection over union (IoU), i.e., all the predicted bounding boxes that resulted in ratios of overlapping areas to the union areas with ground truth bounding greater than 50% are considered, and the remaining are discarded (Yadav et al., 2022). For each IoU threshold, the AP value is calculated, and then the average AP over multiple thresholds from (0.5 to 0.95 with steps of 0.05) is calculated. This gives us the mAP for a single class at a single IoU threshold (Maxwell et al., 2021).

5. Results and discussion

In this section, the results derived from the training of the proposed model and the results obtained by deploying the trained model are discussed.

5.1. Experimental result

Training summary results for YOLOv5s are tabulated in Table 3. The model is also trained with the same training parameters after adding the SENet attention layer. The improved result obtained after training the model with the attention layer is shown in Table 4. It is observed that after embedding the SENet attention mechanism, the training result for mAP@0.5 is increased by 0.1 for All, 0.11 for Bear, 0.19 for deer, 0.12 for elephants, 0.4 for leopard, 0.2 for monkey, 0.02 for tiger, and for wildboar, it is 0.02. The training accuracy has also increased for mAP@0.5:0.95 values. The comparison of average training results for all is shown in Fig. 5. From the data and the comparison, it is observed that the SENet attention mechanism has enhanced the performance of the YOLOv5 model.

The confusion matrix (Fig. 6), F1 curve (Fig. 7), P curve (Fig. 8), PR curve (Fig. 9), R curve (Fig. 10) and final training plots (Fig. 11) are shown for the training of the model with attention layer with epoch-150. Some training batch outputs are shown in Fig. 12, and some validation outputs are shown in Fig. 13.

Fig. 12. Training Output (Batch-1) showing the labels of the detected object.

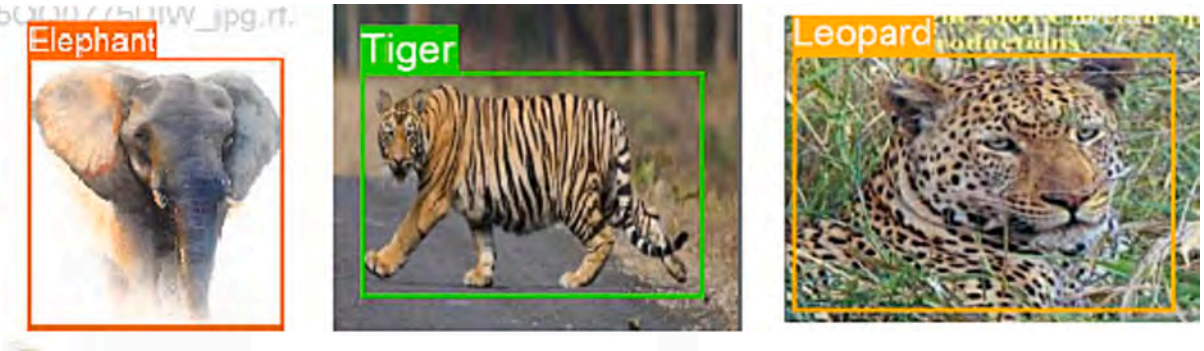

Fig. 13. Validation Output (with images from the dataset). The bounding box and class name are generated and shown for the validated object by the model.

Table 5
Training summary for YOLOv5s with SENet Attention layer at epoch-100. The class names, precision and recall are in the first, second and third columns, respectively. The fourth column is the mean average precision (mAP) calculated at a 50% threshold value for intersection over union (IoU). The fifth column is mAP over multiple thresholds from 0.5 to 0.95 with steps of 0.05. The last column shows the model training time in hours. The first row shows the average values for all classes.

Class	P	R	mAP@0.5	mAP@0.5:0.95	Model Training Time
All	0.896	0.87	0.911	0.593	2.242 h
Bear	0.886	0.86	0.916	0.603	
Deer	0.921	0.89	0.918	0.696	
Elephant	0.855	0.88	0.909	0.68	
Leopard	0.955	0.92	0.957	0.562	
Monkey	0.868	0.81	0.873	0.502	
Tiger	0.947	0.89	0.939	0.662	
WildBoar	0.844	0.84	0.863	0.447	

Table 6
Training summary for YOLOv5s with SENet Attention layer at epoch-200.

Class	P	R	mAP@0.5	mAP@0.5:0.95	Model Training Time
All	0.892	0.88	0.915	0.598	4.32 h
Bear	0.873	0.88	0.927	0.619	
Deer	0.932	0.91	0.941	0.718	
Elephant	0.823	0.86	0.893	0.669	
Leopard	0.944	0.91	0.948	0.569	
Monkey	0.872	0.84	0.868	0.495	
Tiger	0.963	0.9	0.953	0.67	
WildBoar	0.837	0.84	0.872	0.449	

Table 7
Training summary for YOLOv5s with SENet Attention layer at epoch-250.

Class	P	R	mAP@0.5	mAP@0.5:0.95	Model Training Time
All	0.904	0.87	0.919	0.602	5.391 h
Bear	0.899	0.88	0.937	0.618	
Deer	0.919	0.9	0.922	0.704	
Elephant	0.841	0.85	0.896	0.67	
Leopard	0.947	0.92	0.96	0.571	
Monkey	0.913	0.79	0.884	0.517	
Tiger	0.958	0.9	0.945	0.669	
WildBoar	0.852	0.86	0.886	0.461	

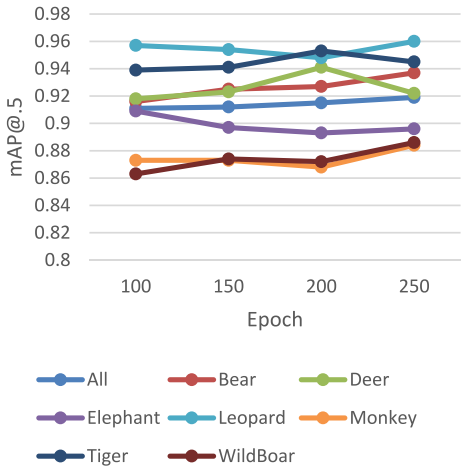

Fig. 14. Change of mAP@0.5 with increase in epoch.

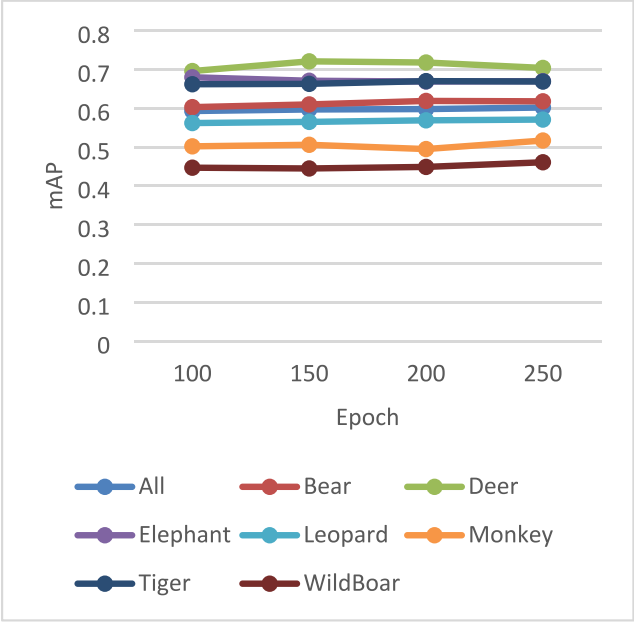

Fig. 15. Change of mAP@0.5:0.95 with rise in epoch.

In order to determine more accurate training results and also to explore the effect of epoch upon training result, apart from 150 epochs, the model with attention layer is trained with epoch values 100, 200, and 250 under a uniform training environment and with the same dataset. The training summary for each of these epochs is shown in Tables 5, 6 and 7 for epochs 100, 200, and 250, respectively. The trends of mAP values with increasing epochs are shown in Figs. 14 and Fig. 15.

The size of the dataset used for custom training is sufficiently large. Further, the DL model used in this work is tuned with optimized hyper parameters for which overfitting and under fitting situations are prevented from occurring. As mentioned above, the effort to over-train the model has been avoided and graded MSE convergence (Figs. 7 to 11) has been adopted to fine-tune the model. Results obtained for different epochs are shown. As the number of epochs increases, mean average precession (mAP) increases.

For the detection of wild animals and conflict situations, the model is tested with real images captured by four different cameras around the NH-37 passing through the KNP. The wild animals are detected successfully, and evolving conflict situations are reported. Deer while crossing roads when vehicles are plying are shown in Fig. 16. Whereas elephants crossing the NH-37 through the animal corridor at the KNP are detected from still images, as shown in Fig. 17. Tigers are also detected from the still images obtained from onsite cameras at the KNP, as shown in Fig. 18. Even the tiger is detected in an image captured during the night where a vehicle is in the background with a headlight on. Fig. 19 shows the detection of elephants from videos.

Apart from testing the model with real images of animals while crossing roads in the KNP, for performance analysis of the model while detecting wild animals, image data is captured using cloud-resident cameras at four different locations in four ranges of the KNP. This way, four datasets are prepared considering wild animals from the Kohora range, Agoratuli range, Bagori range, and Burapahar range. Each of the four trained models with epoch values 100, 150, 200, and 250 are tested with these four datasets. The preprocessing applied is image auto-orientation and re-sizing. The time requirement in millisecond (ms) for preprocessing, inference and Non-Maximum Suppression (NMS) are recorded for each test. The accuracy for each test is noted. The average test result is shown in Table 8. The variation of the test results with epoch values is graphically represented in Figs. 20, 21, 22 and 23.

Fig. 16. Detection of Deer while crossing roads at Kaziranga National Park.

Fig. 17. Detection of Elephant while crossing roads at Kaziranga National Park.

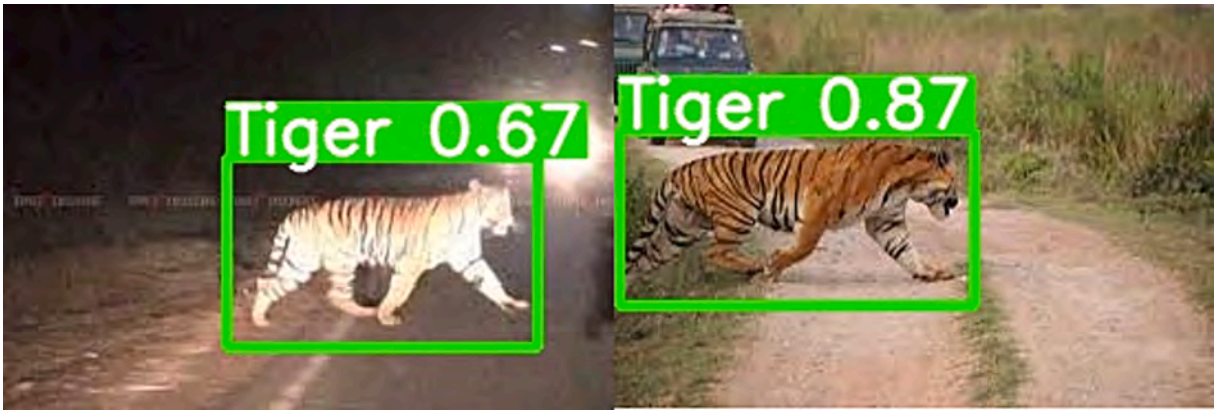

Fig. 18. Detection of Tiger while the tiger is crossing some roads inside the KNP. The image was captured during the night with background lights. The trained model detected the animal under night vision and constrained environment.

5.2. Deployment of the model

The trained model is hosted and run in a cloud-based system. Cloud resident camera is placed in the Kohora range of KNP. The images captured by the camera are fed into the model. The following cloud computing infrastructure is used for the deployment of the system

- Processor: Equivalent to intel core i9 7980XE @ 2.60 GHz
- GPU: Tesla T4 GPU
- Bandwidth: 1Gbps
- Storage: 100GB
- Latency: Maximum 50 ms
- Secured wireless connectivity: WPA3
- Cloud features: Scalable, resource pooling, secured, economic, etc.

The system could detect wild animals with preprocessing – 0.9 ms, inference – 144.2 ms, NMS- 0.8 ms, and accuracy of 95%. The system generated an alarm when it detected wild animals like elephants, tigers, deer, etc. The information could be passed to local forest officials and the local public for alarm and taking appropriate measures. The way of sending the alarm signal is using IoT based alarm generator. The SMS-based system can also be integrated to deliver the detection result to

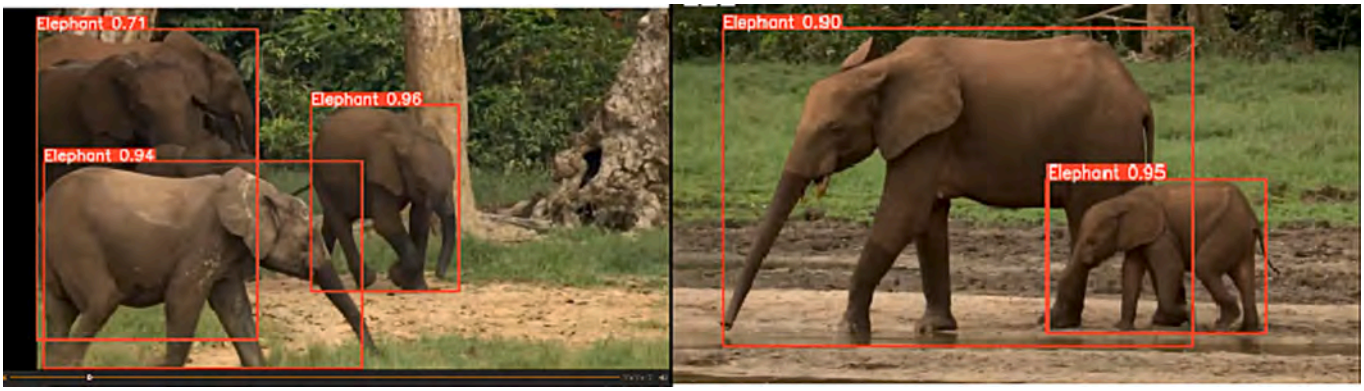

Fig. 19. Detection of Elephant from video. The model detected multi-occurrences of animals from video.

Table 8
Average test of results for four different datasets with four epoch values. Average values for preprocessing, inference and NMS values is in millisecond. The last row is the average accuracy for different epochs.

	Computation time (milliseconds)			
Parameters	Epoch-100	Epoch-150	Epoch-200	Epoch-250
Preprocessing	0.85	0.85	0.875	0.8
Inference	140.68	141.38	145.83	147.95
NMS	0.73	0.8	0.75	0.75
Accuracy	92.6	94.25	94	92.9

the forest official and local community.

5.3. Comparison with previous works

The comparison of the current work with some recently developed models for animal detection and prevention of human-animal conflicts is shown in Table 9. From the comparison, it is observed that the accuracy of wild animal detection with the proposed system is better than the recently developed systems. On the other hand, the minimum accuracy of the proposed system is lesser than the existing three systems, as the proposed system is applied at night time also for detecting wild animals. Therefore, the proposed system detects wild animals at night with low accuracy. The computational complexity (Hu et al., 2019) of the proposed model, as shown by the system, is 15.8 GFLOPs.

Another important point of the proposed DL technique is that even though the images are augmented, the model is able to detect an animal. This is because of the feature extraction process during training, and then the extracted features are applied for detecting images. There are three main methods of feature extraction, namely local, holistic, and hybrid. For example, in the local approach entire face is divided into some small regions and then features are extracted from each small region and then during detection, those extracted features are applied. That is why after changing the images slightly from the original one, either by rotating the image or by changing its contrast, the trained network can work for detecting images.

5.4. Impact analysis

The key novelty of the system is the use of an AI-based automated approach, which provides higher accuracy in detecting human-wild animal conflicts and alarms forest officials and the public continuously throughout the day and night. Forest officials are not required to stand along the boundary of the KNP and monitor the movements of wild animals constantly. Instead, they can attend as notified by the system. It can go a long way in assisting the coexistence of the natural world with humans and minimizing distressing situations.

The impact analysis of the proposed technique looks at how accurate and reliable the system is at finding wild animals compared to similar works that have already been reported. It also looks at the manual ways that people in the area already use to deal with the problem of people and animals getting into fights. This analysis is carried out in order to

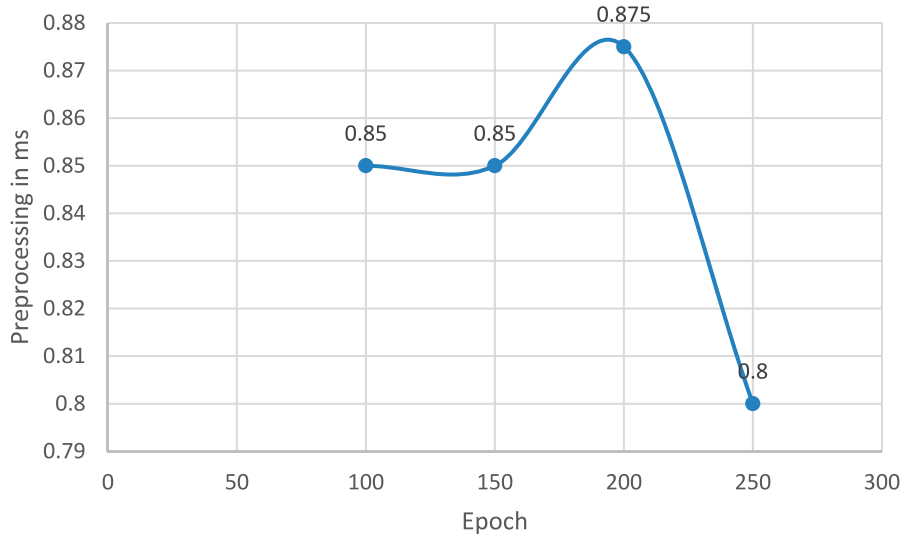

Fig. 20. Variation of preprocessing with epoch.

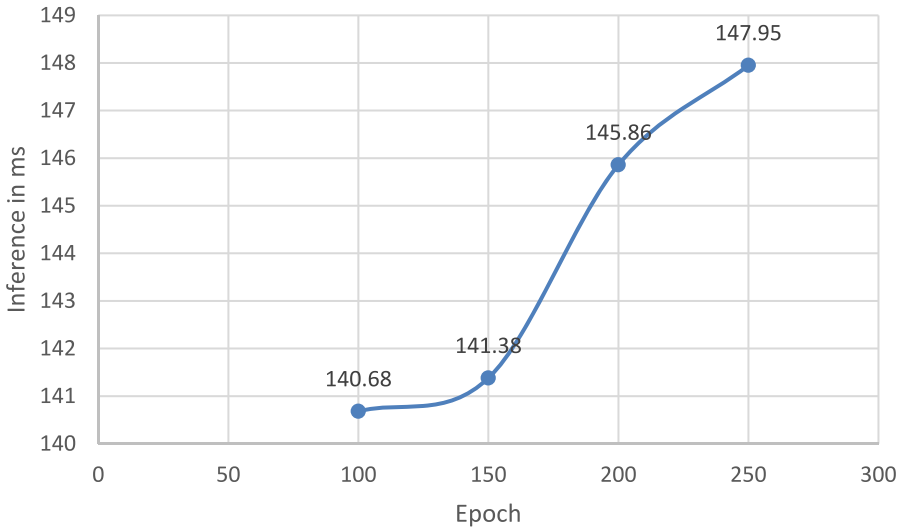

Fig. 21. Variation of inference with epoch.

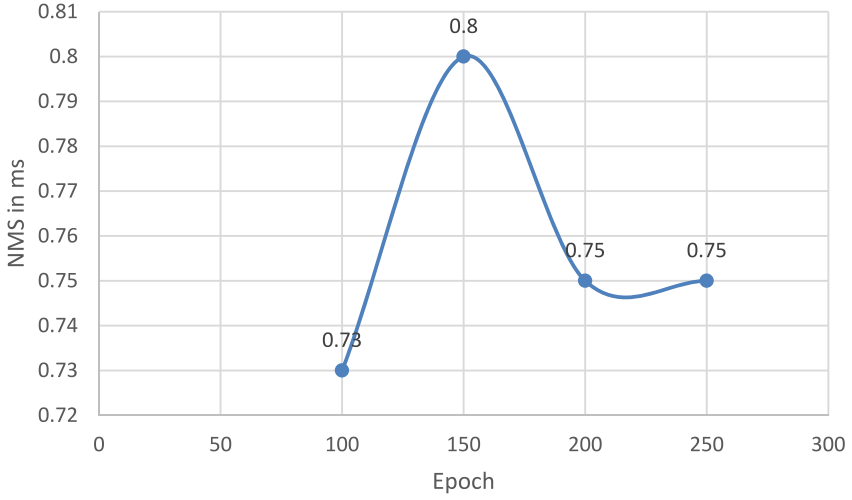

Fig. 22. Variation of NMS with epoch.

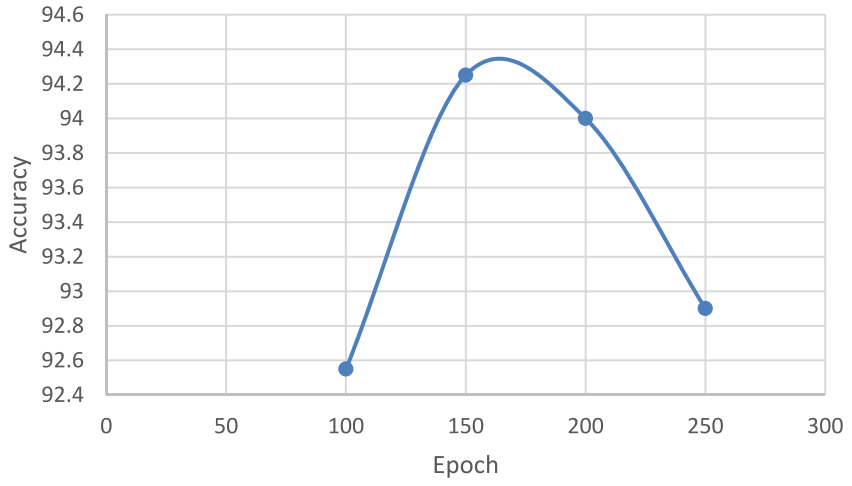

Fig. 23. Variation of accuracy with epoch.

determine how effective the suggested approach is. The system that is being offered is set up to avoid conflicts between humans and animals, as well as to identify a wide variety of animals even when the lighting, distance, and background are all different, while also providing for the automatic generation of alarms. The system has been built and tested in a form that is resident in the cloud, with cameras set at four different

Table 9
Comparison of Results.

Work	Maximum Accuracy	Minimum Accuracy	Model Used	Dataset Used	Size of Dataset
Banupriya et al. (2020)	86.79%	79.25%	CNN model	Elephant train dataset, Cheetah train dataset	55 images
Yuvaraj et al. (2022)	91.00%	86.00%	HOG + CNN	Deer with animals and without animal	Out of 1500 images, 1068 images used
Ghosh et al. (2022)	82.2%	60.20%	CNN model series N1, N2, N3, N4, N5, N6	Own created satellite image dataset SAT1 (10), SAT1(8), SAT1(4)	2628 satellite imaginary images covering area 10 km × 10 km, 8 km × 8 km, 4 km × 4 km
Zhang et al. (2023)	95.6%	91.3%	Improved YOLO-v5	Wild animal dataset comprises tiger, panda, elephant, squirrel, giraffe, butterfly.	6050 images
Proposed model	96.00%	67.00%	YOLOv5s with SENet attention layer	Animal2-v1, comprises images of tiger, beer, leopard, monkey, elephant and wildboar.	About 9952 images

sites, each of which represents a different zone of the KNP with high rates of human-animal conflict. Along the section of the NH-37 that travels through the KNP, as well as in the boundaries towards the human-inhabitant areas of the Park's four different ranges, namely the Kohora range, the Agoratuli range, the Bagori range, and the Burapahar range, positions for the cameras that take pictures of wild animals are being considered. The model receives video and image data taken by the cameras, which it then uses to detect instances of wild animals crossing roadways or entering human habitation or agricultural regions. The model has a high degree of accuracy when it comes to identifying wild animals such as elephants, deer, tigers, and other similar species. Due to the fact that this is an automated system, the model has the capability of eliminating and replacing the manual monitoring system. As a result, the system will become very helpful for conservation efforts as well as for the community. When the system identifies the presence of any wild animal, the information may be shared with forest officials as well as the general public so that appropriate safety measures can be taken. Wild elephants are responsible for the destruction of a significant quantity of crops and rice fields each year in the region surrounding the KNP as well as throughout the state of Assam. The implementation of an automated system that is based on AI, as proposed in this work, might prevent something like this from happening. The application of this paradigm can remove or significantly reduce the risk that humans pose to biodiversity, which is caused by the conflict that arises between humans and wild animals. Because it serves to limit the number of interactions between humans and wild animals, this system may contribute to the sustainable growth of the KNP. The conflict between people and wild animals can increase the risk of disease transfer from wild animals to humans. Therefore, there is hope that the suggested system will be able to stop the spread of disease from wild animals to people. Enhancing human development and decreasing the number of conflicts between humans and wild animals are two ways that societal morals might be raised.

When such a system is put into place, forest officials are not obliged to remain along the boundary of the KNP the entire time—something that is not always possible—in order to monitor the movement of wild animals continuously. This is because the system allows them to do so from a central location. They have the option to attend instead when told by the system. It has the potential to go a long way towards easing the coexistence of humans and the natural world by reducing the occurrence of stressful circumstances.

The manual processes are costly, time-consuming, and prone to error. Systems that are automated and assisted by artificial intelligence, such as the one that is proposed in this work, are beneficial in guaranteeing the safety of both humans and animals. The use of such a technology in elephant corridors that span railway tracks is another beneficial application for it. There have been a significant number of reports of elephant deaths in Assam and throughout India as a result of the animals crossing the railway tracks. If this type of AI-assisted technology is combined with the railway communication and control system, it has the potential to save a great number of lives.

6. Conclusion

Here, an animal detection system is developed using the YOLOv5 and SENet attention model for auto alarm generation to reduce human-animal conflicts and save wild animals from road accidents. The system successfully detects animals and classifies them into seven classes accurately, while preventing human-animal conflicts despite illumination, background, distance, and size variations. The model has been trained from scratch with a large public dataset of wild animals and tested using video feeds and images from four different locations along the NH-37 that pass through the KNP. The system is capable of detecting elephants, deer, tigers, etc., from both still images and videos with high accuracy; up to 96%, even multiple occurrences of such animals are detected and even at night time also. The system, since it is trained from scratch on a custom dataset, is also capable of handling augmented images by slight rotation, changes in contrast, etc. Attempts to improve the accuracy by training and testing the model for different epoch values viz. 100, 150, 200, and 250, respectively, resulted in significant performance improvement. At 250 epochs, the proposed system offered a mean average precision (mAP) of about 0.96 for some images, which is the same as obtained by the YOLO-Z model. The system that is being offered is set up to avoid conflicts between humans and animals, as well as to identify a wide variety of animals even when the lighting, distance, and background are all different, while also providing for the automatic generation of alarms. If this type of AI-assisted technology is combined with the railway locomotive management system, a great number of lives of elephants and other animals might be saved.

Future work

It is very important for wild animal detection systems to respond in time to avoid human-animal conflict. There is scope for speeding up the wild animal detection mechanism of the proposed method as well as there is future scope for speeding up the transmission of the alarm signal of wild animal detection. The challenges like the position of the camera for optimal direction and distance from objects, and illumination, especially for night vision, are to be addressed properly for better recognition of the wild animals. The improvement of the accuracy of the model may be explored by modifying the model architecture. An extended version of the system can be made a part of wildlife sanctuary management systems giving the capability of real-time monitoring and effective surveillance for preservation of the flora and fauna.

CRediT authorship contribution statement

Bijuphukan Bhagabati: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Kandarpa Kumar Sarma:** Conceptualization, Project administration, Supervision, Writing – review & editing. **Kanak Chandra Bora:** Conceptualization, Investigation, Visualization, Writing – review &

editing.

Declaration of Competing Interest

None.

Data availability

The authors are unable or have chosen not to specify which data has been used.

References

Ardovini, A., Cinque, L., Sangineto, E., 2008. Identifying elephant photos by multi-curve matching. Pattern Recogn. 41, 1867–1877.

Arshad, U., 2021. Object detection in last decade - A survey. Sci. J. Inform. 8 (1), 60–70. https://doi.org/10.15294/sji.v8i1.28956.

Banupriya, N., Saranya, S., Swaminathan, R., 2020. Animal detection using deep learning algorithm. J. Critic. Rev. 7, 434–439.

Benjumea, A., Teeti, I., Cuzzolin, F., Bradley, A., 2021. YOLO-Z: improving small object detection in YOLOv5 for autonomous vehicles, arXiv: 2112.11798.2021, Dec 2022.

Bhagabati, B., Sarma, K., 2016. Application of face recognition techniques in video for biometric security: A review of basic methods and emerging trends. Handbook Res. Modern Cryptogr. Solutions Comp. Cyber Security 460–478.

Bhagabati, B., Sarma, K.K., 2022a. A study on significant progress in face recognition and its related techniques towards better achievement for various applications. Intl. Conf. Emerg. Elect. Automat. E2A.

Bhagabati, B., Sarma, K.K., 2022b. Masked or unmasked face detection from online video using learning aided pattern recognition method. In: 2022 Second Intl. Conf. on Computer Science, Engineering and Applications (ICCSEA), pp. 1–4.

Bienstock, D., Muñoz, G., Pokutta, S., 2023. Principled deep neural network training through linear programming. In: Discrete Optimization, Vol 49. https://doi.org/10.1016/j.disopt.2023.100795. ISSN 1572-5286.

Birdlife International Organization Portal. IBA Conservation Status for Kaziranga National Park. http://datazone.birdlife.org/site/factsheet/kaziranga-national-park-iba-india.

Bottou, L., 2012. Stochastic gradient descent tricks. In: Montavon, G., Orr, G.B., Müller, K.R. (Eds.), Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science, vol. 7700. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35289-8_25.

Cao, Y., Bai, Y., Pang, R., Liu, B., Zhang, K., 2023. Vehicle detection algorithm based on background features assistance in remote sensing. Sens. Mater. 35 (2), 607–621.

Caron, M., Touvron, H., Misra, I., J'egou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging properties in self-supervised vision transformers. In: Proc. IEEE Int. Conf. Comput. Vis, pp. 9630–9640.

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J., 2021. Repvgg: making vgg-style convnets great again. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13733–13742. https://doi.org/10.1109/CVPR46437.2021.01352.

Dlužnevskij, D., Stefanovič, P., Ramanauskaitė, S., 2021. Investigation of YOLOv5 efficiency in iPhone supported systems. Baltic J. Modern Comp. 9 https://doi.org/10.22364/bjmc.2021.9.3.07.

Forest Department Website, Government of Assam. Various information on Kaziranga National Park. https://forest.assam.gov.in/portlets/national-park#kar (accessed 24 September 2023).

Ghosh, S., Varakantham, P., Bhatkhande, A., Ahmad, T., Andheria, A., Li, W., Taneja, A., Thakkar, D., Tambe, M., 2022. Facilitating human-wildlife cohabitation through conflict prediction. Proc. AAAI Conf. Artif. Intell. 36 (11), 12496–12502.

Gogoi, J., Hira, B., 2020. Issues and Challenges of Sustainable Tourism Development: A Case Study of Kaziranga National Park of Assam, India.

Guo, M.H., Xu, T.X., Liu, J.J., et al., 2022. Attention mechanisms in computer vision: A survey. Comp. Visual Media 8, 331–368. https://doi.org/10.1007/s41095-022-0271-y.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37, 1904–1916.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hosang, J., Benenson, R., Schiele, B., 2017. Learning non-maximum suppression. arXiv.

Hu, J., Shen, L., Albanie, S., et al., 2019. Squeeze-and-excitation networks, arXiv: 1709.01507v4 [cs.CV] 16 May 2019.

Khalajzadeh, H., Manthouri, M., Teshnehlab, M., 2014. Face recognition using convolutional neural network and simple logistics classifier. Adv. Intell. Syst. Comp. 223, 197–207.

Körschens, M., Denzler, J., 2019. ELPephants: A fine-grained Dataset for elephant re-identification. In: 2019 IEEE/CVF international conference on computer vision workshop (ICCVW), pp. 263–270.

Loos, A., Ernst, A., 2013. An automated chimpanzee identification system using face detection and recognition (cvpr). EURASIP J. Image Video Proc. 2013, 49.

Loos, A., Pfitzer, M., Aporius, L., 2011. Identification of great apes using face recognition. In: In 19th European Signal Processing Conference. IEEE, pp. 922–926.

Maxwell, A.E., Warner, T.A., Guillén, L.A., 2021. Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—part 1: literature review. Remote Sens. 13, 2450. https://doi.org/10.3390/rs13132450.

Medhi, S., 2020. Conflict and compensation in protected areas: A case study of Kaziranga National Park, Assam. E-J. Indian Social Soc. 4 (1), 119–134.

Nakada, M., Han, W., Demetri, T., 2017. AcFR: active face recognition using convolutional neural networks. Proc. IEEE Conf. Comp. Vision Pattern Recog. Workshops 35–40.

Nepal, U., Eslamiat, H., 2022. Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. In Sensors 22, 464. https://doi.org/10.3390/s22020464.

Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62. ISSN 0925-2312,. https://doi.org/10.1016/j.neucom.2021.03.091.

Premarathna, K.S.P., Rathnayaka, R.M.K.T., 2020. CNN based image detection system for elephant directions to reduce human-elephant conflict. In: 13th Intl. Research Conf., General Sir John Kotelawala Defence University, p. 591.

Pretorius, A.M., Barnard, E., Davel, M.H., 2019. ReLU and sigmoidal activation functions. In: In Fundamentals of Artificial Intelligence Research. https://api.semanticscholar.org/CorpusID:211073632.

Project, A.H, 2009. Living with Elephants in Assam: A Handbook. North of England Zoological Society, Guwahati, Assam, India; UK.

Ramaiah, N.P., Ijjina, E.P., Mohan, C.K., 2015. Illumination invariant face recognition using convolutional neural networks. In: 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), pp. 1–4.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Roboflow Dataset. Animal2 Image Dataset page on Roboflow. https://universe.roboflow.com/m-qm8tv/animal2-30h0a/dataset/1 (accessed 24 September, 2023).

Schroff, F., Kalenichenko, D., Philbin, J., 2015. Facenet: A unified embedding for recognition and clustering. In: 28th IEEE Conference on Computer Vision and Pattern Recognition.

Sharma, P., Chettri, N., Wangchuk, K., 2021. Human-wildlife conflict in the roof of the world: understanding multidimensional perspectives through a systematic review. Ecol. Evol. 11, 11569–11586.

Taskiran, M., Kashraman, N., Edem, G.E., 2020. Face recognition: past, present and future (a review). Digit. Signal Proc. 106.

The Deccan Herald. News on the killing of elephants in Assam. News published on 18th October, 2022. https://www.deccanherald.com/india/electrocution-poisoning-and-train-hits-the-major-causes-of-elephant-deaths-in-assam-1154795.html (accessed 24 September 2023).

The Deccan Herald. News on the killing of wild animals by speeding vehicles. News published on 9th June, 2022. https://www.deccanherald.com/india/at-kaziranga-if-your-speeding-kills-or-injures-animals-be-ready-to-pay-rs-5k-1116769.html (accessed 24 September 2023).

Tuia, D., Kellenberger, B., Beery, S., 2022. Perspectives in machine learning for wildlife conservation. Nat. Commun. 13, 792.

UNESCO World Heritage Convention List. Kaziranga National Park in UNESCO WHC List. https://whc.unesco.org/en/list/337/ (accessed on 24 September 2023).

Wang, W., Xie, E., Song, X., Zang, Y., Wang, W., Lu, T., Yu, G., Shen, C., 2019. Efficient and Accurate Arbitrary-Shaped Text Detection with Pixel Aggregation Network. arXiv.

Xu, R., Lin, H., Lu, K., Cao, L., Liu, Y., 2021. A forest fire detection system based on ensemble learning. Forests. 12, 217. https://doi.org/10.3390/f12020217.

Yadav, P.K., Thomasson, J.A., Searcy, S.W., Hardin, R.G., Braga-Neto, U., Popescu, S.C., Martin, D.E., Rodriguez, R., Meza, K., Enciso, J., Diaz, J.S., Wang, T., 2022. Assessing the performance of YOLOv5 algorithm for detecting volunteer cotton plants in corn fields at three different growth stages. In: Artificial Intelligence in Agriculture, vol 6, pp. 292–303. https://doi.org/10.1016/j.aiia.2022.11.005. ISST 2589-7217.

Yuvaraj, M., Antonio, M., Dimitrios, M., Hermilo, H., Miltiadis, A., 2022. Intelligent system utilizing HOG and CNN for thermal image-based detection of wild animals in nocturnal periods for vehicle safety. Appl. Artif. Intell. 36, 1. https://doi.org/10.1080/08839514.2022.2031825.

Zhang, M., Gao, F., Yang, W., Zhang, H., 2023. Real-time target detection system for animals based on self-attention improvement and feature extraction optimization. Appl. Sci. 2023.

Zhou, X., Girdhar, R., Joulin, A., Krähenbühl, P., Misra, I., 2022. Detecting twenty-thousand classes using image-level supervision. In: Proc. Eur. Conf. Comput. Vis, pp. 350–368.

Zhu, L., Geng, X., Li, Z., Liu, C., 2021. Improving YOLOv5 with attention mechanism for detecting boulders from planetary images. Remote Sens. 13, 3776. https://doi.org/10.3390/rs13183776.