International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 12, 2021
www.ijirse.com
ISSN: 2454-9665

YOLO BASED ANIMALMONITORING SYSTEM IN FOREST RANGE AGRICULTURAL LANDS

Akila B[1] , Senthilkumar S[2] , Sivashankar T[3]

[1]*B.com CA, Bharathiyar university ,modakurichi. (India)*
[2]*B.tech IT, DR. Mahalingam college of engineering and technology.(India)*
[3]*B.tech IT, DR. Mahalingam college of engineering and technology.(India)*

Abstract

In order to inform conservation and management choices, efficient and consistent monitoring of wild animals in their natural habitats is critical. Automatic covert cameras, sometimes known as "cameras," are a type of hidden camera that is Because of their usefulness and consistency in gathering data, "traps" are becoming a more common instrument for wildlife monitoring. Wildlife data is collected in an unobtrusive, continuous, and large-scale manner. However, processing such a massive number of photographs and movies might be challenging. Manually capturing images from camera traps is quite costly .It is both time-consuming and boring. This is a significant problem.Biologists and ecologists have an impediment to openly monitoring animal's environment. Taking use of current deep learning improvements in this study, we suggest a new approach in computer vision. Foundation for developing in-the-wild automated animal recognition Seeking for a wildlife monitoring system that is automated. We leverage a single-labeled dataset from citizen scientists' Wildlife Spotter project, as well as state-of-the-art deep convolutional neural network architectures, to train a computational system capable of autonomously filtering animal photos and recognizing species. Our experimental results showed that detecting images containing animals was 96.6 percent accurate, and identifying the three most common species among the set of images of wild animals taken in South-central Victoria, Australia, was 90.4 percent accurate, demonstrating the feasibility of building fully automated wildlife observation systems. As a result, research findings can be faster, citizen science-based monitoring systems may be built more efficiently, and management choices can be made more quickly, potentially having a substantial influence on the field of ecology and trap camera image processing.

Keywords *: animal recognition, citizen science ,deep learning, large scale image classification, wildlife monitoring, yolo .*

INTRODUCTION

In ecology, one of the most important tasks is to observe wild animals in their native habitats. Over-exploitation of natural resources is occurring as a result of the rapid rise of the human population and the never-ending quest of economic progress, resulting in quick, unique, and significant changes to Earth's ecosystems. Human activity has altered the population, habitat, and behaviour of animals across a growing region of land surface. More significantly, numerous wild species have been pushed to extinction on Earth, and many species have been brought into new places where they potentially damage natural and human systems [1]. Monitoring wild animals is also critical because it provides researchers with knowledge to help them make conservation and management decisions that will keep ecosystems diversified, balanced, and sustainable in the face of these changes.

(a) a Reconyx covert camera (b) a camera trap deployed in the wild

Figure 1: An example of camera trap setting in the open space

ISSN: 2454-9665

Radio tracking [2], wireless monitoring, and other modern technologies have all been created for wild animal monitoring. Satellite and worldwide positioning, as well as sensor network tracking [3].system (GPS) tracking [4], [5], and motion-sensitive camera trap monitoring [6]. Motion-activated remote cameras, often known as "camera traps," are becoming increasingly popular. Because of their unique properties, wildlife monitoring is possible. Greater commercial availability, as well as deployment and deployment ease operation. Figure 1 shows a common covert camera model.1) is capable of collecting not just high-resolution photographs in Both during the day and at night, as well as collecting time-related data Incorporated temperature and moon phase into picture data. In Furthermore, tracking is possible because to the camera's extensive and versatile settings. Animals in private and on a regular basis When a battery is completely charged, it may be used for a variety of purposes.

The camera can take hundreds of shots in a row, giving you a lot of options. A vast amount of data Camera traps are made according to these parameters. Ecologists will find this a useful tool because it allows them to document every step of the process.an element of wildlife .If visual data can be acquired, it is a rich source of information that may help scientists answer ecological issues like: what are the geographical distributions? Unique animals, which species are endangered and require protection Bandicoot protection, for example, is a pest species that belongs to the bandicoot family. Red foxes and rabbits, for example, must be kept under control. Are some instances of important questions to consider when learning about wild animals? Populations, ecological linkages, and population dynamics are all important factors to consider. To this goal, ecologists have lately adopted a widely utilized strategy [7].is to put up a number of camera traps in the field to gather data.[6], [7], [8] data on wild animals in their native habitats .Thanks to developments in digital technology, more sophisticated camera traps with automated system components but cheaper purchasing costs are fast being used for wildlife monitoring; yet, the work of evaluating large collections of camera trap photos has been done manually. Despite the fact that the human visual system can interpret pictures quickly and easily [9], manually processing such a large number of images is prohibitively costly. For example, from 2010 to 2013, the Snapshot Serengeti project1 collected 3.2 million photos with 225 camera traps around Tanzania's Serengeti National Park [8].Species Spotter2, a similar initiative, gathered millions of photographs of wildlife shot in Australia's tropical rainforests and arid rangelands. Unfortunately, the great majority of acquired photographs are difficult to interpret, even for humans, due to the automated trap camera snapping mechanism.

As shown in Figure 2a, only a small percentage of the photos obtained are in good condition. Many photographs merely show a portion of the body of work. Animal items (Figure 2d), whereas in others they are The entire body was caught, but it was too far away from the camera (Figure).2b), or occlusion in various viewpoints or deformations (Figure 2g)(See Figure 2f) Furthermore, many of the photographs are grayscale. They were photographed at night with the use of an infrared light (Figure).2e), and a considerable proportion of photos are devoid of animals, as seen in Figure2h (75 percent of the Serengeti Snapshot [8] and 32.26 percent of the Wildlife Snapshot) Photographs with no animal labels were classed as "no animal"), whereas images with animal labels were rated as "no animal" by the spotter. Others may display a variety of goods from various categories. Species. Amounts of data are overwhelming, and image quality is restricted. As a result, the picture analysis is significantly slowed. Process.Volunteers were asked as "citizen scientists" to join the picture analyzing process remotely using Web-based image categorization tools in huge wildlife monitoring initiatives such as Snapshot Serengeti or Wildlife Spotter to share scientists' labor. The effectiveness of citizen science programmers can be shown in the high number of volunteers involved and the species recognition accuracy of 96.6 percent gained on the Snapshot Serengeti dataset [8], which was confirmed by specialists. Figure 2: Various scenarios using the Wildlife Spotter picture collection. The original photos have 1920x1080 or 2048x1536 pixel resolutions. For illustrative purposes, all pictures have been scaled.

(a) a good image

ISSN: 2454-9665

(b) scale variation/far field problem: the bird is far from camera
(c) background clutter: the reptile blends into scene
(d) object is too close to camera
(e) image captured at night with infrared flash
(f) object occlusion problem
(g) object deformation/unclear view
(h) an image without animal, but being presented to citizen scientist to annotate

However, even for specialists, the large collections of photos and the limitations of inadequate image quality have a significant impact on human categorization speed and accuracy [8]. Experts annotated some images in the Snapshot Serengeti dataset as "impossible to identify," over 9,600 images in the Wildlife Spotter dataset of Southcentral Victoria as "something else" or "image problem," and thousands of photos were labelled inconsistently (for example, the same image was classified as different species by different volunteers). Furthermore, even though many volunteers were eager in participating in citizen science programmers, manually analyzing millions of photographs would take a long time. For example, it took more than two months for a group of 28,000 registered and 40,000 unregistered volunteers in the Snapshot Serengeti project to annotate a 6-month batch of photos [8]. As a result of these challenges, there is a demand for wild animal identification automation.We believe that, to the best of our knowledge, there are currently just a few works that have attempted to construct an automated system to analyses and analyses films and photographs recorded in the field for the purpose of environmental monitoring.The massive volumes of data generated by camera traps emphasize the importance of image processing automation. There are some immediate techniques to make wildlife identification automated from a data analysis and machine learning standpoint, such as using a linear support vector machine (SVM) classifier with manual object bounding on hand-crafted features [10], a convolutional neural network (YOLO) model with automatic object detection [11], or fine-tuning YOLO models inheriting model weights pre-trained on a very large scale dataset like the ImageNet [12], [13]. These methods tackled the issue of wildlife monitoring automation and yielded encouraging empirical findings. However, there are two major obstacles that stand in the way of implementing an automated wildlife monitoring programmer in practice. The first barrier is that a tremendous amount of human preprocessing is still necessary to get acceptable image classification accuracy when using photos for recognizing and bounding animal things [10]. The second drawback is that, despite total automation, the wildlife monitoring system performs poorly, necessitating further modifications for practical implementation [11].In this research, we propose a framework for recognizing animals in the wild, with the goal of creating a completely automated wildlife spotting system. The state-of-the-art capabilities of new deep YOLO models for image classification has driven our study, particularly recent proof that automated recognition can outperform humans in some object identification tasks in the ImageNet competition [14]. Experiments are conducted using datasets from the Wildlife Spotter project, which comprise a huge number of photos captured by trap cameras installed by Australian scientists. Because the Wildlife Spotter dataset contains both animal and non-animal images, we split the wild animal identifying automation into two tasks: (1) wildlife detection, which is a binary classifier capable of classifying input images into two classes: "animal" or "no animal" based on the prediction of animal presence in images; and (2) wildlife identification, which is a multiclass classifier capable of labelling each input image with animal presence by a set of criteria. Each job is essentially a deep YOLO-based classifier that has been trained using datasets that have been manually tagged by volunteers.For comparisons, the framework employs a number of different deep YOLO architectures. By automatically filtering out a huge number of non-animal photographs where citizen annotators are currently spending their time, the achievement of Task 1 will have a considerable influence on enhancing the efficiency of citizen science-based programmers (e.g., Wildlife Spotter). Our results on the

ISSN: 2454-9665

Wildlife Spotter datasets suggest that this strategy is practical and can save a significant amount of time and money. As a result, the work's main contribution is that, given enough data and computing infrastructure, deep learning could be used to build a fully automatic image classification system on a large scale, freeing scientists from the burden of manually processing millions of images, as the project managers see it. "It's a task that computers are incapable of performing." Furthermore, our suggested framework may be integrated with an existing citizen science initiative to create a "hybrid" image classifier with an automated component that acts as a recommendation system, offering volunteers with helpful ideas to help them categories images faster.The remainder of the paper is laid out as follows. In Section II, we go over the basics of YOLO and how it may be used to classify images. In this part, we also discuss related work on the issue of automated wildlife categorization, as well as the Wildlife Spotter, a citizen science-based wild animal classification initiative. In Section III, we go over the suggested animal recognition framework, data, and experimental setup. Section IV presents the empirical findings as well as a commentary. Finally, in Section V, we come to a conclusion and make recommendations for further work.

II Related work

In this part, we'll go over the YOLO and how it's used in picture categorization. Then, in recent ImageNet Challenges [14], we discuss numerous YOLO architectures that have proven state-of-the-art performance. Finally, we review known techniques to a specific problem: animal categorization using camera trap photographs in real situations.

A. Convolutional Neural Networks for Image Classification

Because of the diverse and variable features of pictures, visual recognition is a relatively simple process for humans, but it remains a challenge for automated image recognition systems [15]. Variations in location, scale, perspective, backdrop, or illumination can yield an endless number of distinct photographs for any object of interest. Real-world problems such as wild animal categorization from automatic trap cameras, where the majority of collected photos are of poor quality, as detailed in Section I, present more challenges. As a result, building models capable of being invariant to specific input modifications while maintaining sensitivity with inter-class objects is critical for picture classification automation [16].

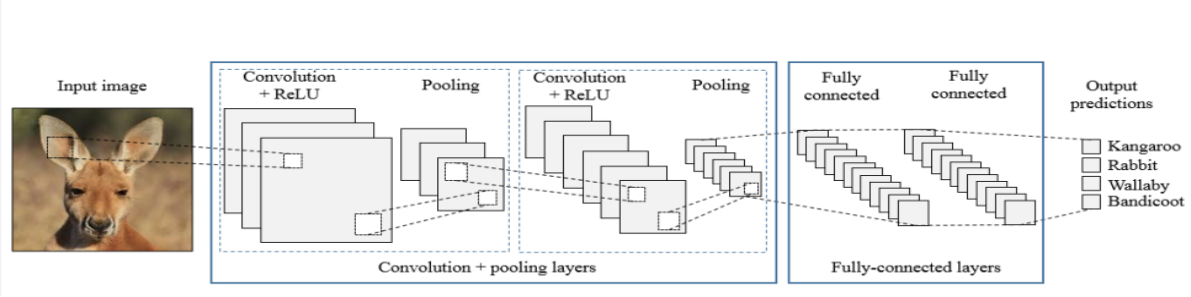

Figure 3: Illustration of a typical convolutional neural network architecture setup.

YOLOs were first proposed by LeCun et al. [17], and have since demonstrated excellent practical performance and have become widely used in machine learning, particularly in the areas of image classification [14], [18], [19], [20], [21], speech recognition [22], and natural language processing [23], [24].Due to recent advances in neural networks, specifically deep YOLOs, and computing power, particularly successful implementations of parallel computing on graphical processing units (GPUs) and heterogeneous distributed systems for learning deep models in large scale, such as Tensor Flow [26], these models have achieved state-of-the-art results that even outperformed humans in image recognition tasks [25].YOLOs are learning models based on neural networks that are especially intended to take into account the spatial structure of input pictures, which are typically in three dimensions: width, height, and depth (the number of colour channels). A YOLO is essentially a sequence of layers that can be divided into groups, each consisting of a convolutional layer plus a non-linear activation function, usually the Rectifier Linear Unit (ReLU) [20], and a pooling layer, usually max pooling; followed by several fully-connected layers, the last of which is the output layer with predictions, as shown in Figure 3. Each neuron in a conventional neural network is entirely linked to all neurons in the preceding layer, and each layer's neurons are completely independent. When used to high-dimensional data like natural

ISSN: 2454-9665

photographs, the total number of parameters might exceed millions, resulting in a major overfitting issue and making training impossible.In YOLOs, on the other hand, each neuron is only linked to a small part of the previous layer, resulting in local connection. The convolution layer computes the outputs of its neurons coupled to particular areas in the preceding layer, with a filter size defining the spatial breadth of this connection. Furthermore, another significant aspect of YOLOs, parameter sharing, decreases the number of parameters and hence computation complexity substantially.As a result, as compared to normal neural networks with similar layer sizes, YOLOs contain many fewer connections and parameters, making them easier to train but marginally degrading performance [20]. These three main characteristics – spatial structure, local connectivity, and parameter sharing – enable YOLOs to convert input images into layers of abstraction, with the lower layers displaying detail features such as edges, curves, and corners, and the higher layers displaying more abstract object features.

B. Wildlife Classification

Camera traps are an effective and dependable way of natural observation because they can capture a high volume of visual data in a natural and cost-effective manner. The wildlife data, which may be totally automated gathered and collected via video traps, is a burden for scientists to study in order to discern whether an animal is present in each image or to determine which species the items belong to. Table I: The most popular and successful YOLO architectures for image classification might be automated by automating this costly and time-consuming human analysis procedure. Reduce a big number of human resources and deliver research results rapidly.

Table I: The most common and successful YOLO architectures for image classification

Model	Alexnet	VGG-16	Googlenet	ResNet-50
Trainable layers	8	16	22	50
Main specification	5 convolutional layers and 3 fully-connected layers. [20]	13 convolutional layers with 3x3 filters, and 3 fully-connected layers. [18]	Developed an Inception Module that dramatically reduces the number of parameters while achieving high accuracy. Average pooling is used at top of YOLO instead of fully-connected layers. [19]	A deep residual learning framework, skip connections and batch normalization. Much deeper than VGG-16 (50 compared to 16) but having lower complexity and higher performance. [21

Few attempts have been made to create an automated wildlife classification system. For picture classification, Yu et al. used enhanced sparse coding spatial pyramid matching (ScSPM) [27], [28]. Animal items are manually recognized and cut out of the backdrop with the entire body, then image characteristics are extracted using the ScSPM to convert an image or a bounding box to a single vector, and lastly a linear multi-class SVM is used to classify them.On their own dataset of 7,196 photos representing 18 species, the average classification accuracy was 82%. Chen et al. suggested a YOLO-based model with picture segmentation preprocessing that was automated [11]. The network is made up of three convolutional layers with filter sizes of 909, each followed by a max pooling layer with a kernel size of 2 2, a fully connected layer, and a softmax layer. Furthermore, unlike [10], the animal object cropping procedure in [11] was carried out automatically using an automated segmentation approach called Ensemble Video Object Cut (EVOC) [29]. Despite the fact that Chen's suggested framework is totally automated and outperforms a typical Bag-of-visual-words model based picture classification method [30], [31], the recognition results obtained on their own dataset were only approximately 38.32 percent, rendering it useless in reality. Gomez et al. [32], [12] used deep YOLO models, which have shown state-of-the-art performances on the ImageNet dataset, to deal with the problem of large-scale wild animal identification on a new open dataset, the Snapshot Serengeti [8]. They were motivated by the success of deep YOLO-based models in recent ILSVRC contests. In particular, all YOLO models were pre-trained with

ISSN: 2454-9665

the ImageNet dataset in [32], [12], and then re-trained on top layers of a fresh dataset, a process known as fine-tuning. This is based on the idea that, in data-driven techniques, a network pre-trained on a big dataset like the ImageNet will have learnt features effectively for most picture classification problems, resulting in greater performance than training on smaller datasets [12].

C. Citizen Science

Many research fields, notably ecology and environmental sciences, rely heavily on citizen science [33], [34], [35], [36]. A citizen scientist is a volunteer who helps science by collecting and/or analysing data as part of a scientific investigation. One of the primary causes responsible for the current expansion of citizen scientific initiatives is significant advancements in digital technology, particularly the Internet and mobile computing [33]. Volunteers can now participate in a project remotely by collecting data or processing introduced data using specified software on their mobile phones or laptops, and then entering it online into centralised, relational databases [36]. Citizen scientists are increasingly involved in a wide range of initiatives, including climate change, invasive species, and other types of monitoring [33], [36]. Furthermore, public participation considerably aids the field of machine learning. Human-labeled datasets, such as Snapshot Serengeti or Wildlife Spotter, are excellent resources for supervised machine learning techniques that require huge volumes of labelled data to train automated models. Many Internet-based apps, such as Google Search, Facebook, and Amazon, employ machine learning techniques to improve their business management by collecting data from public user behaviour.Apart from the great contributions that citizen science makes, working with citizen science data poses a number of obstacles [36]. As a result, two technical concepts should be considered. To begin, citizen scientists' data must be adequately vetted. Second, consistent data gathering and processing techniques and tools must be developed [36].

III. DEEP YOLO FOR ANIMAL RECOGNITION FRAMEWORK

We describe our suggested picture categorization framework in this part, as well as its application to the Wildlife Spotter datasets.First, we'll go through the datasets. Then, for wildlife identification, we present a YOLO-based framework. Finally, we describe some of the YOLO designs that we used in our tests and implementations.

A. Wildlife Spotter Dataset

We're particularly interested in the South-central Victoria Wildlife Spotter collection, which now has 125,621 single-labeled photos. The photographs were acquired in both colour and grayscale settings at 19201080 and 20481536 resolutions using 30 Reconyx HC600 Hyperfirehidden cameras. They were captured in daylight without flash and at night with infrared flash in both colour and grayscale settings. We create a collection of 108,944 tagged photos from this dataset, each annotated by around 5 distinct citizen scientists. A citizen scientist in South-central Victoria, Australia, was instructed to classify an observed animal into one of 15 wildlife species (bandicoot, wombat, rat, brush tail possum, mouse, cat, rabbit, wallaby, ringtail possum, echidna, dog, fox, koala, kangaroo, and deer) and three groupings of species (mammal, bird, and reptile). "No animal" is the term for an image that lacks the look of an animal. The picture is labelled as "something else" or "image difficulty" if the user is unsure about his or her judgement owing to poor image quality or occlusions.We delete samples that are duplicated or inconsistently labelled (i.e., the same picture but labelled differently by different citizen scientists, e.g., photos classified as "something else" or "image issue," or containing the tags "animal" and "no animal" at the same time) as part of the preprocessing. Finally, we have a collection of 107,022 single-labeled photos with 34,524 non-animal samples and 72,498 samples from 18 species, as shown in Table II, accounting for more than 85 percent of the original dataset. Following that, we create two scenarios in which to test our suggested framework on two tasks: wildlife detection and identification. We analyze a binary classification issue in the first case and experiment with both balanced and unbalanced classes. To begin, we look at a common scenario in machine learning algorithm training: the balanced dataset, in which each training class comprises 25,000 samples for training and 8,500 samples for validation. The balanced dataset is created by reducing the size of the superior classes to that of the minority class. In real-life challenges where certain classes are better than others, data imbalance is a common occurrence, and the Wildlife Spotter project is no exception. Bird has the biggest population with 22,145

ISSN: 2454-9665

samples, while deer has the smallest population with only 16 photos. Because classifiers are inclined toward better classes, this extreme imbalance is likely to lead to misclassification. We employ 107,000 labelled pictures for training in the case of an unbalanced dataset, separated into two sub-sets: training set and validation set. The training set has 80,000 photos, 55,000 of which are labelled "animal" and the remaining 25,000 are labelled "no animal"; the validation set contains 18,500 and 8,500 images, respectively, labelled "animal" and "no animal."
Due to the huge number of animals with varying numbers of observations (cf. Table II), we resort to two experimental scenarios for the later case of Wildlife identification or animal recognition: recognizing the three most common species and the six most common species, respectively. When it comes to recognizing the three most common species (bird, bandicoot, and rat), we start with a balanced dataset, in which each class has 8,000 photos for training and 2,000 images for validation. The unbalanced dataset scenario is then used to train and test all samples of these three classes from the dataset. Table II lists the species and the matching number of photos utilized in the unbalanced example. We only look at the scenario of an unbalanced dataset for the most difficult task: recognizing the six most common species (bird, rat, bandicoot, rabbit, wallaby, and mammal). Animals from the Wildlife Spotter dataset in South-central Victoria, Australia, with label data (sorted in descending order of the number of images). 80 percent of the six species will be utilized for training, while the remaining 20% will be used for validation.

species	sample
Mammal	4982
Fox	1217
Kangaroo	322
Rat	11884
dog	204

B. Recognition Framework for Animal Monitoring

The Wildlife Spotter labelled dataset, as mentioned earlier in this section, comprises both animal and non-animal photos in proportions of 67.74 percent and 32.26 percent, respectively. As a result, the Wildlife Spotting system has two tasks: (1) wildlife detection, which determines whether an animal exists in a picture, and (2) wildlife identification, which determines which species the animal objects belong to. In the area of image classification, it has been demonstrated that YOLOs outperform other techniques; consequently, in this paper, we focus on using contemporary state-of-the-art YOLO designs for both detection and recognition.Our suggested recognition system, as shown in Figure 4, consists of two YOLO-based picture classification models that correspond to the two tasks. The first YOLO-based model is used to train a binary classifier called Wildlife detector, while the second YOLO-based model is used to build a multi-class classifier called Wildlife identifier.
1) YOLO Architectures: Our suggested framework uses three YOLO architectures with varying depths: Lite AlexNet, VGG-16 [18], and ResNet-50 [21]. We employ Lite AlexNet, a modified version of AlexNet [20] with fewer hidden layers and feature maps at each layer. The Lite AlexNet, in particular, is made up of three 2-D convolutional layers with ReLU activations and MaxPooling, followed by two fully-connected layers: one with ReLU nonlinear activation plus Dropout for reducing overfitting, and an output layer with sigmoid activation for binary classification in detecting and softmax activation for multiclass classification in recognizing tasks. All of the convolutional layers have a modest filter size of 3 3 pixels, but all of the max-pooling layers have a window size of 22 pixels. VGG-16 and ResNet-50 are two examples of cutting-edge YOLO architectures that not only outperformed the competition on the ILSVRC [20], but also generalized well to other datasets [18], [21]. All YOLO designs take a fixed-size 224224 RGB colour picture as input.

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 12, 2021
www.ijirse.com
ISSN: 2454-9665

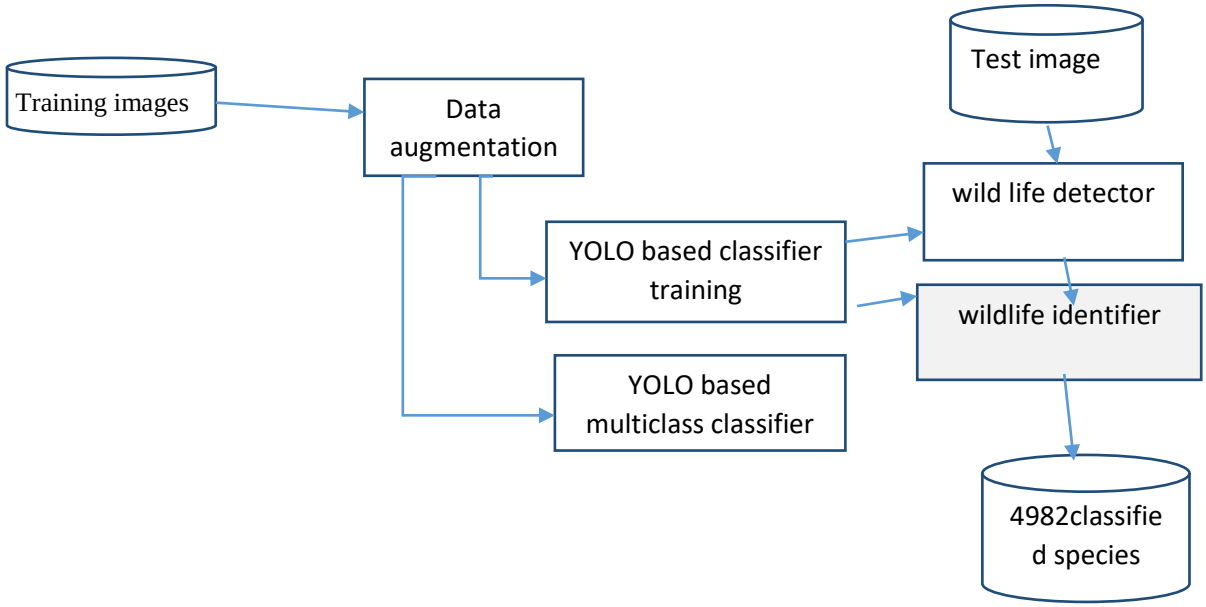

Figure 4: Key steps in the proposed framework for automated wild animal identification.

2) Image Processing: The Wildlife Spotter dataset includes high-resolution photos with resolutions of 1920 1080 and 2048 1536 pixels, however YOLO models require input in fixed dimensions. As a result, all original photos were downscaled to 224224 pixels for training in our trials. This technique was done out in [20] by first rescaling the image's shorter side to a set length, then centre cropping the image to the same length. For the sake of simplicity, we rescale both the width and height of the image at the same time in this work, which may cause visual distortion. The intensity of the pixels is normalized to the range [0,1]. Data quality, which may be improved with augmentation approaches, is critical for data-driven machine learning models; nevertheless, in this study, shearing and zooming were used as data augmentation procedures on training photos.

3) Deep Networks Training: We use Keras [37], a high-level neural networks API, in conjunction with Tensor Flow.[26] backend The Adam optimizer is a gradient-based first-order optimizer. All networks were trained using optimization based on adaptive estimations of lower-order moments [38]. a little amount All tests were run with a minibatch size of 16. We put our employees through rigorous training. Each network runs on four NVIDIA Titan X GPUs.It takes three to five days to complete training.We train YOLO models in two situations for each task: datasets that are both unbalanced and balanced We make a categorization. In both circumstances, precision is essential. Measure is used in addition to accuracy to verify the resilience of the proposed system in the situation of dataset imbalance. The validation set's accuracy is employed as a performance indicator. To test transfer learning, we run Task 2 – Wildlife Identification in two scenarios: building a model from scratch and fine-tuning existing ImageNet pre-trained models. Fine-tuning techniques rely on a network that has been pre-trained on a large dataset, in this case the ImageNet, on the assumption that such a network has already learned useful features for most computer vision problems, and thus can achieve higher accuracy than a model trained on a smaller dataset. Our fine-tuning procedure is divided into three steps: first, the convolutional blocks are created, then the model is trained on new training and validation data, and lastly, the fully-connected model with fewer defined classes is trained on top of the stored features [37].

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 12, 2021
www.ijirse.com
ISSN: 2454-9665

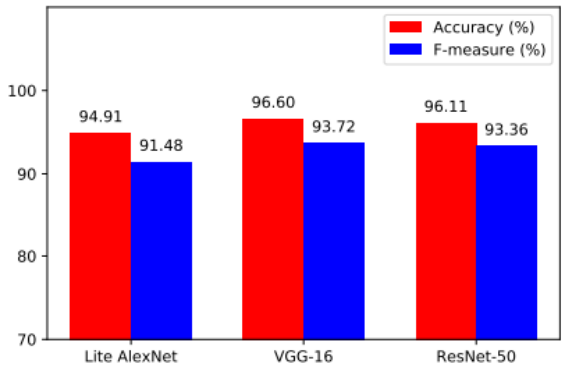

Figure 5: Animal vs. non-animal picture recognition accuracy on the South-central Victoria, Australia Wildlife Spotter dataset.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Results for Recognizing Animal vs. Non-animal Images

On the Wildlife Spotter unbalanced dataset, Figure 5 displays the performance of Task 1 with three alternative YOLO architectures. Overall, all of the models performed admirably. The greatest accuracy is 96.6 percent (with VGG16 architecture), with ResNet-50 coming in second at 95.96 percent, which is only slightly lower. These findings support previous findings that VGG and ResNet models generalise well to various datasets [18], [21]. Furthermore, Lite AlexNet, the most basic architecture with only five learnable layers, performed admirably on this binary classification job.Similarly, excellent F-measure metric performance implies that these models are resistant to unbalanced data. Table III illustrates the results on the balanced dataset presented previously to see if imbalanced data is a significant issue for our animal recognition task. As in the case of data imbalance, we maintain three YOLO designs the same. As can be shown, the performance of all models was only minimally deteriorated, which might be owing to the under sampling procedure, in which the samples of superior classes were significantly lowered in order to achieve a balanced dataset. This confirms the method's promise of being resilient and accurate when it comes to detecting photos with animals.

Table III: Image detection accuracy of animals vs. non-animals using the Wildlife Spotter dataset from South-central Victoria, Australia.The data is balanced, with 25,000 photos for training and 8,500 for validation in each class.

Model	Accuracy (%)
Litealex net	92.68
VGG-16	95.88
Resnet-50	95.65

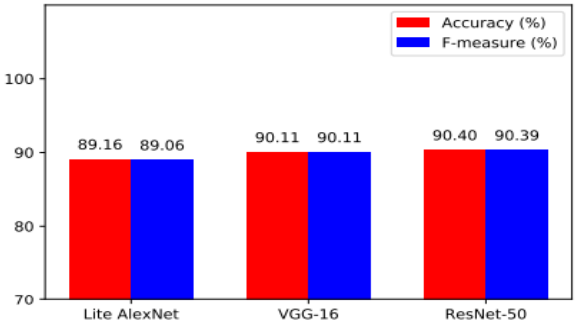

Figure 6: Animal identification accuracy for the three most prevalent species (bird, rat, and bandicoot) using the Wildlife Spotter dataset. As shown in Table II, the training set is unbalanced; 80 percent of each class's photos are utilized for training, while only 20% are used for validation.

ISSN: 2454-9665

B. Animal Identification Results

1) Recognizing the three most common species: In the event of an unbalanced dataset, all samples from these species are utilised to train the model for identifying the three most common species (bird, rat, and bandicoot). The training set comprises 80% of each class's pictures (35,629 photos), whereas the validation set contains the remaining 20%. (8,907 images). Figure 6 shows that the animal recognising task performed exceptionally well across all YOLO architectures, with accuracy ranging from 89.16 percent to 90.4 percent. The most basic model, Lite AlexNet, has the worst performance. ResNet-50, the deepest model, produces the greatest results, but VGG-16, the first runner-up, comes in second. Table IV shows the experimental results of identifying the three most prevalent species in the event of a balanced sample. All three YOLO models perform similarly when trained from scratch, with classification accuracy ranging from 87.80 percent to 88.03 percent. VGG-16 outperformed the others in this scenario, but by a little margin. Task 2 yields lower results for all models when compared to Task 1. The performance decline might be due to two factors: a more sophisticated issue created by an increasing number of classes, and a reduced number of training samples generated by the under-sampling procedure, making it more difficult to fit the models to the datasets.Because these models have pre-trained weights available on ImageNet, and our experimental findings reveal contrary tendencies, the fine-tuning approach was only used on VGG-16 and ResNet-50. While the VGG-16 model improved accuracy by 0.2 percent on a fresh dataset compared to training from scratch, the ResNet-50 model exhibits a significant decline in accuracy from 87.97 percent to 76.43 percent, indicating that overfitting may have occurred. The cost of computation is the most significant contribution the fine-tuning approach made to the framework.

2) Recognizing the six most prevalent species: We study further with the instance of recognizing the six most prevalent classes from the Wildlife Spotter dataset (bird, rat, bandicoot, rabbit, wallaby, and mammal) in order to develop a completely automated wild animal recognition system. We only investigate the scenario of an unbalanced dataset in this case; all samples from these six classes are used as mentioned in Table II; each class is split into two sets, with 80 percent photos used for training and 20 percent used for validation.

Figure 7 depicts the outcomes of the three YOLO architectures in terms of animal identification. For up to six distinct sorts of animal classifications, the technique produces reasonable results. Despite the tiny margin, the ResNet-50 looks to be the best model in this situation, with an accuracy of 84.39 percent, indicating that deeper YOLO architectures might perform better in complicated recognition challenges. Furthermore, the proposed system's resilience against unbalanced data is demonstrated by similar results obtained with F-measure metrics.

C. Discussion

The findings of the experiments showed that photos including animals can be detected with a high degree of precision, with more than 96 percent accuracy. Human annotators' time is precious in citizen science-based projects and systems, and this framework has the potential to dramatically increase the efficiency of these systems. For example, more than 32% of photos supplied to human annotators in the existing Wildlife Spotter system did not contain any animal to be analyzed at the time we gathered this information.

The animal identification findings showed that the most common three and six species of animals were correctly identified. While this performance may not be adequate to develop a completely automatic recognition system, it nonetheless contributes a lot of value to the system by giving early animal labels for human annotators. This performance might be greatly improved in the near future with more data acquired over time and a rapidly expanding capability of deep learning algorithms in computer vision.

Furthermore, we feel that the present dataset from the Wildlife Spotter might be used in a variety of ways.

International Journal of Innovative Research in Science and Engineering
Volume 7, Issue 12, 2021
www.ijirse.com
ISSN: 2454-9665

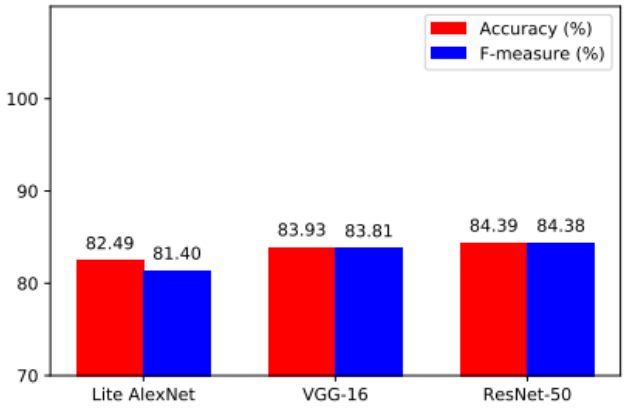

Figure 7: Animal identification accuracy for the six most prevalent species using the Wildlife Spotter dataset. As seen in Table II, the dataset is unbalanced. 80 percent of the photos in each class are utilized for training, while the remaining 20% are used for validation.

Mechanism for enhancing performance To improve the quality of the training data, the training dataset might be further produced and validated by ecological specialists, comparable to the Golden Standard established in the Snapshot Serengeti project.Furthermore, data improvement techniques can be used to improve outcomes; however, only a few data augmentation procedures were used in this study, as stated in Section III. These will be included in our next study and report.

CONCLUSION

Wild animal monitoring in their natural settings must be efficient and trustworthy in order to inform conservation and management choices. We presented and proved the viability of a deep learning technique to creating a scalable automated wildlife monitoring system using the Wildlife Spotter dataset, which comprises a huge number of photos recorded by trap cameras in South-central Victoria, Australia. Our algorithms correctly identified more than 96 percent of animal photos and nearly 90 percent of the three most prevalent creatures (bird, rat and bandicoot). Furthermore, the system has proven to be durable, reliable, and adequate for dealing with photos acquired in the field in various testing scenarios for balanced and unbalanced. By improving the dataset, using deeper YOLO models, and utilizing unique camera trap qualities, we are working on different approaches to increase the system's performance images. Towards fully automated recognition of wild animals system, we'd look into transfer learning to cope with it.Data that is excessively unbalanced is a concern. In the not-too-distant future, we will concentrating on the creation of a "hybrid" wild animal categorization framework with an automated module that acts as a recommendation system for current citizen science-based wildlife monitoring systems Project Spotter.

REFERENCES

[1] P. M. Vitousek, H. A. Mooney, J. Lubchenco, and J. M. Melillo, “Human domination of Earth’s ecosystems,” Science, vol. 277, no. 5325, pp. 494– 499, 1997.

[2] G. C. White and R. A. Garrott, Analysis of wildlife radio-tracking data. Elsevier, 2012.

[3] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, “An analysis of a large scale habitat monitoring application,” in Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, 2004, pp. 214–226.

[4] B. J. Godley, J. Blumenthal, A. Broderick, M. Coyne, M. Godfrey, L. Hawkes, and M. Witt, “Satellite tracking of sea turtles: Where have we been and where do we go next?” Endangered Species Research, vol. 4, no. 1-2, pp. 3–22, 2008. [5] I. A. Hulbert and J. French, “The accuracy of GPS for wildlife telemetry and habitat mapping,” Journal of Applied Ecology, vol. 38, no. 4, pp. 869–878, 2001.

[6] R. Kays, S. Tilak, B. Kranstauber, P. A. Jansen, C. Carbone, M. J. Rowcliffe, T. Fountain, J. Eggert, and Z. He, “Monitoring wild animal communities with arrays of motion sensitive camera traps,” arXiv:1009.5718, 2010.

ISSN: 2454-9665

[7] A. F. O'Connell, J. D. Nichols, and K. U. Karanth, Camera traps in animal ecology: Methods and Analyses. Springer Science & Business Media, 2010.
[8] A. Swanson, M. Kosmala, C. Lintott, R. Simpson, A. Smith, and C. Packer, "Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna," Scientific Data, vol. 2, p. 150026, 2015. [9] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," Nature, vol. 381, no. 6582, p. 520, 1996.
[10] X. Yu, J. Wang, R. Kays, P. A. Jansen, T. Wang, and T. Huang, "Automated identification of animal species in camera trap images," EURASIP Journal on Image and Video Processing, vol. 2013, no. 1, pp. 1–10, 2013.
[11] G. Chen, T. X. Han, Z. He, R. Kays, and T. Forrester, "Deep convolutional neural network based species recognition for wild animal monitoring," in Proceedings of the IEEE International Conference on Image Processing (ICIP), 2014, pp. 858–862.
[12] A. Gómez, A. Salazar, and F. Vargas, "Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks," arXiv:1603.06169, 2016.
[13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 248–255.
[14] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, no. 3, pp. 211–252, 2015.
[15] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?" PLOS Computational Biology, vol. 4, no. 1, p. e27, 2008.
[16] C. M. Bishop, "Pattern recognition," Machine Learning, vol. 128, pp. 1–58, 2006.
[17] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Computation, vol. 1, no. 4, pp. 541–551, 1989.
[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556, 2014.
[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.