IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 2 May 2023, accepted 15 June 2023, date of publication 26 June 2023, date of current version 10 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3289586

RESEARCH ARTICLE

Creating Alert Messages Based on Wild Animal Activity Detection Using Hybrid Deep Neural Networks

B. NATARAJAN[1], R. ELAKKIYA[2], R. BHUVANESWARI[1], (Member, IEEE), KASHIF SALEEM[3], DHARMINDER CHAUDHARY[1], AND SYED HUSAIN SAMSUDEEN[4]

[1]Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai 601103, India
[2]Department of Computer Science, Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai International Academic City, Dubai, United Arab Emirates
[3]Department of Computer Sciences and Engineering, College of Applied Studies and Community Service, King Saud University, Riyadh 11362, Saudi Arabia
[4]Department of Electronics and Communication Engineering, K. Ramakrishnan College of Engineering, Samayapuram, Tiruchirappalli, Tamil Nadu 621112, India

Corresponding author: Kashif Saleem (ksaleem@ksu.edu.sa)

This work was supported by the King Saud University, Riyadh, Saudi Arabia, under researcher supporting project RSPD2023R697.

ABSTRACT The issue of animal attacks is increasingly concerning for rural populations and forestry workers. To track the movement of wild animals, surveillance cameras and drones are often employed. However, an efficient model is required to detect the animal type, monitor its locomotion and provide its location information. Alert messages can then be sent to ensure the safety of people and foresters. While computer vision and machine learning-based approaches are frequently used for animal detection, they are often expensive and complex, making it difficult to achieve satisfactory results. This paper presents a Hybrid Visual Geometry Group (VGG)$-19+$ Bidirectional Long Short-Term Memory (Bi-LSTM) network to detect animals and generate alerts based on their activity. These alerts are sent to the local forest office as a Short Message Service (SMS) to allow for immediate response. The proposed model exhibits great improvements in model performance, with an average classification accuracy of 98%, a mean Average Precision (mAP) of 77.2%, and a Frame Per Second (FPS) of 170. The model was tested both qualitatively and quantitatively using 40, 000 images from three different benchmark datasets with 25 classes and achieved a mean accuracy and precision of above 98%. This model is a reliable solution for providing accurate animal-based information and protecting human lives.

INDEX TERMS Animal detection, VGG-Net, Bi-LSTM, convolutional neural network, activity recognition, video surveillance, wild animal monitoring, alert system.

I. INTRODUCTION

In general, animal activity detection creates numerous challenges for researchers due to the continuous streaming of inputs and the cluttered backgrounds. There are huge varieties of wildlife categories with different facial, nose, body, and tail structures. The detection and classification of such animals in video sequences and the processing of huge feature maps demand the need to develop a robust framework.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhongyi Guo.

Such developments in real-time cases need large-scale video data for training and testing purposes and high GGPU-based computing resources. Moreover, the incorporating techniques should handle the data in an intelligent way to produce plausible results. Hence, there is a high demand for developing such a model to detect animal activities in forest regions. Although numerous advancements have been made in this technological era, research in this area still seeks higher attention to produce a strong model. With this work, we can save humans from sudden animal attacks as well as send alert messages with location information to the forest officers for quick action. These systems offer better monitoring services and help to find the activities of animals and detect if there is any hunting by humans or hindrance to wildlife. These clusters of activities, such as tracking the animal object and finding its activity and generating the alert messages, pose huge complexity in the Deep Learning area. Research on this work, investigates the advancements in video analysis techniques and complex neural network-based architectures. Recent developments in Deep Learning techniques have produced impressive results in image recognition, classification, and generation tasks [1]. Due to these developments, we focus our aim on developing a robust model for monitoring the activities of animals and generating alerts to the forest officers in case of any abnormal activity such as hunting, animals entering into human living areas or agricultural land. The development of the proposed model investigates this problem from multiple angles to provide a better solution.

Object detection techniques play a vital role in understanding the components of images and their associated relationships. In the case of videos, it provides the movement and activity-based details explicitly. The conventional methods use hand-crafted mechanisms [2], [3] for feature extractions and produce tangible results. The development of deep learning models handles this task in an efficient way to reduce the overheads present in earlier studies. Earlier works use traditional machine learning methods to detect objects, but they become stuck when confronted with complex datasets and multimodal inputs. The deep model handles the features of an image effectively to explore the finely tuned investigation on pixels and combines the relevant features to construct feature maps. Feature maps help to predict the patterns, shapes, edges, and contours of objects and learn the structure of objects easily without any manual interventions. Deep learning models are designed to handle such complex data structures and scale large volumes of data. The hyperparameter optimization techniques and regularisation methods regulate the deep neural network performance to produce high accuracy results. Generally, the object detection mechanisms are applied in diverse fields, such as face detection [4], [5], scene understanding [6], and salient object detection [7], [8].

The research studies about animal activity detection are still in their infancy levels. The earlier approaches need to be upgraded and fine tuned to produce plausible results. We have used four different benchmark datasets, which total 40K images. The proposed model has been evaluated qualitatively and quantitatively using reasonable sizes of images and quality metrics. On the other hand, it is obvious that integrating different deep networks in a hybrid way adds additional complexity overhead to the model development. However, the successful implementation of such models produces unimaginable results over a combination of object detection and class prediction tasks [9]. With this motivation, we propose the novel approach termed "hybrid VGG−19+Bi-LSTM networks" to detect the animal activities and create alert messages in case any problem occurs. A novel network is proposed to detect the activities of diverse categories of animals simultaneously, monitoring the locomotion of animals in forest regions and dark areas [10]. The proposed approach uses VGG−19 pre-trained networks to classify the type of animal, and the Bi-LSTM network creates text based alert messages with location information. The surveillance and night vision camera based videos consist of spatial and temporal dynamics. The VGG−19 networks deal with spatial information, and BI-LSTM recurrent networks effectively handle the temporal details [11].

Experimental results are also demonstrated to compare the proposed approach with earlier methods and explore the valid justification results. The details of various levels of development are explained clearly and exhibit the quality of our work [12]. In object detection and classification models, there are huge complexities in finding the expected results. In large scale scenarios, the model performance bottleneck results in low performance and degrades the entire development process. The earlier studies handled these scenarios using a wider range of mechanisms [13]. Although the models produce significant improvements in accuracy, they fail to perform well in testing phases.

The contributions and objectives of the proposed techniques are listed as follows:

1) The proposed Hybrid VGG-19+Bi-LSTM model is built using deep neural networks with fine- tuned hyperparameters to yield greater recognition accuracy results.
2) The proposed model aims to achieve outstanding classification results by incorporating novel hybrid approaches.
3) 'The proposed system offers foresters more accurate prediction performance about animal detection and also supports them with faster alert services via SMS.

The further sections of the paper are arranged as follows: Section II discusses related works and identifies shortcomings in previous developments, while Section III describes the proposed VGG−19+Bi-LSTM system architecture and implementation details. Section IV presents the experimental results of the proposed model evaluation for four different benchmark datasets. Section V concludes the summary of the entire work and its future scope in a wider range of applications.

II. RELATED WORK

The author Zhang et al. proposed wild animal detection using a multi-level graph cut approach for investigating spatial details and a cross-frame temporal patch verification technique for temporal details. The model analyzes the foreground and background details of the camera trap videos. This approach uses a Camera trap and Change Detection net dataset for segmenting the animal object from natural scenes based on cluttered background videos. Although the model produces a high detection rate, fails to perform well in detecting crucial details like location details, and human interruptions. The author [14] proposed animal detection using Convolutional Neural Network (CNN), and the author proposed animal detection using Iterative Embedded Graph Cut (IEGC) techniques to form regions over images and DeepCNN features and machine learning classification algorithms [15] for classification purposes. Although these models verify the extracted patches are background or animal, still need improvements in classification performance. Object Detection using deep learning methods attained new heights in computer vision applications. The detection of objects present in images or videos by using object localization and classification techniques gives higher support in detecting various objects present in an image or video. From the extracted results, we can count the number of objects and their activity. This technique is highly used in video surveillance and security-based applications, tracking objects in hidden boxes, monitoring fraudulent activity in public and crowded areas, traffic monitoring and identification of vehicle theft, vehicle number plate recognition, and Object Character Recognition (OCR) [16].

This paper aims to identify the movements of animals around forest space, provides alert information to the forest officers in case of hunting, crossing the forest lines, any hindrance to villagers and tourists people, and detection of trespassing activity. The development of various methods for employing object detection in different environments and diverse applications shows the progress and importance of object detection in research fields and gained more attention. Moreover, further research works in this area provide useful insights into numerous applications and construct powerful frameworks for detecting objects in different scenarios. The Fast R-CNN techniques [17] are widely used for object detection due to their high accuracy and improved training performance. The introduction of the Faster R-CNN technique [18] rapidly improves the detection performance of the model by employing full image-based convolution features and region-based networks. The Histogram of Oriented Gradients (HOG) feature descriptors [19] uses the Region of Interest (ROI) techniques to identify the objects faster than earlier approaches. The conventional R-CNN technique [20] introduces efficient detection methods by incorporating region proposal networks and ConvNet. This method detects the thousands of object classes in an image or video using annotated information. The R-CNN techniques do not use any approximation techniques and hashing methods for predicting the object regions. R-FCN techniques [21] use weighted full convolution layers to detect object's region and finds ROI to detect the category of objects and its background details.Object detection techniques also sounds good with the help of deep learning techniques in the field of autonomous vehicles [22] and traffic scene object detection [23] also.

The Single Short Detector (SSD) methodology [24] uses bounding boxes based discretization techniques to effectively handle feature map information and large volume data. The Spatial Pyramid Pooling (SPP-net) [25] computes the feature maps in single computations and provides high robustness to the object detection tasks using sub-region-based fixed-length representations. The You Only Look Once (YOLO) architecture achieves faster results by processing 155 frames per second in real-time cases. This technique uses an end-to-end approach to detect the objects using regression and probabilistic computations instead of considering classification approaches and produces remarkable results in object detection with a lower false-positive rate. The detailed investigation is done by the researchers with respect to background subtraction and elimination. The authors used different approaches to detect the background details such as estimating multiple hypotheses, non-parametric model [26], and global statistic-based methods [27], background cut [28]. The inducement of principal component analysis and regression-based models [29], eigen approaches [30], and weightless neural network [31] explores the pixel level analysis and manages the sensitivity score. Recent studies [32], [33] show the outperformance of DCNN networks in object detection, recognition, and classification tasks. To improve the speed of the method the authors [34], [35] proposed fast detection approaches and regression frameworks [36] to plot bounding boxes over the image. The author [37] proposed a fully connected layer-based model to estimate the box coordinate values used for object localization tasks by considering the single object. Later, the model is revised to detect multiple objects by replacing the convolutional layers. The MultiBox [38], approach plots the multiple bounding boxes over the images to detect the different objects present in a single image.

III. PROPOSED SYSTEM

The system architecture of the proposed hybrid VGG−19+ Bi-LSTM model is demonstrated in figure 1. The proposed architecture comprises five phases of development steps, which includes data pre-processing, animal detection, VGG−19 pre-trained model-based classification, extracting the prediction results, and sending alert messages. In the data pre-processing phase, 45k animal images were collected from different datasets such as camera trap, wild animal, and the hoofed animal dataset. The collected images were rescaled to the size of 224×224 pixels and denoised. In the second phase, we pass the pre-processed images into YOLOR object detection model [39], which identifies the animal present in an image using bounding boxes as illustrated in Fig. 4. In the third phase, using hybrid VGG−19+Bi-LSTM model we perform image classification tasks and class label prediction was done and animal details are extracted using LSTM Networks. In the fourth phase, we collect the location information of the animal, and the web server creates a SMS alert and sends it to the forest officers. Finally, remedial action will be taken by the forest officers to save the animals and human lives.

In this phase, we have collected animal images of diverse categories from four different datasets, such as the camera trap dataset [40], the wild anim dataset [41], the hoofed animal dataset [42], and the CDnet dataset [44]. Totally, we collected $45k$ animal images and resized them to 224×224 pixels. Furthermore, we applied denoising over the rescaled images.

The VGG$-$19 pre-trained CNN framework consists of 47 layers total and takes the input of 224×224 pixel based images. It has 16 convolution layers that perform convolution operations and 3 fully connected layers to yield the desired results. The model applies ReLU activation functions to achieve non-linearity and performs Max pooling operations to get the maximum value from the pixel window. The drop out value is set as 0.5 and finally, the softmax layer predicts the probability of class output.

A. THE OBJECT DETECTION-BASED CONVOLUTIONAL NEURAL NETWORKS

The CNN is a very famous deep learning technique, mainly used to perform image recognition, image classification [43], [44], [45], and object detection tasks. CNN networks used real-time applications such as Facebook Face Detection, Google image search, and Amazon product recommendations. The evolution of CNN networks attained various changes in its architecture to produce fantabulous results in image classification tasks. The CNN model basically comprises of three layers, such as the convolutional layer, pooling layer, and fully connected layers.

1) THE CONVOLUTIONAL LAYER

The convolutional layers are an important component in CNN networks. This layer extracts the different features of images, like shapes, contours, edges, and corners. The deeper networks learn the image information very well and produce high accuracy over classification and prediction tasks. The convolution layers perform convolution operations (multiplication) on images using filters. Each time it chooses the limited portions of images and applies convolution to them. The size of the filters is 2×2 or 3×3, depending on the image size. The processing of RGB images uses small height and width but the same depth based filters. The output of this layer contains reduced dimensions of images known as 'feature maps'.

$$s(t) = (x * w)(t) \tag{1}$$

The convolution operation is carried out using the equation 1, where x represents the original image pixel values, t represents time, w represents the filter values, and $s(t)$ represents the output value.The convolution operations are depicted visually in Fig 2.

2) THE POOLING LAYER

The pooling layers reduce the image size drastically and aid in focusing on the dominant features alone. The pooling operations are classified as max pooling, average pooling, and min pooling. The technique for max-pooling operations chooses the maximum value out of the pooling region, the average pooling operations averages the values lying in the pooling area; and the minimum value for min pooling operations. Mostly, the network uses max pooling operations to predict the image structures using important pixels, for example, the recognition of handwritten digits. The detailed illustration example is shown in Fig 3

3) THE FULLY CONNECTED LAYER

The inclusion of layers such as convolution and pooling often in the network helps to flatten the desired features that can be easily processed by fully connected layers. The classification of images is done in this phase, and the model learns the features present in an image deeply and compares the differences between two images. The use of ReLU activation functions provides non-linearity to the model. The weight and bias values are tuned properly to achieve better results. The optimizers are chosen carefully depending on the application suit to produce high accuracy. The use of regularization techniques facilitates the better convergence of the model using the best learning rate and dropout values. These steps make the model learn perfectly and achieve the target function in a faster way.

B. CLASS LABEL PREDICTION AND CREATING THE ALERT MESSAGES

The CNN network produces the class prediction results, which are further processed by the LSTM networks to identify the activity of the animal and create alert messages. The LSTM networks are variants of RNN that produce impressive results in handling sequential data, time series analysis, weather forecasting, image captioning, and text generation tasks. The LSTM networks use a gating mechanism to process the longer sequences in a recursive way and memorize the past information.

The input layers process the present input information and pass it to the network. The network uses cell states and memory cells to keep past information. The forget gate decides the need for particular information and decides the removal of unwanted information. The Bi-LSTM networks are extended versions of LSTM networks that have proved an efficient model for handling textual information. It applies forward and backward approaches and has the capability to read the contents of encoder units from hidden layers of decoder units to produce plausible results in text generation tasks. To obtain accurate results, we employ an attention mechanism over encoder outputs.

The LSTM units are used to process sequence data applications such as longer sequence text processing, image captioning, time series data analysis, and text generation applications. LSTM network is proposed by the researchers [46], [47] to address the complexities of conventional RNN. With the help of LSTM networks, the problems such as handling longer sequences of sentences are addressed. It explores

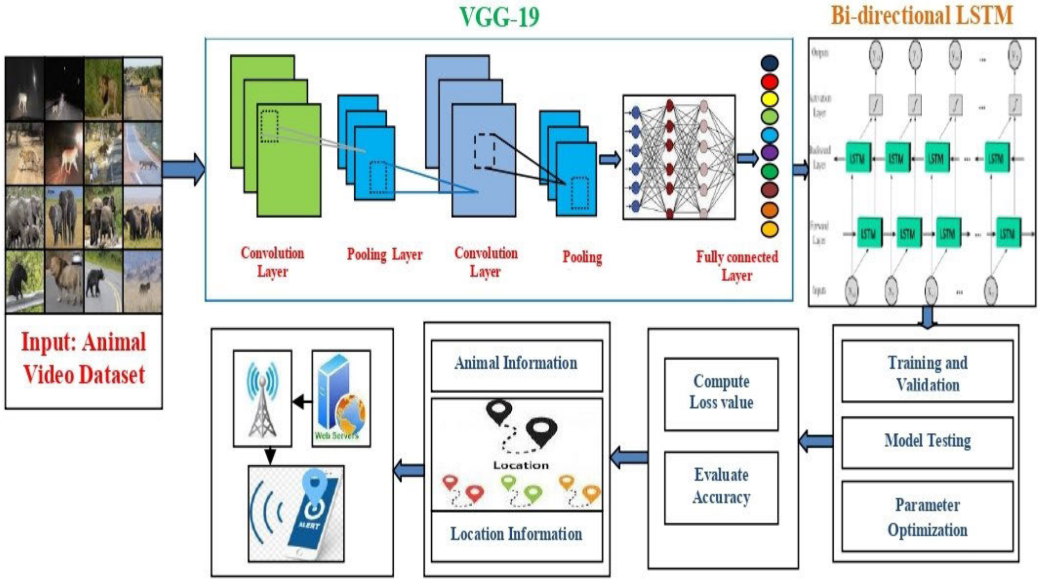

FIGURE 1. The system architecture of the proposed hybrid VGG-19+Bi-LSTM model.

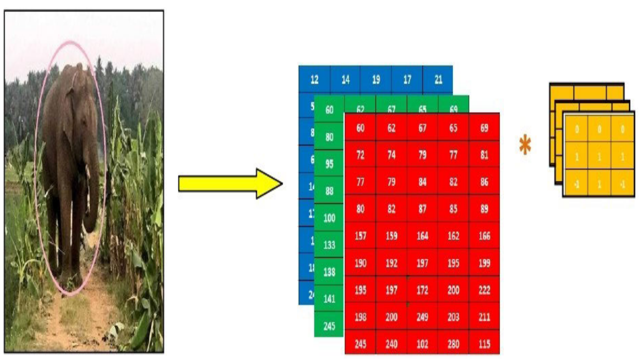

FIGURE 2. Convolution operation on Animal Images.

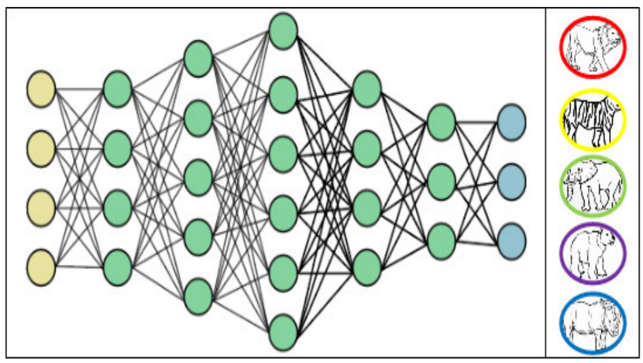

FIGURE 4. Fully connected layers.

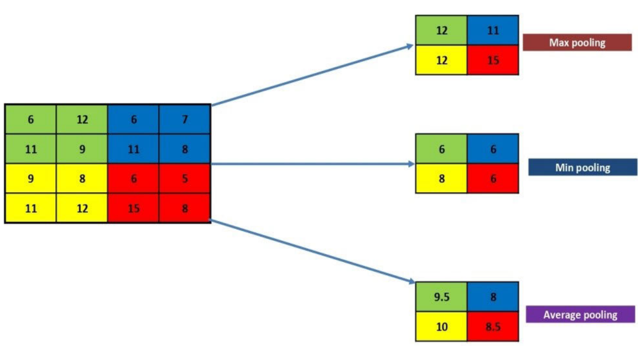

FIGURE 3. Pooling operations on feature map vectors.

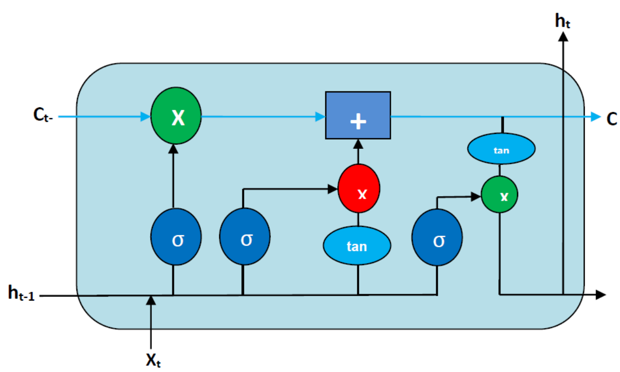

FIGURE 5. Architecture of LSTM Network.

the solutions to the vanishing gradient and exploding gradient issues by enforcing gated mechanisms and cell states. The three gates of LSTM networks are the input gate for processing the input, the forget gate to perform the removal of data which is no longer required, and the output gate to show the results. The input gate decides the current information to be feed into the network. The forget gate removes the unnecessary information content in memory. The output gate

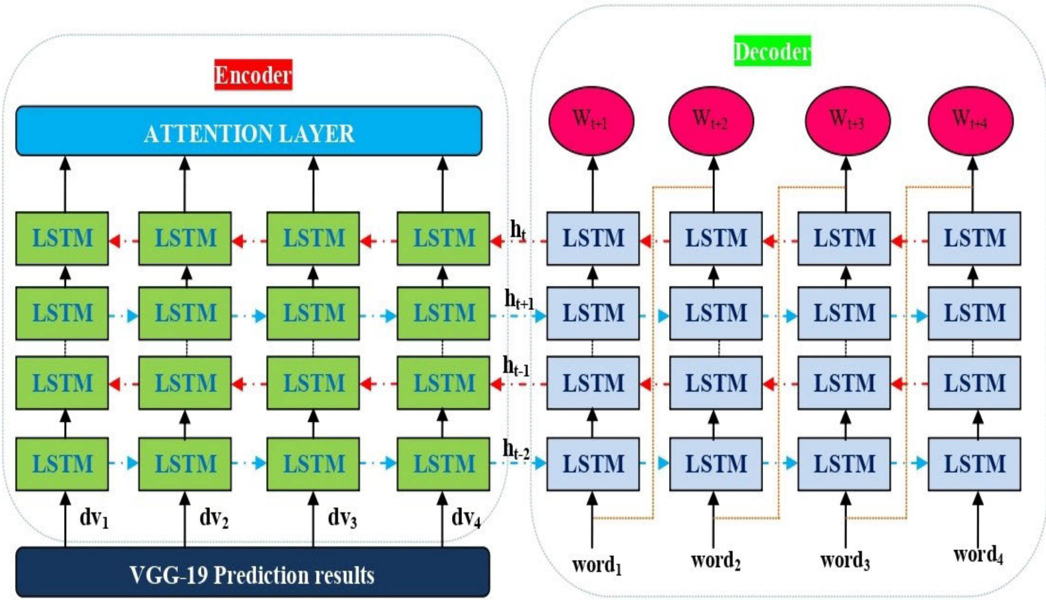

FIGURE 6. Bi-LSTM Network Architecture.

determines the final output to be generated from the network

$$s_t = \sigma_g(W_s k_t + V_s h_{t-1} + b_s) \tag{2}$$
$$m_t = \sigma_g(W_i k_t + v_m h_{t-1} + b_m) \tag{3}$$
$$out_t = \sigma_g(W_o k_t + v_o h_{t-1} + b_o ut) \tag{4}$$
$$j_t = s_t \circ j_{t-1} + m_t \circ \sigma_j(W_j k_t + V_j h_{t-1} + b_j) \tag{5}$$
$$h_t = out_t \circ \sigma_h(j_t) \tag{6}$$

The proposed VGG−19+Bi-LSTM Algorithm explains the detailed procedures of our work in each step and gives a clear overview of the execution procedures of our method. The VGG−19 model prediction results are further processed by Bi-LSTM networks as depicted in Fig. 6 to generate the text output. The LSTM units use separate memory cells. The cell states decide the information to be remembered or forgotten from the memory cells. For each cycle, the LSTM unit consumes the previous hidden state information ($h_t - 1$), current input (x_t), and previous cell state information (m_t). The weight values (W) and bias vectors are updated from model learning to produce good results. The hyperparameters such as learning rate values, number of epochs, and batch size values are tuned properly to get the right set of values to achieve extended performance. We use Adam optimizer and dropout regularization techniques to obtain greater results over the benchmark datasets.

C. THE DATASET

The proposed model was experimented with using four different benchmark datasets such as the camera trap dataset [48], wild animal dataset [49], the hoofed animal dataset [50] and the CDnet dataset [51]. The details of the datasets are stated as follows.

Algorithm 1 Proposed VGG-19+Bi-LSTM Algorithm

Input: Animal Video frames (AVF), Class labels (CL)
Output: SMS text generation(SMS(Activityk))
Procedure
Initialise the random values for weights and bias vectors
function VGG-19(AVF, CL)
    apply convolution (3x3) using the equation (7)
    Show the Feature Map results (FM)
    apply max pooling operation (Maxpool(FM))
    flatten the results using Fully connected Layers(FC)
compute softmax $S(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_i}}$ (7)
repeat
end function
if (Predicted output! = null) then
    for i=l to length(LM-l) do
        for all epochs do
            for all Stacked LSTM cells do
                for all hidden states do
                    for all time steps do
                        Bi-LSTM(ht)
                        update input gate(it)
                        update cell state (ct)
                        update forget gate(ft)
                        update output gate (ot)
                        repeat
                    end for
                end for
            end for
        end for
    end for
end if

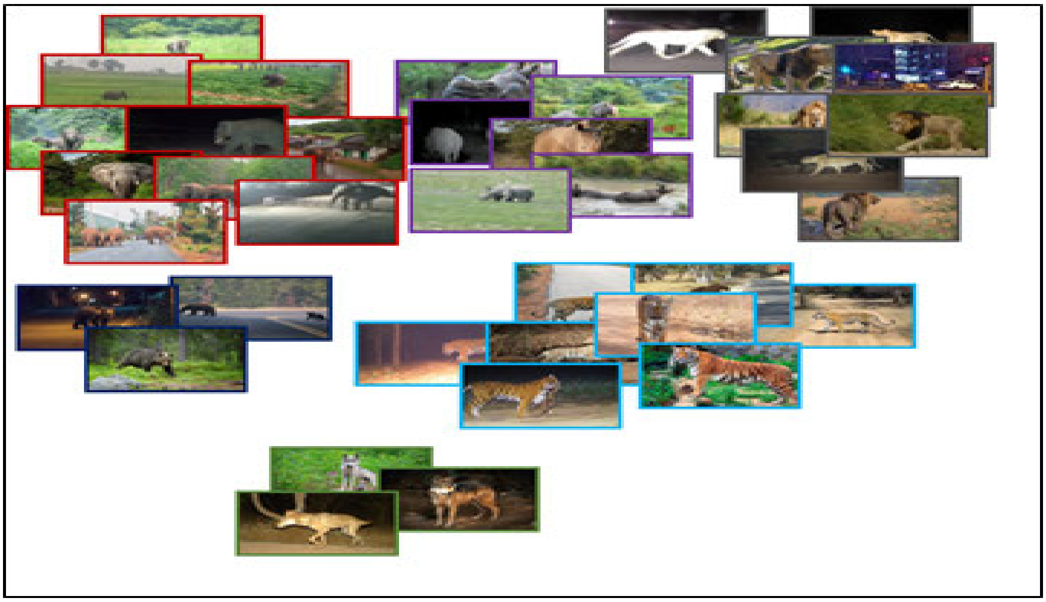

FIGURE 7. t-SNE visualization Images.

FIGURE 8. The analysis of pixel level details using Stacked Dense Flow Difference Image (SDFDI) Computation Method.

1) CAMERA TRAP DATASET

The Camera trap dataset was developed for animal detection and consists of videos of 23 classes of animals. It consists of 800 image sequences about animal movements and activities. All the images are properly labelled and contain bounding boxes. This dataset contains one million images and focuses on cluttered backgrounds.

2) WILD ANIM DATASET

The Wild Anim dataset consists of images collected from Flicker. It has over 5, 000 images of five classes of animals, such as Leopard, Wolf, Bear, Lion, and Elephant. The dataset has been fully labelled and contains RGB images of $250X250$ pixels. This dataset introduces the anamorphic distortions slightly.

3) THE HOOFED ANIMAL DATASET

The hoofed animal dataset consists of 200 images of 6 classes of animals. It consists of images of deer, horses, goats, camels, sheep, and cows. This dataset was developed to provide assistance with object detection, recognition, and segmenting the exact boundaries of animal patches.

4) CDnet DATASET

The CDnet dataset has been developed for providing change detection activity-based video sequences. This corpus consists of 90, 000 frames and 31 categories of video classes. It covers six variants and two modalities. The modalities are thermal, IR, and color. We considered the following categories in our work, such as dynamic background, camera jitter, and shadows.

5) DATASET VISUALIZATION

The multi-class based datasets are visualised using t-SNE algorithms, which help to visualize the different features of the complex data set as well as understand the underlying structure of datasets at different thresholds. The t-Distributed

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
fc1 (Dense)	(None, 4096)	102764544
fc2 (Dense)	(None, 4096)	16781312
predictions (Dense)	(None, 1000)	4097000

FIGURE 9. **The Feature map results produced by VGG-19 layers for wild anim dataset.**

Stochastic Neighbor Embedding (t-SNE) algorithms project the dataset into two or multi-dimensional view points. The conventional SNE algorithm converts the estimated Euclidean distance values for high dimensional datasets into conditional probability values to explore the similarities among the data points.

$$pk|j = \frac{exp(-\frac{\|m_j - m_k\|^2}{2\sigma_i^2})}{\sum_{l \neq k} exp(-\frac{\|m_j - m_l\|^2}{2\sigma_i^2})} \tag{7}$$

The estimation of probability values for comparing the scores between one image and another. These comparison scores yield the similarity among the data points in high dimensional space. It uses the student t-distribution metric to evaluate the relationship between the data points. The student t-distribution metric is computed using the equation 7.

$$t_{dist} = \frac{\bar{y} - \mu}{SD/\sqrt{n}} \tag{8}$$

In equation 8, the $\bar{y}$ represents the sample mean values, represents the mean value of the population. The Standard Deviation (SD) computes the variance among the various data points, and symbol n denotes the size of the data point. The estimation of low-dimensional features of the high-dimensional data points is computed using the condi-

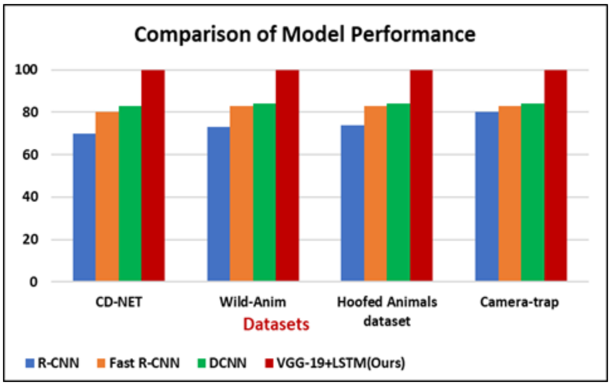

FIGURE 10. **The proposed hybrid VGG-19+Bi-LSTM model recognition accuracy comparison.**

tional probability as follows.

$$q_{k|j} = \frac{exp(-\left\|n_j - n_k\right\|^2)}{\sum_{l \neq k} exp(-\left\|n_j - n_k\right\|^2)} \tag{9}$$

In equation 9, the probability values $q_{k|j}$ use squared Gaussian functions and assign a value of zero to similar data points as $q_{j|j} = 0$. For comparing the results of two different probability computations, SNE uses KL divergence as the cost function represented in equation 10.

$$J = \sum_j KLD(P_j||Q_j) = \sum_j \sum_k p_{k|j} log \frac{P_{k|j}}{q_{k|j}} \tag{10}$$

The KL divergence provides the results of the low-dimensional and high-dimensional similarity based results using $P_{k|j}$ and $q_{k|j}$ values. Finally, the t-SNE visualization results are depicted in Fig 7.

D. ESTIMATION OF STACKED DENSE FLOW DIFFERENCE IMAGE (SDFDI)

We use Stacked Dense Flow Difference Image (SDFDI) methods to track the temporal details of the frame and identify the changes in animal activities. This technique explores the motion flow estimation in a sign video. It compares each frame and all the pixel values with the next frames to adopt a consistent flow using polynomial expansion methods. We used the Farneback method to evaluate the optical flow vectors.

$$SDFDI = \sum_{i=2}^{n-1} iX \left|C_i - C_{i-1}\right| \tag{11}$$

In equation 11, the variable C_i represents the stacking of horizontal, vertical, and magnitude. The results are shown in Fig. 8 which demonstrates the motion flows for the wild anim dataset.

E. TRAINING DETAILS

The proposed hybrid VGG−19+Bi-LSTM model was implemented using a Lenovao Think Station−P920 Tower workstation with dual Intel Xeon processors and it has three NVIDIA Quadro GPUs. The model handles 32k images for training purposes, 3k images for validation and 5k images for testing purposes. The experimental results demonstrate the outperformance of the proposed model compared to earlier approaches. We set the initial learning rate at 0.01, the dropout value at 0.05. The model runs for 1000 epochs with a batch size of 100.

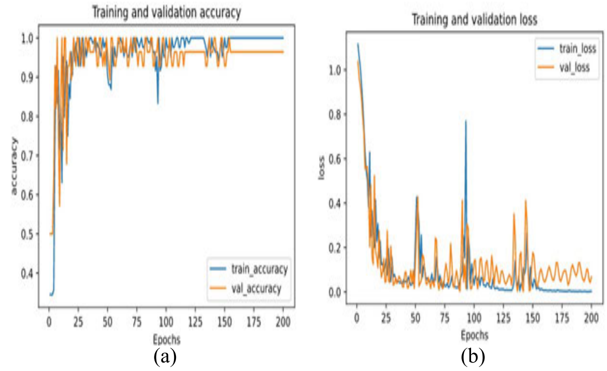

FIGURE 11. **(a) Computing the Accuracy values, (b) Loss estimation.**

The layer details clearly elaborate the sequences of layers present in the VGG−19 model. It comprises the 16 convolutional layers and 3 pooling layers. It applies ReLU activation functions to achieve the non-linearity. The intermediate results of layers called as feature maps provide shapes, edges, and contour details of the given image. The feature map output of animal detection based on our proposed work is illustrated Fig. 9. Furthermore, the integration of LSTM networks processes the class label prediction results and processes the details, capturing the location information and monitoring the activity of the animals. At the end, it creases the short messages and sends them to the forest officers through web servers and mobile networks. This approach drastically improves the deep learning performance in different domain based applications and focuses greater attention on security-based application development. The feature maps are the output of each layer and show the learning process of the network. At each layer, the network learns various features of an image, such as shapes, contours, and edges. These details develop the network to produce better prediction and classification results on new unseen data. We can visualize the feature maps after applying filters and pooling operations.

The proposed model has been evaluated using four different datasets, such as the camera trap dataset, the wild anim dataset, the hoofed animal dataset, and the CDnet dataset. The proposed model was trained using 32K training images, 3K validation images, and 5K testing images. The proposed model performance is evaluated using Mean Accuracy computation methods and F-Score Metrics. The results are listed out in table 1. The improved score values indicate the greater improvement of model performance over diverse domain datasets. In addition to that, we compared the proposed hybrid VGG−19+Bi-LSTM model performance using mean Average Precision (mAP), Frame per Second values.

IV. EXPERIMENTAL OUTCOMES AND DISCUSSIONS

This section, which shows the improved model performance compared with existing YOLO models such as Faster

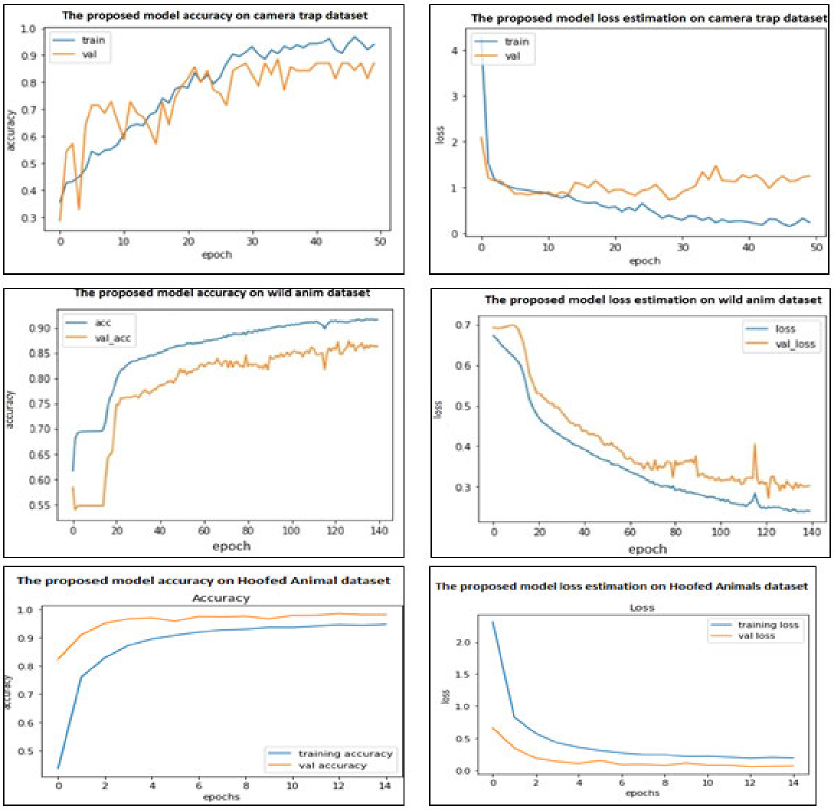

FIGURE 12. Training accuracy and validation accuracy analysis over the benchmark datasets – Camera trap dataset, wild anim dataset and Hoofted Animal dataset.

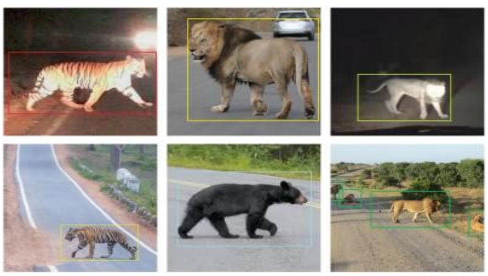

FIGURE 13. The animal detection using bounding box for camera trap dataset and wild anim dataset.

FIGURE 14. Contours of detected Animals of Hoofed Animals dataset.

R-CNN, Fast YOLO model, R-CNN, Fast R-CNN, and YOLOv5. We compared the model performance in terms of recognition accuracy with earlier developments such

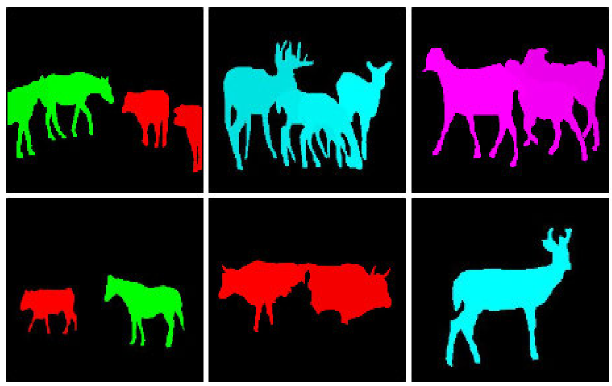

FIGURE 15. Segmentation results of detected Animals of Hoofed Animals dataset.

as R-CNN, Fast R-CNN, and DCNN using four different benchmark datasets-camera trap dataset, wild anim dataset, the hoofed animal dataset, and the CDnet dataset. The results are shown in the bar chart, which indicates the improved recognition accuracy of our model compared with earlier approaches

The proposed hybrid VGG−19+Bi-LSTM model achieves significant improvements in accuracy metrics during training and validation phases. We plotted the accuracy curve of the proposed model in Fig. 11(a) Moreover, the loss computations are performed to analyze the network performance and failure cases. We apply a mean squared error loss

FIGURE 16. Ellipse bounding box based animal detection.

TABLE 1. The proposed model evaluation using Mean Accuracy and F-Score.

	AlexNet		ResNet50		GoogleNet		DCNN		VGG16		VGG19+Bi-LSTM(proposed)	
Dataset/Model	Mean Accuracy score	F-Score Metric	Mean Accuracy score	F-Score Metric	Mean Accuracy score	F-Score Metric	Mean Accuracy score	F-Score Metric	Mean Accuracy score	F-Score Metric	Mean Accuracy score	F-Score Metric
Camera trap dataset (Zhang et al. 2016)	0.799	0.819	0.789	0.816	0.852	0.889	0.862	0.854	0.912	0.926	**1.000**	**1.000**
Wild anim dataset (Okafor et al. 2019)	0.802	0.813	0.761	0.825	0.826	0.825	0.851	0.885	0.911	0.896	**0.989**	**0.999**
The hoofed animal dataset (Ahuja et al. 2008)	0.825	0.841	0.792	0.839	0.855	0.863	0.823	0.871	0.901	0.898	**0.972**	**0.988**
Change Detection Dataset (Goyette et al. 2012)	0.866	0.821	0.813	0.809	0.893	0.852	0.875	0.862	0.912	0.914	**0.991**	**0.949**

TABLE 2. The proposed model performance comparison using mAP and FPS score metrics.

Model	mean Average Precision (mAP)	Frame Per Second
YOLO model	61.3%	70
R-CNN	70.2%	75
Faster R-CNN	71.6%	80
Fast R-CNN	70.3%	86
Fast YOLO model	69.5%	155
YOLOv5	74.3%	140
Hybrid VGG-19+Bi-LSTM (proposed)	**77.2%**	**170**

TABLE 3. Performance comparison of different models with proposed work.

Model	Precision value	Recall value	F1-Score	Accuracy
CNN	0.9552	0.8456	0.8256	0.9144
R-FCN	0.8451	0.9321	0.9758	0.9256
DCNN	0.9145	0.9456	0.8758	0.9412
R-CNN	0.9325	0.8236	0.9354	0.9012
SPP-net	0.9654	0.8452	0.8635	0.9214
YOLO	0.9125	0.9123	0.9415	0.9315
DCNN	0.8326	0.9254	0.9444	0.8462
Faster R-CNN	0.9458	0.9456	0.9546	0.8263
DeepCNN	0.9652	0.9526	0.9642	0.9512
$\mathrm{VGG}-\mathbf{19}+\mathrm{Bi}-\mathrm{LSTM}$	**1.0000**	**0.9843**	**0.9758**	**0.9856**

function to estimate the loss values. Fig. 11 (b) illustrates the loss value computation of the proposed model. The rectangle based bounding boxes for animal detection in an image are depicted in Fig. 13. It shows the efficient way of detecting the animals and the capability of the proposed model. The proposed hybrid VGG−19+Bi-LSTM model

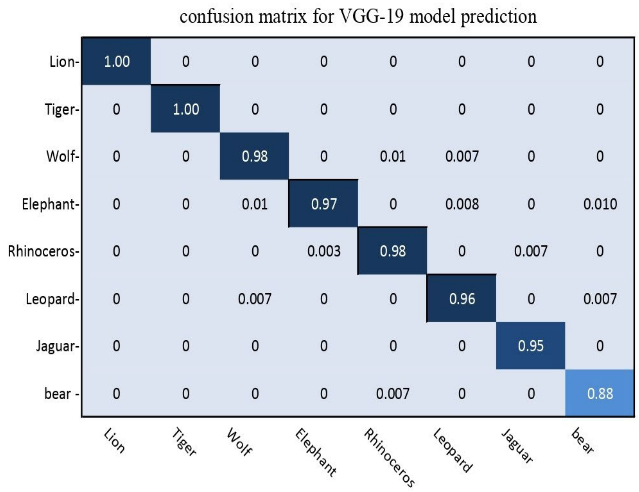

FIGURE 17. Confusion Matrix for evaluating the VGG-19+Bi-LSTM model performance.

detects the animal outlines in an image perfectly and provides the best detection results over different benchmark datasets. Figure 13 demonstrates the bounding box estimation results over the wild anim dataset and camera trap dataset. We further investigated the performance of the model for individual datasets. The results are shown in the following Fig. 13 which provides a detailed view of our analysis over different benchmark datasets such as the camera trap dataset, wild anim dataset, and Hoofed Animal dataset. Fig. 14 represents the contours of the animals which are detected from the Hoofed Animals dataset. And from the detected animals, segmentation was carried out which is used for identifying the types of animals which has been illustrated in Fig. 15.

The training and validation process in deep learning using complex neural networks aims to map the input to the output. This process regularly updates the weights and bias vectors. Based on such updates, the error value of the model gets reduced. We use a total of $32K$ images for testing the model, $3K$ images for validation and $5K$ for testing purposes. We evaluated the proposed model in different dimensions to yield plausible detection results. The datasets are trained and validated separately to identify the various intrinsic complexities in the proposed hybrid VGG$-$19+Bi-LSTM model development. The development of such systems greatly supports the forest officers in identifying animal activities and saving human lives. This drastically increases the accuracy of the model. It reduces the unwanted pixel information by covering the more accurate regions of animals in an image. Further, we analyzed the accuracy of the proposed VGG$-$19+Bi-LSTM model performance by comparing it with the existing approaches.

We evaluated the performance of the proposed VGG$-$19+ Bi-LSTM model using confusion matrices. The confusion matrices' rows represent the true labels and their columns represent the predicted results. The improved scores denote the better prediction results. We compared $40k$ images of benchmark animal dataset images and depicted the classification results. The Confusion Matrix was found to be a useful method for evaluating the classification or prediction model performance. Although the model performance is good, it still sputtered out poor performance in generated results due to misclassification results. This greatly affects the reliability of the model to provide support in medical imaging, disease prediction, COVID$-$19 detection and so on. The Confusion Matrix uses NXN matrices to compare the classification results with target values. Based on that, it further identifies the defects present in the model. The following rules assist in finding the various views to explore the holistic view of the confusion matrix over classification tasks.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

$$Specificity = \frac{TN}{TN + FP} \tag{13}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Recall = \frac{TP}{TP + FN} \tag{16}$$

$$F1Score = 2 * \frac{Recall * Precision}{Recall + Precision} \tag{17}$$

$$TruePositiveRate = \frac{TP}{TP + FN} \tag{18}$$

$$FalseNegativeRate = \frac{FN}{FN + TP} \tag{19}$$

V. CONCLUSION

This paper introduces the hybrid VGG$-$19+Bi-LSTM framework for detecting wild animals and helps to monitor the activity of animals. This hybrid approach greatly helps to save the animals from human hunting and humans from animal sudden attacks by sending an alert message to the forest officer. This model introduces novel approaches to upgrade the performance of deep learning techniques in wider applications and real time cases. The proposed model has been evaluated on four different benchmark datasets that contain animal based datasets—camera trap dataset, wild anim dataset, hoofed animal dataset, and CDnet dataset. The experimental results show the improved performance of our model over various quality metrics. The proposed hybrid VGG$-$19+Bi-LSTM model achieves above 98% average classification accuracy results and 77.2% mean Average Precision (mAP) and 170 FPS values. Henceforth, the proposed hybrid VGG$-$19+Bi-LSTM model outperforms earlier approaches and produces greater results with lower computation time.

REFERENCES

[1] S. Aarthi and S. Chitrakala, "Scene understanding—A survey," in *Proc. Int. Conf. Comput., Commun. Signal Process. (ICCCSP)*, Jan. 2017, pp. 1–4.

[2] N. Ahuja and S. Todorovic, "Connected segmentation tree—A joint representation of region layout and hierarchy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[3] T. A. Assegie, P. K. Rangarajan, N. K. Kumar, and D. Vigneswari, "An empirical study on machine learning algorithms for heart disease prediction," *IAES Int. J. Artif. Intell. (IJ-AI)*, vol. 11, no. 3, p. 1066, Sep. 2022.

[4] N. Banupriya, S. Saranya, R. Swaminathan, S. Harikumar, and S. Palanisamy, "Animal detection using deep learning algorithm," *J. Crit. Rev.*, vol. 7, no. 1, pp. 434–439, 2020.

[5] M. Cheng, Z. Zhang, W. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3286–3293.

[6] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," 2016, *arXiv:1605.06409*.

[7] M. De Gregorio and M. Giordano, "Change detection with weightless neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 409–413.

[8] B. Natarajan, E. Rajalakshmi, R. Elakkiya, K. Kotecha, A. Abraham, L. A. Gabralla, and V. Subramaniyaswamy, "Development of an end-to-end deep learning framework for sign language recognition, translation, and video generation," *IEEE Access*, vol. 10, pp. 104358–104374, 2022.

[9] W. Dong, P. Roy, C. Peng, and V. Isler, "Ellipse R-CNN: Learning to infer elliptical object from clustering and occlusion," *IEEE Trans. Image Process.*, vol. 30, pp. 2193–2206, 2021.

[10] R. Elakkiya, P. Vijayakumar, and M. Karuppiah, "COVID_SCREENET: COVID-19 screening in chest radiography images using deep transfer stacking," *Inf. Syst. Frontiers*, vol. 23, pp. 1369–1383, Mar. 2021.

[11] R. Elakkiya, V. Subramaniyaswamy, V. Vijayakumar, and A. Mahanti, "Cervical cancer diagnostics healthcare system using hybrid object detection adversarial networks," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1464–1471, Apr. 2022.

[12] R. Elakkiya, K. S. S. Teja, L. Jegatha Deborah, C. Bisogni, and C. Medaglia, "Imaging based cervical cancer diagnostics using small object detection—Generative adversarial networks," *Multimedia Tools Appl.*, vol. 81, pp. 1–17, Feb. 2022.

[13] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Computer Vision—ECCV*. Dublin, Ireland: Springer, Jun. 2000, pp. 751–767.

[14] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2155–2162.

[15] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Proc. 13th Scandin. Conf. (SCIA)*. Halmstad, Sweden: Springer, Jul. 2003, pp. 363–370.

[16] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[18] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection.Net: A new change detection benchmark dataset," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 1–8.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[20] J. Imran and B. Raman, "Evaluating fusion of RGB-D and inertial sensors for multimodal human action recognition," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 1, pp. 189–208, Jan. 2020.

[21] F. Kahl, R. Hartley, and V. Hilsenstein, "Novelty detection in image sequences with dynamic background," in *Statistical Methods in Video Processing*, Prague, Czech Republic: Springer, May 2004, pp. 117–128.

[22] T. Liang, H. Bao, W. Pan, and F. Pan, "Traffic sign detection via improved sparse R-CNN for autonomous vehicles," *J. Adv. Transp.*, vol. 2022, pp. 1–16, Mar. 2022.

[23] T. Liang, H. Bao, W. Pan, X. Fan, and H. Li, "DetectFormer: Category-assisted transformer for traffic scene object detection," *Sensors*, vol. 22, no. 13, p. 4833, Jun. 2022.

[24] G. Li and Y. Yu, "Deep contrast learning for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 478–487.

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision—ECCV*. Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.

[26] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831–843, Aug. 2000.

[27] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.

[28] W. Ouyang and X. Wang, "Joint deep learning for pedestrian detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2056–2063.

[29] M. Monnet, Paragios, and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, 2003, pp. 1305–1312.

[30] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.

[32] P. E. Rybski, D. Huber, D. D. Morris, and R. Hoffman, "Visual classification of coarse vehicle orientation using histogram of oriented gradients features," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2010, pp. 921–928.

[33] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*.

[34] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 512–519.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[36] J. Sun, W. Zhang, X. Tang, and H.-Y. Shum, "Background cut," in *Computer Vision—ECCV*. Graz, Austria: Springer, Aug. 2006, pp. 628–641.

[37] X. Sun, P. Wu, and S. C. H. Hoi, "Face detection using deep learning: An improved faster RCNN approach," *Neurocomputing*, vol. 299, pp. 42–50, Jul. 2018.

[38] C. Szegedy, S. Reed, D. Erhan, D. Anguelov, and S. Ioffe, "Scalable, high-quality object detection," 2014, *arXiv:1412.1441*.

[39] H. Mechria, M. Gouider, and K. Hassine, "Breast cancer detection using deep convolutional neural network," in *Proc. 11th Int. Conf. Agents Artif. Intell.*, 2019, pp. 327–338.

[40] C.-Y. Wang, I.-H. Yeh, and H.-Y. Mark Liao, "You only learn one representation: Unified network for multiple tasks," 2021, *arXiv:2105.04206*.

[41] Z. Zhang, Z. He, G. Cao, and W. Cao, "Animal detection from highly cluttered natural scenes using spatiotemporal object region proposals and patch verification," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 2079–2092, Oct. 2016.

[42] ID Zone. (2015). *Histogram of Oriented Gradients (HOG) Descriptor*. Accessed: Jul. 28, 2016. [Online]. Available:https://software.intel.com/en-us/node/529070

[43] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 1–9.

[44] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, Sep. 2013.

[45] G. Hinton and L. van der Maaten, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008.

[46] R. Bhuvaneswari and S. G. Vaidyanathan, "Classification and grading of diabetic retinopathy images using mixture of ensemble classifiers," *J. Intell. Fuzzy Syst.*, vol. 41, no. 6, pp. 7407–7419, Dec. 2021.

[47] T. Vignesh, K. K. Thyagharajan, R. B. Jeyavathana, and K. V. Kanimozhi, "Land use and land cover classification using recurrent neural networks with shared layered architecture," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2021, pp. 1–6.

[48] B. Natarajan, R. Elakkiya, and M. L. Prasad, "Sentence2SignGesture: A hybrid neural machine translation network for sign language video generation," *J. Ambient Intell. Humanized Comput.*, vol. 14, pp. 9807–9821, Jan. 2022.

[49] R. Bhuvaneswari and K. Thaiyalnayaki, "Kernel collaboration with perceptron for facial emotion recognition," *Int. J. Comput. Appl.*, vol. 69, no. 13, pp. 29–32, May 2013.

[50] P. Yadav, N. Menon, V. Ravi, S. Vishvanathan, and T. D. Pham, "EfficientNet convolutional neural networks-based Android malware detection," *Comput. Secur.*, vol. 115, Apr. 2022, Art. no. 102622.
[51] H. Theivaprakasham, S. Darshana, V. Ravi, V. Sowmya, E. A. Gopalakrishnan, and K. P. Soman, "Odonata identification using customized convolutional neural networks," *Exp. Syst. Appl.*, vol. 206, Nov. 2022, Art. no. 117688.

B. NATARAJAN received the B.E. degree, in 2011, and the M.E. degree, in 2015. He is currently pursuing the Ph.D. degree with the School of Computing, SASTRA University, Thanjavur. He is a Faculty Associate with the Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India. He has eight years of teaching experience and has published many articles in leading international journals. His research interests include computer vision, machine learning, deep learning, and sign language development.

R. ELAKKIYA received the Ph.D. degree from Anna University, Chennai, in 2018. She was an Assistant Professor with the Department of Computer Science and Engineering, School of Computing, SASTRA University, Thanjavur. She is currently an Assistant Professor with the Department of Computer Science, Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai International Academic City Dubai, United Arab Emirates. She got three patents and has published more than 35 research papers in leading journals, conference proceedings, and books, including IEEE, Elsevier, and Springer. Her research interests include deep learning and computer vision. She is a Lifetime Member of the International Association of Engineers. She is also an Editor of the *Information Engineering and Applied Computing* journal. She has organized various events, including a Workshop on "Cyber Security" with the Agni College of Technology, Chennai, from 2014 to 2017, sponsored by IIT Bombay, and a Software Development Workshop, "EKTIA" with the Jerusalem College of Engineering, Chennai, in 2012.

R. BHUVANESWARI (Member, IEEE) received the Ph.D. degree from Anna University. She is currently an Assistant Professor with the Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India. She has 18 years of teaching experience in the field of engineering. She has authored over many publications on international journals and international conferences and coauthored a book on computer graphics. Her research interests include machine learning and deep learning for image processing applications.

KASHIF SALEEM received the B.Sc. degree in computer science from Allama Iqbal Open University, Islamabad, Pakistan, in 2002, the P.G.D. degree in computer technology and communication from Government College University, Lahore, Pakistan, in 2004, and the M.E. degree in electrical and electronics engineering and telecommunication and the Ph.D. degree in electrical engineering from the University of Technology Malaysia, in 2007 and 2011, respectively. Since 2012, he has been with the Center of Excellence in Information Assurance (CoEIA), King Saud University, Saudi Arabia, where he is currently an Associate Professor. He is also an Adjunct Professor with the Department of Computer Sciences and Engineering, College of Applied Studies and Community Service, King Saud University. He is professionally certified by the Massachusetts Institute of Technology (MIT) in cybersecurity, the University of the Aegean in information and communication security, Institut Mines-Télécom in queuing theory, IBM in security intelligence analyst, and Microsoft and Cisco in computer networks. He acquired several research grants in Saudi Arabia, EU, and the other parts of the world. He has authored or coauthored over 100 papers in refereed journals and international conferences. His research interests include ubiquitous computing, mobile computing, the Internet of Things (IoT), machine to machine (M2M) communication, wireless mesh networks (WMNs), wireless sensor networks (WSNs), mobile adhoc networks (MANETs), intelligent autonomous systems, information security, and bioinformatics. He has served as a technical program committee member and organized numerous international workshops and conferences. He is providing services as an Associate Editor mainly to *Alexandria Engineering Journal*, *Journal of Multimedia Information System* (JMIS), IEEE Access, *International Journal of E-Health and Medical Communications* (IJEHMC), and *International Journal of Cyber-Security and Digital Forensics* (IJCSDF).

DHARMINDER CHAUDHARY received the M.Tech. degree in computer science and data processing and the Ph.D. degree in cryptography and network security from the Indian Institute of Technology, Kharagpur, India. He is currently an Assistant Professor (Senior Grade) with the Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India. He has published 27 SCI/Scopus indexed articles in the areas cryptography and network security, Internet of Drones security, and security/privacy issues in Industrial Internet of Things/robotics.

SYED HUSAIN SAMSUDEEN received the B.E. degree in electronics and communication engineering from the Latha Mathavan Engineering College, Madurai, India, in 2011, and the M.E. degree in communication systems from the Mookambigai College of Engineering, Trichy, India, in 2015. He is currently a former Lecturer and an Assistant Professor with the Department of Electronics and Communication Engineering, K. Ramakrishnan College of Engineering, Samayapuram, Trichy, which is an autonomous college under Anna University, Chennai, India. He has supervised many undergraduate student projects. He has also authored or coauthored many quality papers in various Scopus and Web of Science journals. His current research interests include wireless networks, antenna design, and medical electronics.