

Multi-shift spatio-temporal features assisted deep neural network for detecting the intrusion of wild animals using surveillance cameras

Jeen Retna Kumar R¹, Berakhah F Stanley²

¹Electronics and Communication Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi Chennai, India. jejinrsrch@gmail.com

²Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi Chennai, India.

Abstract. The coexistence of human populations and wildlife in shared habitats necessitates the development of effective intrusion detection systems to mitigate potential conflicts and promote harmonious relationships. Detecting the intrusion of wild animals, especially in areas where human-wildlife conflicts are common, is essential for both human and animal safety. Animal intrusion has become a serious threat to crop yield, impacting food security and reducing farmer profits. Rural residents and forestry workers are increasingly concerned about the issue of animal assaults. Drones and surveillance cameras are frequently used to monitor the movements of wild animals. To identify the type of animal, track its movement, and provide its position, an effective model is needed. This paper presents a novel methodology for detecting the intrusion of wild animals using deep neural networks with multishift spatio-temporal features from surveillance camera video images. The proposed method consists of a multi-shift attention convolutional neural network model to extract spatial features, a multi-moment gated recurrent unit attention model to extract temporal features, and a feature fusion network to fully explore the spatial semantics and temporal features of surveillance video images. The proposed model was tested with images from three different datasets and achieved promising results in terms of mean accuracy and precision.

1 Introduction

It is evidently clear from various reports that wild animals entering human living areas have caused significant losses in terms of crops, livestock, and pose a threat to human lives. The continuous stream of data and cluttered backgrounds make it challenging for researchers to observe animal behaviour [1]. There are numerous distinct types of wild animals, each with its unique body, tail, and facial features. Detecting and classifying these animals in video sequences, along with the processing of extensive feature maps, necessitates the development of a reliable framework. The demand for vast video data for training and testing, as well as substantial computational resources, has become a pressing requirement for advancements in real-time applications [2]. To produce credible results, integrated

approaches must also handle data intelligently. Hence, there is a strong need for creating a model to identify animal activity in forest areas. Although the modern era has witnessed many break-throughs, more attention should be directed toward this field of study to develop a robust model [3].

Detecting the intrusion of wild animals presents a significant complexity challenge for deep learning. This research examines developments in sophisticated neural network-based architectures and video analysis methods. Impressive results in image identification, classification, and generation tasks have been achieved recently due to advances in Deep Learning algorithms [4]. These advancements have led us to focus our efforts on creating a reliable model for tracking animal behaviors and alerting relevant officials in the event of unusual activity. This involves the intrusion of animals into residential or agricultural areas and hunting animals [5]. To provide a better solution, the proposed model's development explores this issue from various perspectives.

The proposed strategy is contrasted with past approaches using experimental results, and the results of the valid justification are also explored. The clarity with which the intricacies of different stages of development are discussed demonstrates the calibre of our work. It is also required to localise the animals inside the image in order to efficiently follow and observe their movements in addition to detecting their presence. The job of object detection is to do this. Systems for object detection identify areas of interest in images and classify the objects that are found there. Since this is the case, detecting the object is the best option for the system suggested in this re-search. In many instances, the models for this detection that were suggested in prior studies was not successful in many cases. Therefore, the spatio-temporal features are extracted to explore the semantic spatial characteristics as well as the moving temporal motion features in the proposed work. The following are the main contributions of the proposed method:

- i. Detecting animal intrusion from surveillance video is a challenging task and by using low level spatial features alone is much difficult to perform the detection task effectively. Hence, in the proposed work a robust and effective method by considering spatial and temporal features using deep learning approach.
- ii. Frames with moving objects alone are subjected to further processing because the moving object recognition technique has been modified to avoid processing of undesired and insignificant frames.
- iii. The spatial features are extracted from each frame using multi-shift attention CNN model and multi-moment GRU attention model is used to probe the temporal characteristics between adjacent frames.
- iv. Finally, the feature fusion module is used to utilize the combined characteristics of spatial and temporal features to effectively detect animal intrusion

The organization of this manuscript is depicted as below. The related works is re-visited in section 2 and detailed elaboration of the proposed method is depicted in section 3. Section 4 constitutes the experimental results and its analysis whereas section 5 gives the conclusion of the paper.

2 Related Works

Many methodologies have been developed for the detecting the intrusion of wild animals with the use of different datasets. Recently, deep learning methods proves a high impact in the task of image classification with accurate results [6]. From the camera trap images, the segmentation of animal features is done with the combination of Alexnet and histogram of orientation features and is described by zhang et. al. [7]. This can be accomplished by the utilization of multilevel iterative graph cut for identifying the region of interest. Erhan et. al. [8] proposed a model for the detection of wild animals from camera images using iterative

embedded graph cut along with deep learning and machine learning algorithm for classification. Although these models affirm background or animal in the extracted patches, improvement is need in the classification algorithm for better performance. Deep learning techniques have enabled object detection in computer vision applications to reach new heights [9]. Using object localization and classification algorithms, it is possible to more accurately identify different items in images and videos that are present. The objects count and their activity can be easily manipulated from the gathered results. The advancement of numerous techniques for object detection in various settings and applications demonstrates the progress and significance of object detection in research domains and its increased attention [10].

A novel animal detection model by combining deep learning algorithm with back ground modelling is proposed by yousif et. al. [11] for detecting animals in camera trap images. A technique involving scale invariant feature transform (SIFT) in addition with speeded up robust feature (SURF) for extracting the features and support vector machine (SVM) for image classification is developed by matuska et al [12]. Faster R-CNN is one of the widely used algorithm for animal detection. The traditional R-CNN method [13] introduces effective detection techniques by combining ConvNet and region proposal networks. Using annotated data, this method finds the hundreds of object classes that can be found in an image or video. The R-CNN algorithms do not employ any hashing or approximation techniques to estimate the regions of the objects. Weighted full convolution layers are used in R-FCN approaches [14] to discover ROI and detect the category of objects as well as the information of their surroundings. With the use of deep learning algorithms, object detection approaches also seem promising for autonomous vehicles [15] and traffic scene object detection [16]. The You Only Look Once (YOLO) architecture processes 155 frames per second in real-time cases to produce quicker results. Instead of taking into account classification methodologies, this strategy uses an end-to-end approach using regression and probabilistic computations to detect the objects [17]. It detects objects with good accuracy and a decreased false-positive rate. The researchers conduct a thorough analysis with regard to background subtraction and elimination. A multi stage pipeline for animal detection using YOLO is proposed by Parham et. al. [18]. Multiple vibration sensors are also used by various researchers for detecting wild animal and locating their movement.

When CNN is used for extracting spatiotemporal features, accuracy of the corresponding algorithm is improved. Utilising spatiotemporal approaches, the dynamic features of moving objects are properly handled [19]. The identification of salient regions in the absence of any signals can be difficult, despite the fact that spatial attention modules have been used extensively in the past for video analysis. The application of object detectors and a few regularizes encourages spatial attention [20]. Spatial features can be extracted by convolutional neural network and the extension of this in time dimension paves the way for the extraction of spatio-temporal features. A typical recurrent neural network (RNN), Gated recurrent unit (GRU) has the ability to preserve spatial information as well as temporal information by choosing important information in time frames. In this paper, the spatial information's are retrieved by multishift CNN model by retaining the orientation sensitive values. The temporal features are acquired by implementing CNN in various time frames and important features in time frame is chosen by GRU. The attention mechanism is followed in each module by the inclusion of attention network. In particular for complicated inputs, by focusing on the most important parts of the input, attention approaches allow the model to focus, which can improve accuracy and robustness. Furthermore, by simply analysing the most crucial portions of the data, it can lower computing costs and memory use.

3 Proposed Methodology

The proposed wild animal intrusion system embraces a pre-processing and moving object detection module, a multi- shift attention CNN module, multi-moment GRU attention module and feature fusion model. The systematic flow of the proposed work is described in figure 1. The proposed model is explained and discussed in this section.

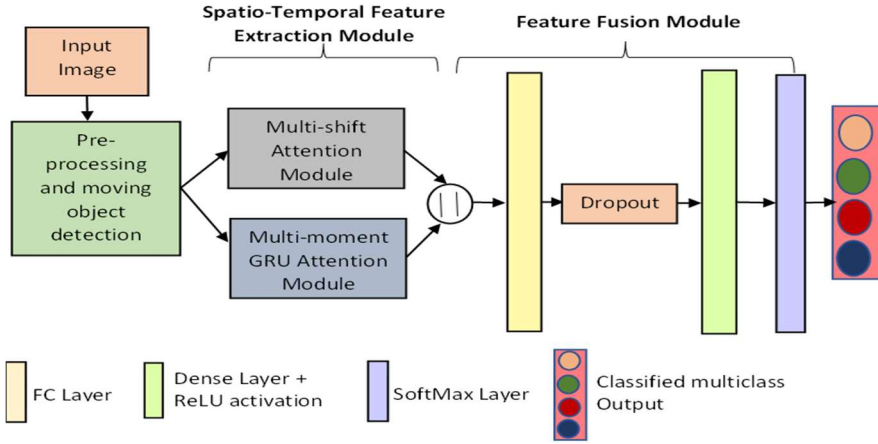


Fig. 1. Overall Block Diagram of proposed System

3.1 Preprocessing and Moving Object Detection

The input image needs to be pre-processed before training. Image smoothing is a component of the initial pre-processing stage that lowers camera image noise. Image smoothing can also be used to remove transient environmental noise from images captured by an outdoor camera, such as snow and rain. Each modal image is normalised before the characteristics are extracted because the contrast and other problems in images of different modes vary. With z-score normalisation technique, each image is standardised using the mean and standard deviation values and it is utilized in the proposed method. The normalized image is obtained using the formula

$$z_i = \frac{R_i - \mu}{\sigma} \quad (1)$$

where R_i denotes the input image from surveillance video, z_i is the normalized output image, μ and σ denotes the mean and the standard deviation of input image.

Background information is not required for the recognition of wild animals' intrusion and will be regarded as noise. Therefore, it needs to be removed from the input image so that appropriate features may be extracted from the foreground im-age to gain acceptable accuracy. By determining the Euclidean distance between two adjacent images in a video sequence, the motion features are first derived. Let R^m and R^{m+1} be the two consecutive image in the video sequence and the boundary im-age B^m is obtained from,

$$B_i^m = \sqrt{(R_i^{m+1} - R_i^m)^2} \quad (2)$$

The sharpness of the generated boundary features from boundary image should be dilated using four convolution and two pooling layers. The retrieved dilated boundary characteristics can be represented as

$$D^m = \delta(\text{conv}(\text{pool}(B^m))) \quad (3)$$

where δ denotes the sigmoid function, the convolution and max pooling layers are represented as conv and pool respectively. The kernel size of convolution layer is 7x7 and the kernel size of the pooling layers is 3x3. The element wise multiplication of dilated

boundary image D^m to the image I^{m+1} gives the boundary detected output image and is given in equation 4.

$$S^m = I^{m+1} \odot D^m \tag{4}$$

where \odot indicates the elementwise multiplication. The moving object detected im-age taken from the video sequence of camera trap dataset is shown in figure 2.

3.2 Multishift Attention Module

A multi shift convolution operation is applied to the pre-processed images in terms of horizontal up \uparrow , horizontal down \downarrow , vertical right \rightarrow and vertical left \leftarrow respectively, in order to extract spatial information for better recognition. The applied shift module has no parameters and no calculation cost. By applying the one-dimensional Shift operation in the $\uparrow \downarrow \rightarrow \leftarrow$, dimensions to capture spatial information from different perspectives and learning spatial features using convolution operations, the recognition model is both effective and high performing. The spatial features are extracted using one-dimensional shift operations with respect to the horizontal and vertical dimensions.



Fig. 2. Image clips taken from camera trap dataset with moving object detected

All the elements of an activation map are shifted one-dimension along all spatial directions using shift operations. After the shift operation, the output is subjected to a convolution layer with a kernel size of 1×1 and stride of 1. Batch normalization and a rectified linear unit (ReLU), which eliminates the image's negative value and substitutes it with zero, come after the convolution layer. For data aggregation, the output convolved image from the horizontal and vertical shift operations are concatenated while spatial attributes document the change in space as a result of the movement, spatial characteristics are retrieved by the multi-shift operation. This multi- shift module makes it simple to extract spatial information, and figure 3 depicts the mod-ule's flow.

3.3 Multi-moment GRU Attention Module

In a variety of machine learning tasks, recurrent neural networks have demonstrated excellent results. It has been used in a variety of situations, including language modelling, speech recognition, text summarization, video tagging, time-series prediction, and more. Gated Recurrent Units (GRUs) are one of the closely related versions of sophisticated recurrent units. Gated recurrent unit networks, a subset of the recur-rent neural network, are capable of processing memories of sequential data by storing past inputs in the internal state of the networks and planning from the history of prior inputs to target vectors. A GRU with temporal-wise attention module is de-signed to extract the temporal features from the input

video images. GRU is a recur-rent neural network with a reset and update gate used to preserve time specific in-formation over time.

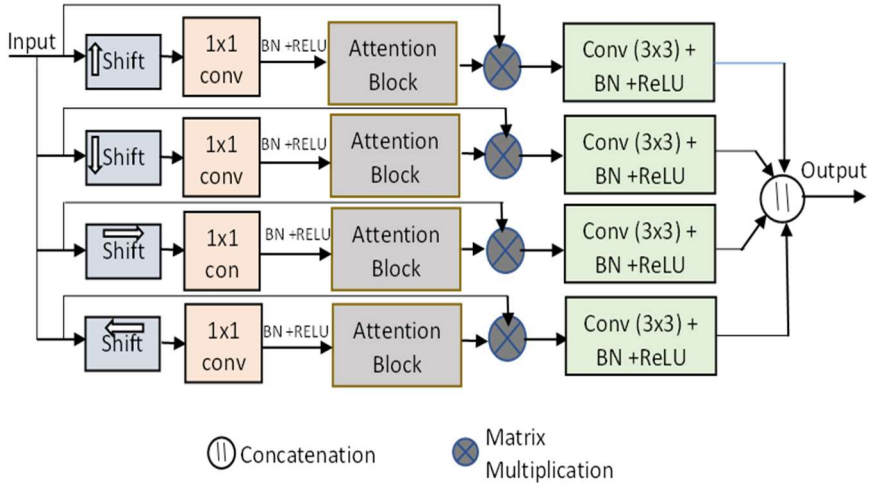


Fig. 3. Multi Shift Attention Module

An attention model is constructed after GRU to determine which frame is crucial, in contrast to earlier attention models that primarily focused on areas in each frame. Back propagation is used to train the attention model, which results in dynamic attention weights throughout the video sequence. The multi-moment GRU attention block employed in the suggested approach is shown in Figure 4. Through the attention mechanism, this module selectively extracts the discriminant traits from each frame, influencing the overall consideration. By choosing the important features gained from each frame, a promising improvement is made. The multi-moment attention module introduces feature maps from which local information can be depicted from low-level feature map and information rich semantic data can be depicted from high-level feature map. A feature fusion model is designed to explore the semantic spatial and temporal together. These two networks' attention models resulted in feature vectors that each included distinguishing qualities and had internal relationships. For each type of feature vector, general approaches for action recognition usually train two independent classifiers, but this will overlook the relationships. In order to investigate their inherent relationships, we process two vectors concurrently using a single classifier. The classifier will be trained more robustly for recognition with twice as much vector input.

The feature fusion module is shown in figure 1. This module consists of a fully connected (FC) layer and a SoftMax layer. To avoid the model over-fitting, a drop-out technique is performed between the FC and SoftMax layers. The feature fusion creates different class scores by treating each feature vector as a distinct classification task with the same label. Losses from each task are combined to create the ultimate loss during training. This method greatly enhances the deep learning performance in various domain-based applications and gives security-based application development more attention. The output of each layer, the feature maps, reflect the network's learning process. The network picks up on different aspects of an image, such as forms, contours, and edges, at each layer. These specifics help the network mature so that it can predict and categorise fresh, unexplored data more accurately. Following the application of filters and pooling processes, we may visualise the feature maps.

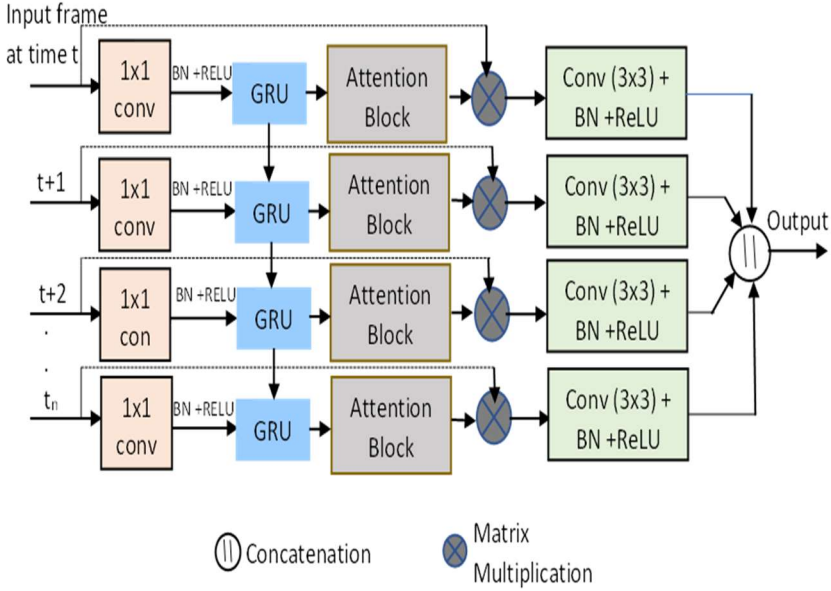


Fig. 4. Multi-moment GRU Attention Module

The attention feature module is obtained from equation 5,

$$F(Z_i) = \sigma (f_{7 \times 7}(f_{p(3 \times 3)}(Z_i))), (i=1,2,3,4) \tag{5}$$

where the sigmoid activation function is denoted as σ , $f_{7 \times 7}$ is a convolution layer of kernel size 7×7 , f_p indicates the maximum pooling layer with kernel size of 3×3 . The rectified linear unit is obtained from

$$\text{ReLU}(x) = \max(0,x) + p \cdot \min(0,x) \tag{6}$$

where p indicates the learnable parameter. The calibrated features are then obtained using the attention map via element-wise multiplication, which produces the helpful features by removing the irrelevant characteristics.

$$T_i = Z_i \otimes F(Z_i), (i=1,2,3,4) \tag{7}$$

where \otimes indicates the elementwise multiplication, $F(Z_i)$ indicates the attention feature map and T_i is the redefined feature map obtained after attention block. The redefined feature map is smoothed using batch normalisation and a ReLU unit in the final convolution layer. The redefined feature maps from each scale are then concatenated, averaged, and input into the dense layered sigmoid function for classification. These feature maps contain both high-level and low-level information.

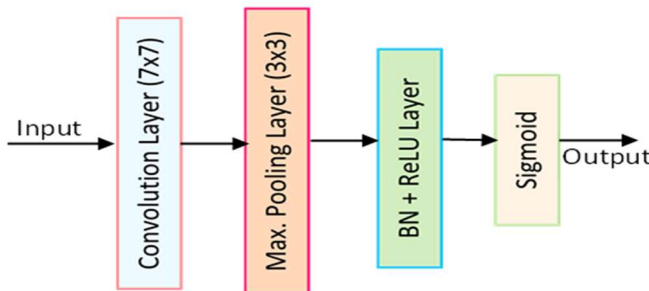


Fig. 5. Attention Block Layers

4 Experimental Results

4.1 Datasets

The experimentation of proposed system is done with the help of two public available datasets wild animal dataset [21] and camera trap dataset [22]. The sample images taken from the datasets are given in figure 6. Wild animal dataset contains 5000 images of five class of animals such as bear, elephant, leopard, lion, and wolf. The dataset is split into five subgroups, each of which has 1000 total images for training and testing. The dataset contains 256 x 256 pixel images that have been further cropped to 128 x 128 pixels. Camera trap dataset consist of 800 video sequence that contain 23 classes of animals. However, due to highly similar characteristics, eight animal categories are used in this work. They are deer, european hare, wild boar, agouti, collared peccary, paca, ocelot and mouflon. We manually selected 400 images for each class and cropped them to 128 x 128 pixels.

4.2 Results

The experiments were conducted by python tensor flow 2.0. The operating system utilized is windows 10 with intel i5 core, 8 GB RAM and NVIDIA GeForce GTX 940 graphical processing unit. The confusion matrix obtained for wild animal dataset and camera trap dataset is shown in figure 7 & 8 respectively. It is evident from the results obtained from wild animal dataset that, elephant class got high results followed by leopard class and others. This is because the spatio-temporal features obtained from elephant image is distinct from others. Bear and wolf get low accuracy since these two will get confused more due to their non distinct nature compared to other classes. The results obtained from the camera trap dataset shows that, deer got good accuracy compared to other and European hare got low accuracy results. The semantic information is retrieved by the spatial features and the transient informed are retrieved from temporal features. The performance metrics obtained in terms of accuracy, precision and F1 score for the wild animal dataset is described in table 1. The accuracy attained is 96.34%, precision is 95.38% and F1 score is 97.18%.



Fig. 6. Wild animal image samples taken from datasets

	Bear	Elephant	Leopard	Lion	Wolf
Bear	94.3	1.5	0.8	0.8	2.6
Elephant	0.6	98.2	0.5	0.5	0.4
Leopard	0.8	0.1	97.7	0.1	1.3
Lion	1.3	0.4	2.2	95.9	0.2
Wolf	2.2	0.5	1	1.8	94.5

Fig. 7. Confusion Matrix obtained for wild animal dataset

The proposed method is compared with other models and is listed in table 2. Our suggested methodology significantly attains best results than other methods in the table in terms of accuracy. Our method's accuracy in the wild animal dataset is 96.34%, which is respectable when compared to other approaches. Most state-of-the-art approaches in the wild animal dataset have good accuracy values, and our pro-posed method produces 93.4% accuracy, which is high when compared to other methods. It demonstrates both the efficiency of the proposed approach and how effectively animals are detected.

	deer	euopean hare	wild boar	agouti	collared peccary	paca	ocelot	mouflon
deer	98.2	0.9	0.2	0.2	0.3	0	0.1	0.1
euopean hare	2.3	91.3	0	0.3	0.5	4.8	0.6	0.2
wild boar	0.1	0.3	92.5	0.1	5.2	1.6	0	0.2
agouti	0	0.4	0.6	93.6	0.4	3.2	0	1.8
collared peccary	1.6	0	0.3	6.1	91.8	0.1	0.1	0
paca	1.5	1.3	0	0.2	0	93.2	0.1	3.7
ocelot	0.4	0.6	0	1.9	0.5	2.8	93.8	0
mouflon	0	0.8	0.4	3.1	2.2	0.5	0.3	92.7

Fig. 8. Confusion Matrix obtained for camera trap dataset

Table 1. Performance metrics of proposed system

Class	Accuracy	Precision	F1 Score
Bear	94.3%	92.5%	93.6%
Elephant	98.2%	96.5%	99.4%
Leopard	97.7%	99.1%	98.5%
Lion	95.9%	93.2%	98.9%
Wolf	94.5%	95.6%	95.5%
Avg	96.34%	95.38%	97.18%

Table 2. Comparison of proposed method with state-of-the-art method

Model	Wild animal dataset	Camera trap dataset
CNN	90.12	86.3
R-CNN	91.44	89.1
R-FCN	92.56	89.9
YOLO	93.15	91.6
Faster R-CNN	92.63	90.5
CNN+GRU	95.12	92.1
Spatio-temporal + CNN + GRU (proposed method)	96.34	93.4

5 Conclusion

This study contributes to the development of a system for identifying wild animals and their intrusion from camera images. The proposed framework consists of three stages, with preprocessing and moving object detection as the first stage. Spatiotem-poral features are extracted in the second stage using a multi-shift attention module and a multi-moment GRU attention module. To obtain more focused and dynamic features for further categorization, a feature fusion module is designed to combine the spatial and temporal features. The proposed method has yielded promising ex-perimental results in terms of accuracy using camera trap and wild animal datasets. In the future, there are plans to propose edge intelligence for wild animal detection with the utilization of IoT to achieve fast, robust, and reliable responses.

References

1. Bradley J Cardinale, J Emmett Duffy, Andrew Gonzalez, David U Hooper, Charles Perrings, Patrick Venail, Anita Narwani, Georgina M Mace, David Tilman, David A Wardle, et al. Biodiversity loss and its impact on humanity. *Nature*, 486(7401):59–67, 2012.

2. Agnieszka Miguel, Sara Beery, Erica Flores, Loren Klemesrud, and Rana Bayrakcismith. Finding areas of motion in camera trap images. In 2016 IEEE international conference on image processing (ICIP), pages 1334–1338. IEEE, 2016.
3. Michael A Tabak, Mohammad S Norouzzadeh, David W Wolfson, Erica J Newton, Raoul K Boughton, Jacob S Ivan, Eric A Odell, Eric S Newkirk, Reesa Y Conrey, Jennifer Stenglein, et al. Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: Mlwic2. *Ecology and evolution*, 10(19):10374–10383, 2020.
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. G, P., Bagal, M.U.: A smart farmland using raspberry pi crop vandalization prevention & intrusion detection system. *International Journal of Advance Research and Innovative ideas in Education* 1(S), 62–68 (2016).
6. Kellenberger, B., Volpi, M., Tuia, D.: Fast animal detection in uav images using convolutional neural networks. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 866–869 (2017).
7. Zhang, Z., He, Z., Cao, G., Cao, W.: Animal detection from highly cluttered natural scenes using spatiotemporal object region proposals and patch verification. *IEEE Transactions on Multimedia* 18(10), 2079–2092 (2016).
8. 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.
9. 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
10. A. B. Amjoud and M. Amrouch, "Object Detection Using Deep Learning, CNNs and Vision Transformers: A Review," in *IEEE Access*, vol. 11, pp. 35479-35516, 2023, doi: 10.1109/ACCESS.2023.3266093.
11. Yousif, H., Yuan, J., Kays, R., He, Z.: Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification. In: 2017 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–4 (2017).
12. Matuska, S., Hudec, R., Benco, M., Kamencay, P., Zachariasova, M.: A novel system for automatic detection and classification of animal. In: 2014 ELEKTRO, pp.76–80 (2014).
13. S. K. L and A. Edison, "Wild Animal Detection using Deep learning," 2022 IEEE 19th India Council International Conference (INDICON), Kochi, India, 2022, pp. 1-5, doi: 10.1109/INDICON56171.2022.10039799.
14. Tan M, Chao W, Cheng J-K, Zhou M, Ma Y, Jiang X, Ge J, Yu L, Feng L. Animal Detection and Classification from Camera Trap Images Using Different Mainstream Object Detection Architectures. *Animals*. 2022; 12(15):1976. <https://doi.org/10.3390/ani12151976>.
15. T. Liang, H. Bao, W. Pan, X. Fan, and H. Li, “DetectFormer: Categoryassisted transformer for traffic scene object detection,” *Sensors*, vol. 22, no. 13, p. 4833, Jun. 2022.
16. 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

17. 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.
18. Parham, J., Stewart, C., Crall, J., Rubenstein, D., Holmberg, J., Berger-Wolf, T.: An animal detection pipeline for identification. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1075–1083 (2018).
19. H. Gammulle, S. Denman, S. Sridharan, and C. Fookes, “Two stream LSTM: A deep fusion framework for human action recognition,” in Proc IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar. 2017, pp. 177–186.
20. Zhaoyang Niu a, Guoqiang Zhong a, Hui Yu b. “A review on the attention mechanism of deep learning”, *Neurocomputing*, Sep. 2021, pp. 48 – 62.
21. Kai Chen, Hang Song, Chen Change Loy, Dahua Lin. “Discover and Learn New Objects from Documentaries”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017*.
22. Kays, Roland, et al.: 'Monitoring wild animal communities with arrays of motion sensitive camera traps.' arXiv preprint arXiv:1009.5718 (2010).