An Efficient Elephant Detection Strategy using Visual Attention Network (VAN) in Custom Dataset improved YOLOv7 Model

Rabin Kumar Mullick[1,*,†], Rakesh Kumar Mandal[1]

[1]University of North Bengal, Raja Rammohanpur, Darjeeling, West Bengal, India

Abstract

Manual detection is crucial for managing human-elephant conflict, especially on roads or railway or human localities. Cloud-based elephant detection using YOLOv7 could mitigate conflict and provide surveillance for elephant transgression. Combining units could address larger issues. To address this issue, two key tasks are proposed: First: Detecting elephants on the railway track or highway or nearby human localities using improved YOLOv7 model, which is integration of Visual Attention Network (VAN) layer with the YOLOv7 to recognize elephants in real-time. Second: Notifying the relevant authorities. This paper examines the effectiveness of object detection methods for identifying elephants, focusing on different variants belonging to YOLO (You Only Look Once). After comparing these versions, the improved YOLOv7 model demonstrated superior performance on a custom Elephant Detection (ED) dataset. The model was trained on a combination of free and custom elephant datasets, with cloud-based cameras capturing images from multiple locations. The model attained an impressive validation accuracy of 97%.

Keywords

Elephant Detection, Google CoLab, YOLOv7, Visual Attention Network (VAN), Webcams

1. Introduction

Initial investigations indicate most collisions take place in particular 'hotspot' areas where elephant pathways cross roads or railway tracks. Often, these elephant/vehicle collisions occur because drivers do not have enough time to react at sharp turns, at night driving or under adverse weather conditions. A vision-based detection system was designed and tested in a prototype early warning system to address this problem. Driven by the initial results, detection accuracy is shown to be satisfactory under extremely varying lighting conditions under the assumption of having extensive training datasets that capture many challenging scenarios. The prototype has been shown to be robust and reliable as a whole.

This paper is structured as follows: Section 1 outlines the introduction, Section 2 reviews related works, while Section 3 introduces the proposed system with its components, while Section 4 presents the experimental results, and finally, Section 5 represents conclusion followed by references.

2. Related Works

Ecological balance depends on the presence of wild animals in the Earth. Many studies have been done in this topic yet more studies are still needed. One such problem is the problem of animal vehicle collisions (AVCs), a serious problem for biodiversity, which Saxena A., et al proposed for an intervention that uses deep learning techniques for wildlife detection and avoidance of collisions [1]. Work near the railway tracks also done and a model is proposed to detect elephants [2] [3]. For crop fields, a grid-based perceptron model to detect elephants efficiently is proposed [4]. Although camera-based methods have been recently used by many researchers to detect animals on roads, there are many

The 2024 Sixth Doctoral Symposium on Intelligence Enabled Research (DoSIER 2024), November 28–29, 2024, Jalpaiguri, India

*Corresponding author.

†These authors contributed equally.

dr.rabin.kumar.mullick@gmail.com (R. K. Mullick); rakesh_it2002@yahoo.com (R. K. Mandal)

0009-0009-7289-3198 (R. K. Mullick); 0000-0002-0471-6925 (R. K. Mandal)

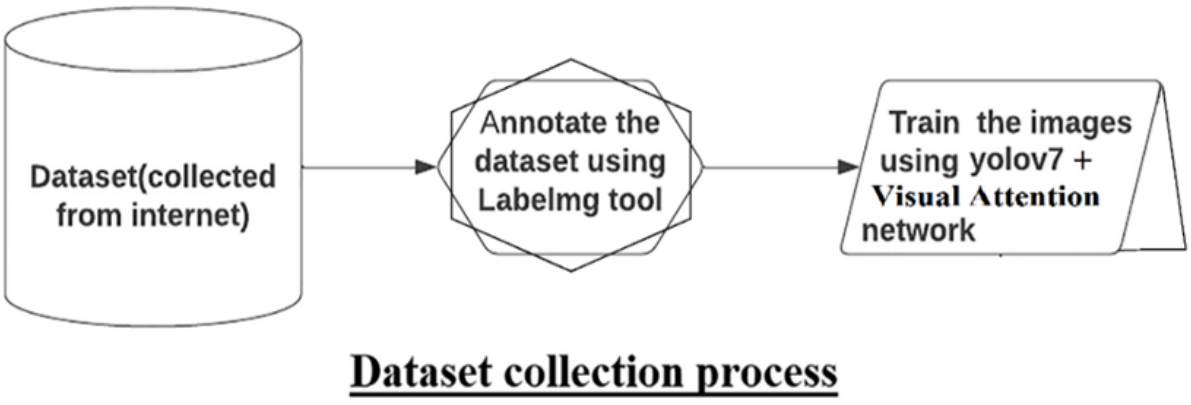

Figure 1: Illustrates the process of data collection and annotation.

limitations inherent in these techniques. However, advanced deep learning techniques have successfully been used to detect animals in colored images. Solution for detecting animal on roadways is necessary and has to be accomplished in a short timeframe of time so that it is time efficient.
Sugumar and Jayaparvathy [5] designed a system for elephants that relies on the extraction of visual features. With the help of computer vision, computers and machines are able to trained to recognize people actions, behavior, as well as dialect in a manner similar to people. Visual computing is a part of Machine Learning (ML), which attempts to find patterns in videos as well as images, through coding computers to analyze and understand the visual content that is encoded into digital data.
In image and video processing and image captioning, deep learning is widely used [6]. In deep learning (DL), artificial neural networks imitate the working belonging to individual intellect. Like people intellect, device is also grasp by machine learning with neural networks. Over the last few years DL has been applied to an incredibly wide spectrum of machine learning problems and the 'DL ecosystem' is rapidly evolving. Object detection is the 'locating' and 'classifying' of objects. Object detection includes locating relevant elements, hailing bounding boxes over them and then classify everything. This can be accomplished using machine learning (ML) and deep learning (DL) methods, with implementation of techniques such as YOLOv7 [7].
A comparable result on various tasks, including image classification, object detection, semantic segmentation, panoptic segmentation, pose estimation, etc. achieved by VAN [8]. Implementing an automated system for detecting animals plus providing warnings can assist to minimize vehicle-animal clashes on roads as well as highways [9] [10] [11] [12] [13]. Van Gemert et al. [14] uses automatic animal counting and warning system. El Abbadi et al. [15], Tan et al. [8], Ulhaq et al. [16] identify and classify the animals. Jawaharlalnehru, Arunnehru, et al. [17] proposed an improved YOLO with integration of SSD as its foundational model.
Each algorithm has its advantages and disadvantages, and the choice of which algorithm to use depends on the specific requirements of the problem. This paper also explores detection methods based on deep learning that are used to identify elephants on the streets, railway tracks or human localities as well as sending an informational alert to the concern authorities, such that appropriate action might be taken.

3. Proposed system and its components

3.1. Dataset construction

The dataset utilized in this research was sourced through the internet for the purpose of elephant detection. Images were collected featuring elephants in diverse orientations, lighting conditions, and backgrounds. Various deep learning techniques can be applied to enhance these images. Additionally, a Python script was employed to extract images from videos. Figure 1 illustrates the process of data collection and annotation for the elephant dataset using the LabelImg tool to prepare for training.

3.2. Dataset

Basically, there are freely available datasets of elephant's images are present on some open-source free websites, like: "The Aerial Elephant Dataset" [18], "Wild Elephant Dataset" [19] and "Asian vs African Elephants" [20].

3.3. Data pre-processing

Data preprocessing is one step to improve the quality of data. However, this process involves organizing plus processing raw data to generate results that are readable as well as easily available. Complexity, accuracy and sufficiency are common challenges to image data. However, the importance of the image data processing is something that has been under investigated in data science. This includes grayscale conversion, normalization, data augmentation and image standardization. In this study, the data augmentation is employed to expand as well as adjusted the dataset size plus resized the images accordingly.

Towards preparing the training dataset, the frames are extracted from the videos. However, most of these frames did not contain any elephants. To ensure robust results, the frames that only retained where confirmed elephants included, discarding the others.

3.4. Data annotation

To annotate the dataset, the "LabelImg" open-source tool is used. Users draw bounding boxes around areas of interest and label the classes. After saving, a text file is generated, where the first decimal value indicates the class ID is listed first, followed by the x-axis and y-axis centers, width as well as height. The dataset is subsequently split into training as well as testing portions for additional processing. LabelImg can export annotations in various formats, including YOLO. If the dataset is labeled with LabelImg, annotations can be directly exported in YOLO format, simplifying the conversion process.

LabelImg provides key features for efficient image annotation:

- User-Friendly Interface: Built with Python and Qt.
- Annotation Formats: Supports PASCAL VOC XML, YOLO, and CreateML.
- Annotation Modes: Detection, segmentation, and classification.
- Pre-defined Classes: Easily create and manage classes for labeling.
- Keyboard Shortcuts: Quick actions for bounding boxes, saving, and navigation.
- Remote Access: Annotate images directly from remote servers.

3.5. Cloud computing

The vast amount of data collected from Internet of Things (IoT) devices must be stored on a secure server, with cloud computing playing a crucial role in this process. Once the data is processed and analyzed, it helps identify electrical faults, errors, and other system issues more effectively.

3.6. Network connection

An internet connection is crucial for communication, with each physical object assigned an Internet Protocol (IP) address. However, as device usage grows, the limited number of available IP addresses will become inadequate, prompting to explore alternative identification methods.

3.7. Deep learning

Deep learning is a subset of machine learning within artificial intelligence that utilizes neural networks to learn from unstructured or unlabeled data. Also known as deep neural learning or deep neural networks, it automates learning processes and is inspired by the structure of the human brain.

3.8. YOLOv7 architecture overview

YOLOv7, built on the Extended Efficient Layer Aggregation Network (E-ELAN), enhances speed and accuracy with an optimized layer design. Key features include:

- Model Scaling: Customizes depth, width, and resolution for various tasks.
- Auxiliary Head: Provides extra supervision during training.
- Enhanced Loss Function: Boosts training efficiency.
- High Performance: Achieves faster, more accurate inference than YOLOv5 and YOLOv4, with fewer parameters and lower computational costs.

3.9. Proposed model

This work proposes a framework designed for real-time processing elephant identification plus alert creation to protect both humans and elephants. Cloud services are utilized to connect cameras deployed in various hotspots, with captured images kept in a database for model inspection. For actual deployment of this trained model, a local cloud is utilized to host the model and establish the connection with the area's cameras, ensuring protection against elephant attacks. This local environment is then incorporated into a global cloud via internet, forming an indispensable component of the system.

The model consists of two phases. In the first phase, the YOLOv7 model is optimized with adjustable metrics using samples from an open-source database, incorporating a Visual Attention Network (VAN) layer. The developed models are evaluated using data from AI-connected cameras positioned near localized hotspots, generating different datasets from various locations. Performance variations from the optimized models are analyzed. During the second phase, the trained model is implemented alongside a local cloud server to detect elephants in the area. Upon detection, alerts are sent to local forest authorities and the community as a warning. This Proposed Model states that an integration of YOLOv7 with Visual Attention Network (VAN) will yield an improved version of YOLOv7.

In real time also the model will work efficiently with real time data captured through webcam. This deployment is depicted in Figure 2.

The experimental results compile various studies conducted over the past years in animal re-identification and attribute prediction, unifying them under a common framework. The recognition systems developed for different animal species share a similar algorithmic design approach. While the original publications in Machine vision focused on technical aspects belonging to these algorithms, this work examines them from an application perspective. It contextualizes these studies in relation to one another, emphasizing their shared approach and their capacity to incorporate lifelong learning techniques that can enhance the decision models used.

New results are presented for identifying elephants in camera trap videos, showcasing advanced capabilities for monitoring animals beyond individual identification. These advancements include predicting individual attributes and implementing lifelong learning with human is the approach. The latter is increasingly important for real-world applications and long-term monitoring.

3.10. YOLOv7 implementation details

To train a model on Google Colab, start by creating fresh computer notebook, setting change runtime type to GPU, and running code to clone this "YOLOv7 repository" plus install necessary segments. Download and extract the dataset, then acquire the "YOLOv7" pretrained weights (yolov7.pt) for faster training. Ensure paths are correct inside 'data.yaml' record within 'yolov7/dataset' directory. Once training completes, optimal weights are stored within the 'runs' directory. With this weight file ready, you can proceed with evaluation and inference using the 'detect.py' script to identify elephants in images.

To train the YOLOv7 model, utilize Google Colab by creating fresh computer notebook, setting change runtime to GPU, and running this provided code to clone this YOLOv7 repository plus install necessary segments. Download and extract the dataset into a folder, then obtain the "YOLOv7" weight

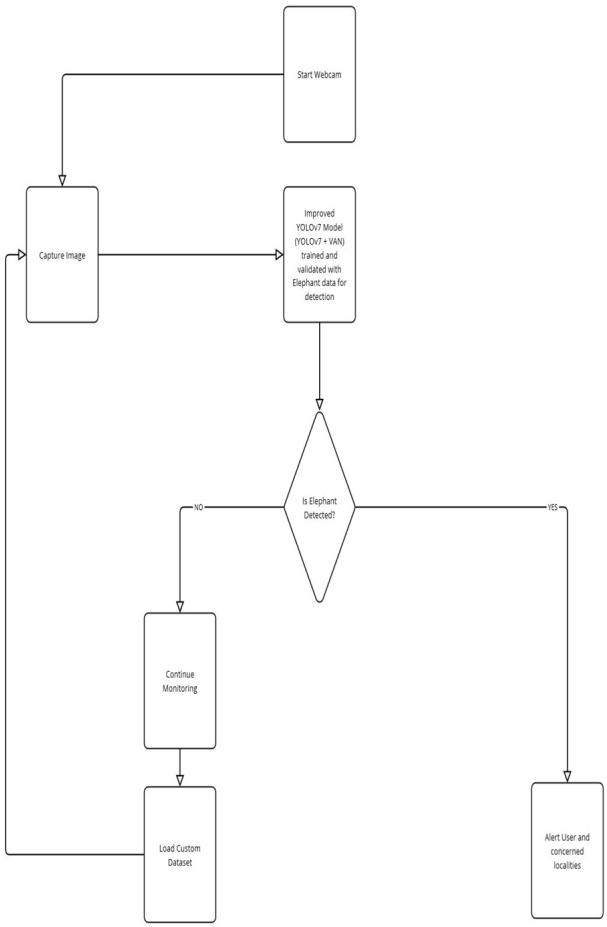

Figure 2: Flow Processes.

file, “yolov7.pt," for finetuning pretrained weights instead of starting from scratch. Create a data.yaml file to specify class names and verify that paths inside data.yaml record located in yolov7/dataset directory. After training, best weights will be stored inside ’runs’ directory. After obtaining this weight file, one can apply “detect.py” script to perform inference plus identify elephants in images.

3.11. Improved YOLOv7 with Visual Attention Network (VAN)

The attention framework is widely employed in machine learning plus deep learning algorithms. [8]. Integration of VAN with YOLOv7 has been derived in the Figure 3. To visualize YOLOv7’s integration with the Visual Attention Network (VAN), a diagram can illustrate data flow and process steps.

Components overview:

1. Input Image: The original image or video frame to be processed for object detection.
2. YOLOv7 Object Detection Module: Detects objects in real-time, outputting bounding boxes and class labels for detected objects.
3. Extract Regions of Interest (ROIs): Crops areas of detected objects based on YOLOv7’s bounding boxes for focused analysis.
4. Visual Attention Network (VAN): Applies attention mechanisms to ROIs to enhance feature extraction, focusing on relevant areas and improving representation.
5. Post-Processing: Combines VAN’s outputs with YOLOv7’s results, refining detection accuracy with attention maps or enhanced bounding boxes.
6. Final Output: An image with detected objects and VAN’s enhancements, such as attention maps or refined bounding boxes.

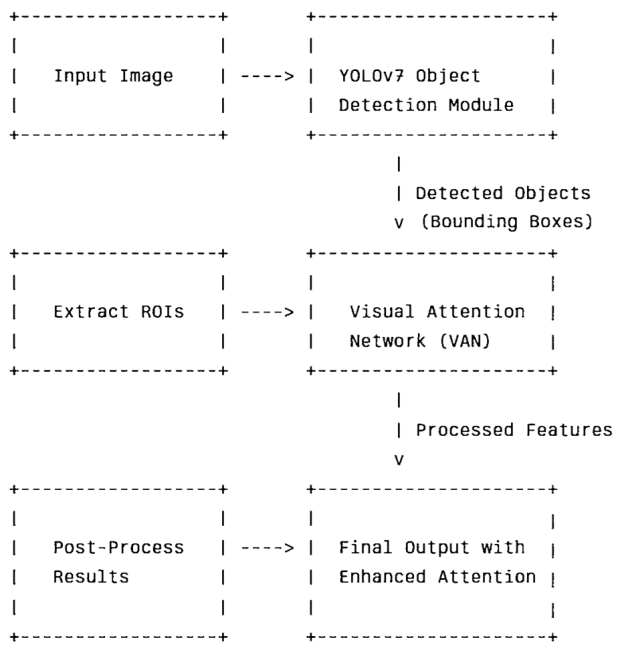

Figure 3: The integration of VAN with YOLOv7.

Conclusion: This integrated approach leverages YOLOv7's real-time detection and VAN's advanced attention mechanisms to improve object detection by isolating ROIs and applying targeted feature extraction, enhancing accuracy and performance in computer vision applications.

4. Results and discussion

Various hyperparameters, such as scaled image size and batch size, can significantly influence the identification accuracy of the Improved YOLOv7 model. Metrics such as precision, recall, F1-score, and accuracy are commonly used to evaluate the model's performance. Precision measures the ratio of correct predictions to the total number of positive predictions (false positives included), while recall assesses the proportion relating to correct predictions to the total actual positives (including false negatives). Accuracy is calculated based on both the count of accurate predictions also the count of error ones. Overall accuracy of the proposed model was computed using the formula provided in Equation (1).

An object detection module was developed to identify elephants in localized hotspot areas. The accuracy and speed of the proposed Improved YOLOv7 model were compared to those of the earlier version of YOLOv7 for elephant detection. Before testing, the YOLOv7 algorithms were trained and validated using an elephant detection dataset. The confusion matrix of elephant detection in improved YOLOv7 and YOLOv7 has been shown in Figure 4 and Figure 5 respectively.

There are mainly two classes i.e., "elephant" and "not an elephant". The total 5,535 images of prepared customed dataset have been taken for evaluating the models, which includes 4,324 elephant images and 1,211 non elephant images.

The models have been trained with 3,165 number of original elephant images and 710 number of non-elephant images. The image dataset details for models evaluation has been shown in Table 1.

The testing or validation of the models have been done with 1,660 images, out of which 1,159 images are positive, which contains elephants and 501 images are negative, which does not contain any elephant portrait. Table 2 depicts the results of the experiments.

The accuracy of improved YOLOv7 during training shows 98%, while during testing or validation shows only 97%. The detailed individual results of the models have been shown in the Table 3 for YOLOv7 model and Table 4 for the improved YOLOv7 model.

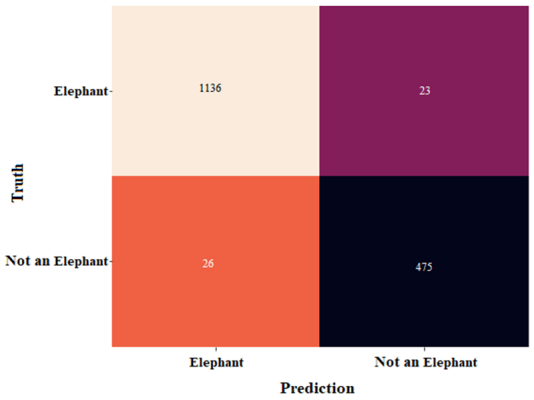

Figure 4: Confusion matrix of elephant detection in improved YOLOv7.

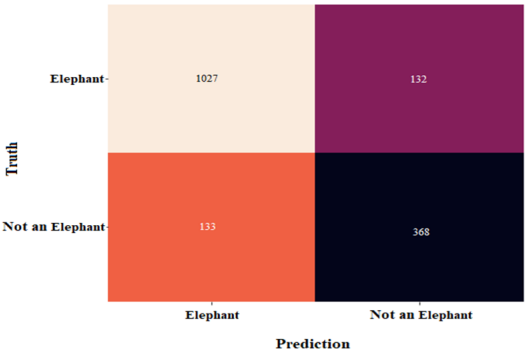

Figure 5: Confusion matrix elephant detection in of YOLOv7.

Table 1
Image Dataset Details

Image	Training	Validation	Total
Positive (Elephant)	3,165	1,159	4,324
Negative(Non_elephant)	710	501	1,211
Total	3,875	1,660	5,535

Correct predictions means true positives plus true negatives of the model during testing phase. Whereas, total predictions means summation of all predicted values. The object detected image of elephant is shown in Figure 6 evaluated in YOLOv7 model.

The object detected image of elephant evaluated in improved YOLOv7 model is shown in Figure 7.

Accuracy: It is defined as the ratio of correct predictions made by the proposed model to the total number of predictions. This metric is particularly effective when the classes of the target variable are balanced within the dataset. It can be represented as follows.

$$Accuracy = \frac{Correct\ predictions}{Total\ predictions} = \frac{1611}{1660} = 0.97\ accuracy \tag{1}$$

5. Conclusion

In this paper an integration of YOLOv7 with Visual Attention Network (VAN) is stated, that yields an improved version of YOLOv7. Incorporating VAN into YOLOv7 enables fast inference backed by the ability to understand what it sees. Combined, accuracy, reliability, and velocity optimize in complex conditions, for various actual-time domains. The deep learning algorithm is combined with a Visual

Table 2
Result Details and Comparison of Algorithms

Parameter(/Measure)	Improved_YOLOv7 (Proposed Model)	YOLOv7
Total predicted	1,660	1,660
True positive	1,136	1,027
False positive	23	132
False negative	26	133
True negative	475	368
Precision	0.980	0.886
Recall	0.977	0.885
F1 score	0.978	0.885
epochs	50	70
Accuracy	0.97	0.84

Table 3
Statistics of YOLOv7

Dataset	True Positive	False Positive	False Negative	True Negative	Total (5,535)
Train (93% correct)	3,042	123	125	585	3,875
Test (84% correct)	1,027	132	133	368	1,660

Table 4
Statistics of improved YOLOv7

Dataset	True Positive	False Positive	False Negative	True Negative	Total (5,535)
Train (98% correct)	3,127	38	39	671	3,875
Test (97% correct)	1,136	23	26	475	1,660

Figure 6: Elephant image evaluated in YOLOv7 model.

Attention Network to improve the control of parameters, selection of parameters, and convergence rate to improve the image detection and classification models. In order to prove the efficacy of the proposed method, it was implemented on an open-source elephant dataset. Both the experimental outcome and computational analysis suggest that the enhanced YOLOv7 has robust optimization features and is capable of excelling in elephant recognition and differentiation notably in the concerns of precise, grasp, and F1 measure. The proposed model achieved training accuracy of 98% and validation (or testing) accuracy of 97%.

Figure 7: Elephant image evaluated in improved YOLOv7 model.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] A. Saxena, D. K. Gupta, S. Singh, An animal detection and collision avoidance system using deep learning, Advances in Communication and Computational Technology: Select Proceedings of ICACCT 2019. Springer Singapore 668 (2021) 1069—-1084. doi:10.1007/978-981-15-5341-7_81.

[2] R. K. Mandal, A prototype model to detect elephants near the railway tracks, Advances in Modelling and Analysis B 63.1-4 (2020) 7–9. doi:10.18280/ama_b.631-402.

[3] R. K. Mandal, D. D. Bhutia, A proposed artificial neural network (ann) model using geophone sensors to detect elephants near the railway tracks, Advanced Computational and Communication Paradigms: Proceedings of International Conference on ICACCP 2017, Volume 2, Springer Singapore 2 (2018) 1–6. doi:10.1007/978-981-10-8237-5_1.

[4] R. K. Mullick, R. K. Mandal, A proposed grid-based elephant detection model using artificial intelligence (ai) to prevent crop damage in farming fields, Doctoral Symposium on Intelligence Enabled Research, Singapore, Recent Trends in Intelligence Enabled Research, DoSIER 2023, Advances in Intelligent Systems and Computing, Springer Nature Singapore 1457 (2023) 55–66. doi:10.1007/978-981-97-2321-8_5.

[5] S. J. Sugumar, R. Jayaparvathy, Automated unsupervised elephant image detection system as a solution to human elephant conflict, Proceedings of the International Conference on Multimedia Processing. Communication and Information Technology, MPCIT (2013).

[6] U. Sirisha, S. C. B, Semantic interdisciplinary evaluation of image captioning models, Cogent Engineering 9 (2022). doi:10.1080/23311916.2022.2104333.

[7] C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2023) 7464–7475. doi:10.48550/arXiv.2207.02696.

[8] M. Tan, W. Chao, J.-K. Cheng, M. Zhou, Y. Ma, X. Jiang, J. Ge, L. Yu, L. Feng, Animal detection and classification from camera trap images using different mainstream object detection architectures, Animals 12.15 (2022). doi:10.3390/ani12151976.

[9] H. R. Sodagar, M. E.-P. Rezaee, R. Shekarian, T. Rahmati, E-learning of router applications to drivers in order to reduce collisions and road accidents with wild animals, Interdisciplinary Journal

of Virtual Learning in Medical Sciences 13.1 (2022) 63–65. doi:10.30476/ijvlms.2022.94592.1139.

[10] R. Gandhi, A. Gupta, A. K. Yadav, S. Rathee, A novel approach of object detection using deep learning for animal safety, 2022, 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE (2022) 573–577. doi:10.1109/Confluence52989.2022.9734225.

[11] D. Sato, A. J. Zanella, E. J. X. Costa, Computational classification of animals for a highway detection system, Brazilian Journal of Veterinary Research and Animal Science 58 (2021) 1–10. doi:10.11606/issn.1678-4456.bjvras.2021.174951.

[12] Y. Munian, A. Martinez-Molina, M. Alamaniotis, Intelligent system for detection of wild animals using hog and cnn in automobile applications, 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA), IEEE (2020) 1–8. doi:10.1109/IISA50023.2020.9284365.

[13] S. U. Sharma, D. J. Shah, A practical animal detection and collision avoidance system using computer vision technique, IEEE access 5 (2016) 347–358. doi:10.1109/ACCESS.2016.2642981.

[14] J. C. van Gemert, C. R. Verschoo, P. Mettes, K. Epema, L. P. Koh, S. Wich, Nature conservation drones for automatic localization and counting of animals, Computer Vision-ECCV 2014Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part I 13, Springer International Publishing (2015) 255–270. doi:10.1007/978-3-319-16178-5_17.

[15] N. K. E. Abbadi, E. M. T. A. Alsaadi, An automated vertebrate animals classification using deep convolution neural networks, 2020 International Conference on Computer Science and Software Engineering (CSASE), IEEE (2020) 72–77. doi:10.1109/CSASE48920.2020.9142070.

[16] A. Ulhaq, P. Adams, T. E. Cox, A. Khan, T. Low, M. Pau, Automated detection of animals in low-resolution airborne thermal imagery, Remote Sensing 13.16 (2021) 3276. doi:10.3390/rs13163276.

[17] A. Jawaharlalnehru, T. Sambandham, V. Sekar, D. Ravikumar, V. Loganathan, R. Kannadasan, A. A. Khan, C. Wechtaisong, M. A. Haq, A. Alhussen, Z. S. Alzamil, Target object detection from unmanned aerial vehicle (uav) images based on improved yolo algorithm, Electronics 11.15 (2022) 2343. doi:10.3390/electronics11152343.

[18] The aerial elephant dataset, 2024. URL: https://zenodo.org/records/3234780, [Accessed in May 15, 2024].

[19] Wild elephant dataset, 2024. URL: https://www.kaggle.com/datasets/gunarakulangr/sri-lankan-wild-elephant-dataset, [Accessed in May 15, 2024].

[20] Asian vs african elephants, 2024. URL: https://www.kaggle.com/datasets/vivmankar/asian-vs-african-elephant-image-classification, [Accessed in May 15, 2024].