A Survey on Augmentation of Computer Vision and Image Processing Techniques with Zoological Research for Anatomical and Speciating Speculations

Mr.Aakash Ravindra Shinde[1], Mr.Kshitij Hemant Patil[2], Ms.Praneeta Gopal Dumbre[3], Ms.Ruchira Kailas Borkar[4], Dr.Amol V. Dhumane[5]

[1-4]*Pursuing Bachelor in Computer Engineering, SPPU*
[5]*Head of Department, NBNSSOE, SPPU*

Abstract— *One of the powerful sensory tools provided to mankind is its ability to see which helps us to visualize and distinguish, these sensory prowess are applied to artificial intelligence-based systems over the virtue of Image recognition and Computer Vision technology proving eyes for the computers. Image-based recognition systems are being developed for more than a decade and have been profoundly useful in multiple technological application like image-based language translating devices, facial detection, etc. With an introduction to better processing power AI systems have better used them in fields providing better models for image processing projects and their development. It would be farfetched to see this revolutionary technology making its way into zoological research helping in studying and conserving fauna around this planet. This paper reviews various manuscripts providing a solution to three major problems identification of animals in the image and its species recognition, pose estimation for animals and individuals' recognition in Patterned Species, and provide a solution inculcating computer vision, AI, and Image processing with zoological aspect. This helps in understanding the ways in which we can use technology for the conservation of the living organisms on this planet.*

Keywords: Computer Vision, Image Recognition, Pose Estimation, CNN, Deep Neural Network, Pattern based Recognition.

1. INTRODUCTION

Wildlife conservation is fundamental for a balanced ecosystem, wildlife monitoring is hence essential for tracking animal habitat location, migratory patterns, the population in certain areas, poaching incidents, etc. Human interaction with nature has transformed an increasing area of the land mass, altered wildlife population, desolated habitats, and transformed animal behavior. This has driven several species closer to extinction, and migration into new areas has been acquitted by several species causing disruption in the local ecology [1]. To understand the natural ecosystem complexities and better management of the fauna, extensive knowledge about the numbers, behavior, and location of animals in the ecosystem is vital [2]. Documentation and research of the zoological entities have been done pictographically for an extensive amount of time and with a significant number of documentaries regarding the same, use of image-based methods for processing of this vast amount of data could beneficial for research and conservational purposes [3]. Several technologies are being introduced for the assessment of the wildlife and gathering of essential data, like RFID tags for radio tracking [4]. GPS-based tracking equipment [5][6][7], motion-sensitive cameras, wireless sensor network [8][9], satellite for tracking animals, animal-mounted video [10], etc. These technologies have been beneficial in many instances for solving tracing problems and gathering essential information but they require a lot of human interaction for analyzing the data as most of the work is done manually. Even so with certain scenarios concerning RFID tags, GPS based devices there is a need for physical interaction with the entity and attaching of devices onto them. Trap cameras and motion-activated devices are installed at a location that does collect a vast amount of data without any physical contact with the entity but still manual labor is required for analyzing the collected data which requires a tremendous amount of time viewing, analyzing, and processing collected data for any productive utilization [11]. Furthermore, data gathered collected by different individuals and research groups using cameras and other sensory devices are scattered in different location and collected at time, consisting of varying formats and sizes, or sequestered [10].

To overcome several of these drawbacks of tracking systems and their human interaction concerning the longevity of the process various ways are proposed in order to be helpful for conducting research and gathering valuable information. We are reviewing several of the proposed solutions proving to be helpful in zoological research, these reviewed solutions venture into the domain of AI and image processing. In this manuscript we shed light on several experiments concerning problems like detection of animals in images, pose estimation of animals, the total number of individual entities in the image, and pattern-based individual animal detection from available data. We also are reviewing several technical terms in relevance to AI models, Image Processing, and Computer Vision giving an idea of techniques used for the execution of solutions to the problems faced. The structure of this manuscript consists of two sections: Section 1 giving and brief about the technology and techniques that are essential and repeatedly used for the execution of image-based projects. Section 2 concerns with three major problems and

their solution proposed over several research papers and evaluating their proposed solution for their merits and demerits. Section 3 covers various drawbacks for implementation of computer vision systems and what future prospects could be gained for further research and all the system improvements required for the betterment of the technology. Finally, the manuscript provides with conclusion over the discussed methods, approaches, and solutions.

II. SECTION 1

Section 1 provides with a technical overview of the technologies and common techniques used in the solution discussed in the section 2 later. This section would provide with a brief introduction to various concepts and topics that would prove helpful in understanding this papers purpose. A brief idea regarding the techniques used for executing the proposed solution by the papers reviewed in this manuscript, is provided in this section.

A. Artificial neural networks (ANNs)

ANN also termed as neural networks (NNs) consists of inter-connected nodes called artificial neurons, which are loosely based on the neurons present in a human brain. An artificial neuron processes signal which can be considered as a real number that neurons receives and transmits information to connected neurons. The output generated by neuron is computed by some non-linear function of the sum of its inputs. Neurons and edges(connections), based on weight assigned adjusts their learning process, increase or decrease in strength of the signal at a connection is dependent on weights assigned. Into several layers neurons are aggregated, different layers in a NN model performs several transformations on its inputs generating results dependent on these layers [12].

B. Convolutional Neural Network (CNN)

CNNs are non-linear regression models optimally selected for performing complex prediction tasks, learning via optimization from real-world data using set of parameters which are hierarchically–organized. Depending upon the way in which a CNN model is specified and trained, both advantages and disadvantages could be encountered hence a CNN model for prediction task must be well selected and optimized. The possibilities of architectural models and optimization schemes are pretty extensive for CNN because of its flexibility, it is favorable for adapting to any computer vision task [14].

C. Deep Convolutional Neural Network (Deep CNN)

Classification of hyperspectral images directly in spectral domain is done by using Deep convolutional neural networks which is a typical feedforward neural networks[15]. Deep CNNs are basically BP algorithms for adjustments of the parameters (weights and biases) for reduction of the cost function value in a network. What so ever, compared with the traditional BP networks it is quite difficult to bifurcate in four new conceptions: shared weights, local receptive fields, combination of different layers, and pooling [16].

D. Support Vector Machine (SVM)

SVM is a supervised machine learning technique best known for pattern and image classification. SVM is Inception-v3, a convolutional neural network developed by Google and used for objects classification that consists of multiple layers. This model is one of the pre-trained models on the TensorFlow environment and uses the dataset that is already trained on ImageNet model. Transfer learning is used to fine tune the top layer of the model and avoidance of over-fitting in the model for data augmentation technique is applied. Fine-tuning and data augmentation techniques were employed to improve the recognition accuracy for each individual category of animals [3].

E. Data Augmentation

Deep neural network is best compared to the traditional machine learning algorithms. However, over-fitting results in performance degradation of the model. Data Augmentation (DA) is the most commonly used technique to enhance the accuracy of the model by increasing the amount of training images and also avoids over-fitting in image classification tasks which are as important in reducing errors. A data augmentation technique includes scaling, translation, rotation, flipping, and adding noise to the images [3].

III. SECTION 2

Previous section overviewed major concepts, ideas and models essential for understanding the implementation of solutions proposed for the problems mentioned in this section. In this section we are going to focus on three problem statements overviewed by many and that have seen several solutions using the computer vision. Section 2 covers these solutions, as they provide with a solution that provides with a near human accuracy in solving the problem and requires fewer manual hours to be spent in processing data to generate same information. Three problem statements focused are:

i. Identification of animal in image and its species recognition

ii. Pose estimation for animals

iii. Recognition of Individuals in Patterned Species

Each of these problems show greater importance in zoological research-based phenomenon and solving those manually takes a lot of manual hours in process. Section 2 sheds light on these problems and discuss various implementations their virtues and drawbacks over each other.

A. Identification of animal in image and its species recognition

Positioning motion sensor and motion activated cameras in animal habitats for over last two decades has benefitted wildlife conservation and ecology also helped in understanding the behavior of animals [17]. Motion-sensing cameras enabling ecologist to study population sizes and understanding population distribution over an area, resulting to become an essential tool for ecologists [18] ,cameras also prove helpful for evaluation of habitat usage [19]. In order to generate data form those devices one needs to go through loads of data in order to use that data for research hence a person needs to sit and evaluate each frame identifying the animal or entity in the frame and also to recognize the species of the animal. To overcome this tedious task many algorithms

and model-based solutions have been suggested In [2] et al. using 3,200,000 images from Snapshot Serengeti (SS) dataset they trained deep CNN for identifying, counting and describing the behaviors of 48 species of animals. As per [2] et al. for labeling half a year batch of collected images form the SS, requires couple of months for a group of thousand "citizen scientist" to label each and every image. Computer Vision's power was harnessed by [2] et al. in order to automatically extract information like species, number of entities in image, younglings in the image and behavior (i.e., eating, moving or resting) of the animal based on the prelabeled SS data, which is seen to be a challenging task for humans as well. They automated information extraction from motion sensing camera images by combining modern supercomputing, deep neural network (DNN) and the data available from the SS project dataset. SS project with continuously running 225 camera traps in Serengeti National Park, Tanzania, since 2011 is the world's largest published camera-trap project up till today [20] which contain 1.2 million capture events of 48 different species. As per research presented by [2] et al. almost 75% images were empty events this shows the time invested for just classifying the empty image set for human would be catatonic and as the SS dataset is labeled by volunteers' human error was also seen to be marginal, also it was found that using individual images ended up with more accurate results as there were 3 times more labeled data for training example. [2] et al. trained their model using the multitask learning technique and using two stage pipelines in first stage network solves the task of whether the animal is present or not while in second stage network handles the task of reporting information of the image. This method proved to be beneficial as there was relation among tasks and they could share weights that encode features common to all tasks, also this resulted in fewer model parameters solving the tasks quicker and being extensively energy efficient, and simpler to store and transmit. In [2] et al. nine different modern DNN architectures were tested to find highest performing network which were AlexNet consisiting of 8 layers, NiN with 16 layers, VGG architecture with 22 layers, 32 layers for GoogLeNet Architecture, and ResNet variants of 18,34,50,101,152 layers.

For model training and testing datasets contained 1.4 million and 105,000 images respectively. As per stated in [2] et al. for classification of presence or absence of animal in image all models were able to achieve accuracy > 95.80% out of which VGG model got the best accuracy. Following the Table 1. Provides with the models' accuracy. Top-1 and Top-5 accuracy are two traditional computer vision parameter fields, Top-1 accuracy corresponds to only correct value prediction at top with highest prediction percentage and Top-5 accuracy corresponds with correct value prediction at top 5 places in regards to prediction percentage. In [2] et al. for identification of species the accuracy of ensembled model resulted to be top-1 with accuracy of 94.9% and top-5 accuracy to be 99.15%. For counting of animals in an image et al. represented numbers by classifying them into values as 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11–50, or +51 individuals, which resulted in top-1 accuracy around 63.1%, and predictions percentage were 84.7% within ±1 bin. This experimentational results showcased system saving 99.3% of manual labor (>17,000h) just for identification of animals while performing similar to human volunteers 96.6% accurate this was achieved with limited data. Similar performance was seen for detection of species.

Architecture	Top-1 accuracy, %	Architecture	Top-1 accuracy, %
AlexNet	95.8	ResNet-34	96.2
NiN	96.0	ResNet-50	96.3
VGG	96.8	ResNet-101	96.1
GoogLeNet	96.3	ResNet-152	96.1
ResNet-18	96.3	Ensemble of models	96.6

Table 1. Accuracy of different models on detecting images that contain animals [2].

Similar to research reviewed before for wild animal classification, species recognition algorithm was proposed by[10] et al. on camera-trap imagery data, for which they devised a novel deep CNN. They collected and annotated over 14,346 training and 9,530 testing images for discrimination of 20 common species in the North American continent using a standard camera-trap image dataset for training the proposed model. For this research [10] et al. physically installed camera trap over more than 1,000 locations with help of group of volunteers for analysis of behavior of animals and monitoring their population in the vicinity. For development of the proposed deep CNN-based species recognition algorithm, bag of visual words model as the baseline model [21][22]and was used for comparison.

Region of interest (ROI) are selected by the tight bounding box around the segmented region, treating the species recognition problem as image classification problem on the ROI [10]. Two major algorithms for the process to resort to were a model-based image classification algorithms named Bag of visual Words (BOW) [21][22] and image classification algorithms [23] based on Deep Convolutional Neural Network (DCNN). In consideration with the availability of the training dataset size [10] et al. opted for a 3 convolutional layered and 3 max pooling layered DCNN. Each layer consisted of 9×9 size convolutional kernels, while the pooling layers has 2×2 size kernels. Input size by [10] et al. consisted of 128 × 128 layers size. Data collected over camera-trap was split into training images of size 14,346 and around 9,530 testing images consisting of 20 species for classification which are Spiny Rat, White-nosed Coati, Ocelot, Red Squirrel, Tinamou, Mouflon, White Tailed Deer, Red Deer, Agouti, Collared Peccary, Paca, Common Opossum, Bird spec, Red Brocket Deer, European Hare, Wood Mouse, Wild Boar, Red Fox, and Coiban Agouti. For Bag-of-words models classification [10] et al. divided image into overlapping 8 by 8 small blocks getting a block of every 3 pixels. The BOW model accuracy of 38.315% overall for species recognition DCNN and accuracies for K = 1000, 2000, 3000 are around 33.192%, 33.507%, and 33.485% respectively, [10] et al. also emphasized on the high learning capacity of DCNN and its performance could be improved with more availability of data. Even if the results do not provide with a jaw dropping surprise with great accuracy but considering the data and size of the available data results are proven to be promising.

For construction efficient monitoring systems, speed up research findings and subsequent management decisions for the world of ecology [24] et al. proposed a high-tech deep CNN architecture for animal images filtration and species identification automatically by training a computational system using a Wildlife Spotter project's single labeled dataset. For implementation of the idea [24] et al. faced two primary challenges, first is to attain applicable accuracy for image classification, manual preprocessing is still required in an enormous amount to input images for bounding and detecting animal objects [25] et al. and second is unsatisfactory performance from the wildlife monitoring system, much more improvements are required in spite of complete automation in regards to practical application [26]. For designing an animal recognition in the wild framework,[24]et al. divided the task into two parts: (1) Detection of wildlife, which is basically a binary classification on existence of the animals in the designated images; and (2) Identification of wildlife, for which a multiclass classifier is for labeling and specifying the species in each input image with presence of animal in detected images [24].Wildlife Spotter dataset of South-central Victoria dataset [24] provides a lists of 108,944 labeled images verified by 5 different citizen scientists. For classification problem [24] et al. considered six of the most common listed species in Wildlife Spotter dataset were considered of which 80% data was used for training and 20% data for validation. Wildlife Spotter labeled dataset consists of 67.74% animals and 32.26% no animals' images with labels majorly consisting of Bird, Rat, Bandicoot, Rabbit, etc. summing the total images available up to 72498 images of varying species. For addressing two tasks [24] et al. created two models one for each task to train a binary designed on the basis of a CNN classifier model is named as Wildlife detector and another named Wildlife identifier using a multiclass classifier. Three CNN architecture with variable depth were employed on the proposed framework by [24] et al., specified as VGG-16 [27], Lite AlexNet and ResNet-50 [28]. The Wildlife Spotter dataset consist of 1920 × 1080 and 2048 × 1536 pixels high resolution images, whereas fixed dimensions are important for input in the CNN models. Therefore, [24] et al. downscaled original images to 224×224 pixels for training. All proposed models by [24] et al. for animal or no animal image detection achieved very high results, with VGG-16 architecture had best accuracy of 96.6%, while ResNet-50 has accuracy of 95.96%, even Lite AlexNet consisting of 5 learnable layers showed better results of 92.68% accuracy. Very high performance was achieved for identification of 3 most common species with accuracy ranging from 89.16% to 90.4% for all CNN architectures. Identification of 6 most common species problem using the similar model also proved to be working exceptionally well. In this case presented the best model with accuracy at 84.39% appears to be ResNet-50, even with small gap, better performing Deep CNN could be developed for complex recognition problems [24]. Models proposed by [24] et al. for recognizing images achieved more than 96% and bird, rat and bandicoot are three most common animals in dataset for which identification close to 90% was achieved. While this performance may not be extraordinary but promising enough to add greater value in improvisation of the system by automatically labeled animals for human annotators [24].

Concerning with the monitoring of animal behavior along with recognition of the species of the animal [29] et al. proposed an extension of CNN and VGG split into three branches one for the whole shape recognition using VGG19 model and for the muzzle and part of shape recognition two VGG16 models. Experiments were conducted by [29] et al. using the Ergaki national park produced dataset. After analysis of the dataset and exclusion of non-recognized images [29] et al. deciphered and discovered on four major sublets first sub-set consist of well represented animal muzzles, second sub-set contained good animal shapes representation, a part of the shapes were held in third sub-set, and the whole objects were involved in forth sub-set of images. Based on the prior observations [29] et al. devised a procedure, as per the mentioned sub-sets categorizes the images, and CNN architecture combining three parallel branches, separately recognizing the part of shape, whole shape and muzzle. Experiments were conducted by utilizing the TensorFlow 1.5 framework encoded in Python 3.5 that supports CNN and conducted on a 3.2 GHz, i7 CPU, RAM size:16 GB and 8 GB graphics drive with Windows 7 operating system [30]. Ergaki national park, Russia ,2012-18 dataset consisted of more than 40,000 images captured in both daylight and night time, all season and with varying picture quality. Whatsoever, approximately rejection rate of

images was 30% due to reasons of being void, consisting of bad quality or unknown.

[29]carried out the first experiment was conducted with close to 28,000 total images using the highly unbalanced E1 dataset. Using dataset E2 second experiment was carried out by [29] et al. and dataset was forged by synthesizing more samples for species from small number of animal images. After which, the unbalanced and balanced datasets and were divided into 80% and 20% for training set and testing set respectively. This resorted that [29] et al. from unbalanced dataset E1 for the CNN training they utilized 22,400 images and5,600 images were utilized for the CNN testing. Form balanced dataset E2 for the CNN training 27,200 images were utilized and 6,800 images were used for the CNN testing as dataset consisted of foreground images only[29]. Testing process proved that dataset E2 provided with better accuracy compared to E1 with Top-1 accuracy of 80.6% and Top-5 accuracy of 94.1%. The best results obtained were 80.6% Top-1 and 94.1% Top-5 accuracy respectively for the balanced dataset. They obtained Top-1 accuracy of 38.7% and Top-5 accuracy of 54.8% for the unbalanced training dataset by [29].

B. Pose estimation for animals

Pose estimation corresponds with estimating the physical posture or skeletal representation of an animal based on the image provided. Pose estimation is important from several aspects as animals provide with a wide range of motion and pose estimation may provide with vital information regarding their anatomical behavior. Pose estimation has focused on the human perspective have been implemented several times providing with promising results [31][32][33][34]. Pose estimation in regards with animals provides with a numerous usage in fields of zoology, ecology, biology and entertainment. Understanding posture of animals might prove helpful for animation and digital processing of CGI and creating life like imagery from the researched phenomenon. For measuring behavior pose estimation is an essential tool, and thus widely used in technology, medicine and biology [35].

Major problem for devising an ML algorithm for pose estimation is the lesser availability of data regarding to prep and device an efficient algorithm. To overcome this problem [36] et al. built a dataset of animal poses for training and evaluation. As the task of labelling each and every species and their pose estimation would prove to be rather tedious [36] et al. proposed for rather a novel method of cross-domain adaptation for transforming knowledge of animal pose from labeled to unlabeled animal classes. [36] et al. started formation of the model based on anthropomorphic data and then designed a WS-CDA (weakly- and semi supervised cross-domain adaptation) scheme is utilized for elicitation of common cross-domain features which consisted of trifactor parts namely keypoint estimator, domain discriminator and the feature extractor. From input data features are extracted by feature extractor, the keypoint estimator predicts the keypoints while the domain discriminator segregates the domain they come from. Dataset majorly consisted of four legged mammals with five classes namely dog, cat, sheep, horse, cow. A dataset of VOC2011 for pose-labeled instances is publicly available on [37][38] building on which and using the five selected classes it proved to be helpful for process as mentioned in [36] et al., for better leveraging knowledge from human data, align annotation format with a popular human keypoint proved to be helpful [36]. Dataset consisted of 5517 instances of the specified 5 classed with image distribution of more than 3000 pictures[36]. Such animal pose annotation can be aligned to that defined in popular COCO dataset by selecting within 17 keypoints [39]. 18 bones were defined by [36] et al. for having similar explanation as on COCO dataset. Even calculating the relative length and taking average of the various described classes. [36] et al. design WS-CDA (Weakly- and Semi- Supervised Cross-domain Adaptation) to better learn cross-domain shared features and scheme to alleviate such flaw later they introduce to PPLO (Progressive Pseudo-Label-based Optimization) strategy referring to 'pseudo-labels' for data augmentation boosting the model's performance on target domain. The final model is boosted under PPLO and pre-trained through WS-CDA [36]. WS-CDA consisted of four modules, first all data is fed to feature extractor that generates the feature maps, second the feature maps passing through domain discriminator segregates the input feature map generated form specified domain, in next module from the pose-labeled samples for supervised learning of pose estimation keypoint estimator receives the feature maps, finally to convert the feature maps a domain adaptation network is inserted for better representation in pose estimation on animal instances [36]. A "progressive pseudo-label-based optimization" (PPLO) was designed by[36]et al. a self-paced 'pseudo-label' selection method and an alternating training method are introduced to boost model performance by bringing target domain data into training with 'pseudo-labels'. Using these techniques [36] et al. provided us with a dataset with a novel prior knowledge to generate algorithms for training and creation of similar purposed methods.

Pose estimation for animals that are currently available have several limitations like speed, robustness, and usability. [14] et al. introduced an open-source software toolkit, DeepPoseKit, for addressing the limitations mentioned. Experiments were conducted on several dataset by [14] et al. comparing the proposed method with [40] and [41]et al., confirming the results of performing considerably better than the comparing parameter. [40] and [41] et al. were pioneers to introduce and propagate the use of Convolutional Neural Networks for estimation of pose in animals. Models introduced by them were used for measuring the postures of animals by network training, leading to transformation of image into plausible estimate of keypoints location stated as confidence maps describing the body postures of one or more individuals. [14] et al. devised a two model-based implementation consisting of a novel model named Stacked DenseNet and another method named

as subpixel maxima which resulted in fast and accurate results with position of subpixels. Solution proposed by [14] et al. the DeepPoseKit built on the popular Keras deep-learning package and is written using the Python programing language [42] using TensorFlow as a backend [43]. Three experiments were conducted by [14] et al. for testing and optimization of the model. First [14] et al. compared integer based global maxima with subpixel maxima layer using Stacked DenseNet model. Further [14] et al. tested training for improvement in accuracy by using the model predicting the global geometry, finally model was compared with models from [41] and [44]. Dataset used for experimentation consisted of using the "fruit" fly dataset provided by [41] et al. also compared model performance using previously unpublished posture data sets consisting 240 groups of desert locusts filmed under laboratory setting, 241 and herds filmed in the wild of Grévy's zebras. Hence [14] et al. presented DeepPoseKit a toolkit for automatically measuring animal posture in images.

[35]et al. developed a 30 horses novel dataset that allowed for both "within-domain" and "out-of-domain" (unseen horse) benchmarking, also [35] et al. probed the generalization ability with three architecture classes [35] et al. developed a novel dataset consisting of 30 Thoroughbred horses labeled by experts in 8,114 images for each of their 22 body parts. Complexity was added by adding collection of Thoroughbred from several farms with various coat colors.
[35]et al. provided with 2 major insights first ImageNet performance generality for both on out-of-domain data and within domain for estimation of pose and second, while [35] et al. confirmed that task-training can catch up with fine-tuning pre-trained models given sufficiently large training sets [35] et al., showing this is not the case for out-of-domain data. Thus, transfer learning improves robustness and generalization. Dataset developed consisted across 30 different horses captured with around 8114 frames for 4-10 seconds videos using GoPro cameras . [35] et al. created 3 divisions with pictures of 10 randomly selected training horses each. we took a subset of 5% for each training set, and for training 50% of the frames, and then performance on the training, test, and unseen horses was evaluated the [35] et al. recorded two major findings, first pretrained ImageNet networks offer known advantages: less data requirements, shorter training times and as well as a novel advantage: rugged on out-of-domain data, & better generalization was observed with networks that have higher ImageNet performance, if pretrained.

C. Recognition of Individuals in Patterned Species

Tracking and determining individuals is one of the important and difficult tasks for many biologist and zoologist for which several methods are used which are far more tedious and requires physical contact with the animals, this also bring in concern with animals that resent human approach. Solution as far specified comes with the solution for the image-based recognition which still is complex for finding of a particular individual from the herd captured and all the available data. Ecologist and evolutionary researchers modulated a way to determine individuals in patterned a better way [45] et al., as a fingerprint is unique to humans similarly the body pattern of animals like zebras, tigers and leopards are unique to those certain species. Using this feature along with image-based recognition it would be easy to collect data of individual patterned animals. This inhibit the use of methods like radio collaring, GPS tracking and field manual monitoring, these approaches minimize the subjective bias are less stressful, cost effective, safer and repeatable for the human and animal [46].

To tackle this problem [46] et al. used Faster-RCNN object detection framework for efficient detection of animals in image further features are extracted by using animals flank's AlexNet and trained a classifier for individual identification based on logistic regression. [46] et al. primarily evaluate and tested proposed framework on a camera trap tiger image dataset that contains images that vary in overall image quality, animal pose, scale and lighting also evaluating proposed recognition system on zebra and jaguar images for showcasing generalization to other patterned species. Addressing the problem [46] et al. split the process into two halves first detection and localization of patterned species in a camera trap and later uniquely identifying the entity of the same species over the existing database. Figure 1 below describes the architecture used in the process in a subtle manner.

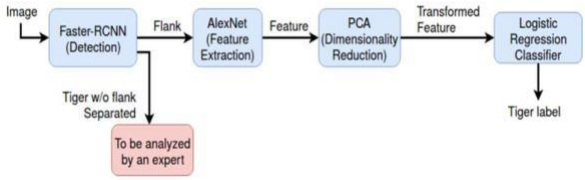

Figure 1. Proposed framework by [46] for detection of animal and entity recognition.

Experimentation was done by [46] et al. on i7-4720HQ device with Nvidia GTX-950M and GPU 3.6GHz processor, programmed on Python programming language. For training three different datasets were used and generated model and compare results with HotSpotter [47] et al., as compared to Wild-ID suggested model showed superior performance [48] and StripeSpotter [49]. Three datasets used were Tiger dataset, Plains Zebra dataset and Jaguar dataset. Tiger dataset consisted of 770 images, acquired from Wildlife Institute of India (WII) generated from camera traps. StripeSpotter dataset consisted of Plains Zebra [49]. As compared to tiger the stripe patterns in jaguars are less discriminative, the images in this dataset have appearance variations and little viewpoints several images were taken seconds apart from each other [46].

As described in [46] et al. the division of the data was a split of 75% and 25% respectively for training and testing

for a disjoint set of tigers and a total of 1032 (516 × 2) images used for data augmentation in the training set and testing set consisting of 171 images were generated. Accuracy was compared among 227×227 images on proposed model, resized images on the same proposed model and HotSpotter model. For Tiger dataset had 76.5 ± 2.2 for 227×227 set and 80.5 ± 2.1 for resized version and 75.3 ± 1.2 for HotSpotter, 73.5 ± 1.8 was acquired accuracy on the Jaguar dataset on 227×227 set and 78.6 ± 2.3 for the resized version while HotSpotter achieved 92.4 ± 1.1 accuracy on the same dataset, finally for the Zebra dataset 227×227 images set acquired an accuracy of 91.1 ± 1.2 as for resized version had accuracy of 93.2 ± 1.4 and HotSpotter with accuracy of 90.9 ± 0.8. [46] et al. utilized the CNN based object detector Faster-RCNN [50] and for the purpose of detecting the whole body and the flank of the tiger fine-tuned it, from a pre-trained AlexNet they further used the detected flanks and extracted features [23] to train a logistic regression classifier for classification of tiger individuals similar processing was also performed on zebras and jaguars for individual recognition task.

IV. SECTION 3

In previous section we discussed several ways in which use of Computer vision and Image processing techniques prove to be beneficial for solving various problems by showing ways to element physical involvement of humans for gathering data of animals, solution assisting humans for doing task in lesser time than previous man hours needed to be invested and helped in ways of understanding the anatomical behavior. Even with so many perks for implementation of the solutions suggested, there are certain drawbacks in order to create and implement of systems with at human par capabilities for eliminating the human intervention in data collection part. Major problem faced by systems for improvement is the lack of available data. To be specific data documented from zoological person is not limited but in vast quantity but isn't well structured as there is no well revised, processed data for training of models. Models could be improved exponentially well with introduction of better training data and dedicated data for various solutions in order to display better results in various defined tasks. Most of the image-based solution stated so far only limited themselves with using the data from camera traps only using the public generated data for trail might make the model more diverse and perform on a better scale for wide range of images set. Identification of individuals in a species is limited with patterned species only as lesser discriminative features are available for pattern less species. Other ways needed to be devised for identification of individual entities in image and cross verifying them with a centralized database which would prove helpful for migratory species data collection.

Even with several of the mentioned setbacks for the systems hope of betterment resides in future solving the current problem. With utilization of more powerful processing devices better algorithms can be devised working on them improving the reaction time and training time could also be similarly reduced. Using the pose estimation techniques and the individual detection of patterned animal certain algorithms could be devised that might prove helpful for creating algorithm that helps to identify individuals in pattern less specimens. Improvement in the camera quality might be helpful to use more sharper images for training models by including those in training datasets. Use of the anatomical models would prove extremely helpful for animation and CGI as graphics could be generated more easily rather than physically tracking the motion of animals for the same process.

CONCLUSION

Through this manuscript we tried to understand several applications of Computer Vision in fields of zoology, ecology and biology, and ways they could prove helpful for research and developments in these fields. We focused on three major problem statements which are identification of animal in image and its species recognition, pose estimation for animals and individual recognition in Patterned Species, narrowing the perspective too surveillance and tracking of the wild animals. The papers reviewed provide with a wide range of solutions for the problems and through light on the limitations noticed and ways that could improve the proposed solutions if limitations are overcome. Major problems faced are mostly regarding with the collection and creation of data to train models upon, with proper labelled and ample images better models could be created. Solution proposed in the papers reviewed so far are beneficial for eliminating a lot of time spent for data preprocessing by humans but lack in certain aspects accuracy to be completely human independent a certain level of human monitoring is still seeming too be necessary. Novel ways of research could be developed over using of image-based research rather than having physical intervention with animals and mounting equipment just for acquiring data. These ways of research would also prove beneficial for prevention of poaching activities by keeping track of the animals using surveillance equipment. Experimenting and researching in these fields might result in conservation of fauna and deepen our understanding of the creatures residing on this planet.

REFERENCES

[1] P. M. Vitousek, H. A. Mooney, J. Lubchenco, and J. M. Melillo, "Human domination of Earth's ecosystems," Science (80-.)., vol. 277, no. 5325, pp. 494–499, Jul. 1997, doi: 10.1126/science.277.5325.494.

[2] M. S. Norouzzadeh et al., "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," Proc. Natl. Acad. Sci. U. S. A., vol. 115, no. 25, pp. E5716–E5725, 2018, doi: 10.1073/pnas.1719367115.

[3] R. Thangarasu, V. K. Kaliappan, R. Surendran, K. Sellamuthu, and J. Palanisamy, "Recognition of animal

species on camera trap images using machine learning and deep learning models," Int. J. Sci. Technol. Res., vol. 8, no. 10, pp. 2613–2622, 2019.

[4] L. D. Mech, "A Handbook of Animal Radio-Tracking." 1983.

[5] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking," ACM SIGPLAN Not., vol. 37, no. 10, pp. 96–107, Oct. 2002, doi: 10.1145/605432.605408.

"Animal Tracking Basics - Jon Young, Tiffany Morgan - Google Books." https://books.google.co.in/books/about/Animal_Tracking_Basics. html?id=HTJMTit3OJ8C&redir_esc=y (accessed Jan. 02, 2021).

[6]"Animal Tracking Basics - Jon Young, Tiffany Morgan - Google Books." https://books.google.co.in/books/about/Animal_Tracking_Basics.html?id=HTJMTit3OJ8C&redir_esc=y (accessed Jan. 02, 2021).

[7] I. A. R. Hulbert and J. French, "The accuracy of GPS for wildlife telemetry and habitat mapping," J. Appl. Ecol., vol. 38, no. 4, pp. 869–878, Aug. 2001, doi: 10.1046/j.1365-2664.2001.00624.x.

[8] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey." Accessed: Jan. 02, 2021. [Online]. Available: www.elsevier.com/locate/comnet.

[9] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, "An analysis of a large scale habitat monitoring application," in SenSys'04 - Proceedings of the Second International Conference on Embedded Networked Sensor Systems, 2004, pp. 214–226, doi: 10.1145/1031495.1031521.

[10] G. Chen, T. X. Han, Z. He, R. Kays, and T. Forrester, "Deep convolutional neural network based species recognition for wild animal monitoring," in 2014 IEEE International Conference on Image Processing, ICIP 2014, Jan. 2014, pp. 858–862, doi: 10.1109/ICIP.2014.7025172.

[11] G. Harris, R. Thompson, J. L. Childs, and J. G. Sanderson, "Automatic Storage and Analysis of Camera Trap Data," Bull. Ecol. Soc. Am., vol. 91, no. 3, pp. 352–360, Jul. 2010, doi: 10.1890/0012-9623-91.3.352.

[12] "Artificialneural network - Wikipedia." https://en.wikipedia.org/wiki/Artificial_neural_network (accessed Jan. 02, 2021).

[13] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553. Nature Publishing Group, pp. 436–444, May 27, 2015, doi: 10.1038/nature14539.

[14] J. M. Graving et al., "Fast and robust animal pose estimation," bioRxiv, p. 620245, 2019, [Online]. Available: https://www.biorxiv.org/content/10.1101/620245v1.

[15] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," J. Sensors, vol. 2015, 2015, doi: 10.1155/2015/258619.

[16] X. Li, G. Zhang, K. Li, and W. Zheng, "Deep Learning and Its Parallelization," in Big Data: Principles and Paradigms, Elsevier Inc., 2016, pp. 95–118.

[17] A. F. O'Connell, J. D. Nichols, and K. U. Karanth, Camera traps in animal ecology: Methods and analyses. Springer Japan, 2011.

[18] L. Silveira, A. T. A. Jácomo, and J. A. F. Diniz-Filho, "Camera trap, line transect census and track surveys: A comparative evaluation," Biol. Conserv., vol. 114, no. 3, pp. 351–355, 2003, doi: 10.1016/S0006-3207(03)00063-6.

[19] A. E. Bowkett, F. Rovero, and A. R. Marshall, "The use of camera-trap data to model habitat use by antelope species in the Udzungwa Mountain forests, Tanzania," Afr. J. Ecol., vol. 46, no. 4, pp. 479– 487, Dec. 2008, doi: 10.1111/j.1365-2028.2007.00881.x.

[20] 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," Sci. Data, vol. 2, Jun. 2015, doi: 10.1038/sdata.2015.26.

[21] D. M. Blei, A. Y. Ng, and J. B. Edu, "Latent Dirichlet Allocation Michael I. Jordan," 2003.

[22] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 2005, vol. II, pp. 524–531, doi: 10.1109/CVPR.2005.16.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks." Accessed: Jan. 02, 2021. [Online]. Available: http://code.google.com/p/cuda-convnet/.

[24] H. Nguyen et al., "Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring," Proc. - 2017 Int. Conf. Data Sci. Adv. Anal. DSAA 2017, vol. 2018-Janua, no. October, pp. 40–49, 2017, doi: 10.1109/DSAA.2017.31.

[25] X. Yu, J. Wang, R. Kays, P. A. Jansen, T. Wang, and T. Huang, "Automated identification of animal species in camera trap images," Eurasip J. Image Video Process., vol. 2013, no. 1, p. 52, Dec. 2013, doi: 10.1186/1687-5281-2013-52.

[26] G. Chen, T. X. Han, and Z. He, "DEEP CONVOLUTIONAL NEURAL NETWORK BASED SPECIES RECOGNITION FOR WILD ANIMAL MONITORING Univeristy of Missouri Electitral and Comptuer Engineering Department Columbia , MO 65203 , USA Roland Kays , and Tavis Forrester North Carolina State University Depart," pp. 858–862, 2014.

[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2015, Accessed:

Jan. 02, 2021. [Online]. Available: http://www.robots.ox.ac.uk/.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Dec. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[29] M. Favorskaya and A. Pakhirka, "Animal species recognition in the wildlife based on muzzle and shape features using joint CNN," Procedia Comput. Sci., vol. 159, pp. 933–942, 2019, doi: 10.1016/j.procs.2019.09.260.

[30] "Image classification | TensorFlow Core." https://www.tensorflow.org/tutorials/images/classification (accessed Jan. 02, 2021).

[31] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional Multi-person Pose Estimation," Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-October, pp. 2353–2362, Nov. 2016, Accessed: Jan. 02, 2021. [Online]. Available: http://arxiv.org/abs/1612.00137.

[32] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Nov. 2017, vol. 2017-January,
pp. 1302–1310, doi: 10.1109/CVPR.2017.143.

[33] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded Pyramid Network for Multi-Person Pose Estimation," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.,

pp. 7103–7112, Nov. 2017, Accessed: Jan. 02, 2021. [Online]. Available: http://arxiv.org/abs/1711.07319.

[34] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 1653–1660, Dec. 2013, doi: 10.1109/CVPR.2014.214.

[35] A. Mathis, T. Biasi, B. Rogers, M. Bethge, and M. Weygandt Mathis, "ImageNet performance correlates with pose estimation robustness and generalization on out-of-domain data," no. Figure 2, 2020.

[36] J. Cao, H. Tang, H. S. Fang, X. Shen, C. Lu, and Y. W. Tai, "Cross-Domain adaptation for animal pose estimation," arXiv, pp. 1–14, 2019.

[37] L. Bourdev and J. Malik, "Poselets: Body Part Detectors Trained Using 3D Human Pose Annotations *." Accessed: Jan. 02, 2021.

[Online]. Available: http://pascal.inrialpes.fr/data/human.

[38] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," Int. J. Comput. Vis., vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.

[39] T. Y. Lin et al., "Microsoft COCO: Common objects in context," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), May 2014, vol. 8693 LNCS, no. PART 5, pp. 740–755, doi: 10.1007/978-3-319-10602-1_48.

[40] A. Mathis et al., "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," Nat. Neurosci., vol. 21, no. 9, pp. 1281–1289, Sep. 2018, doi: 10.1038/s41593-018-0209-y.

[41] T. D. Pereira et al., "Fast animal pose estimation using deep neural networks," Nat. Methods, vol. 16, no. 1, pp. 117–125, Jan. 2019, doi: 10.1038/s41592-018-0234-5.

[42] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions."

[43] M. Abadi et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems." Accessed: Jan. 02, 2021. [Online]. Available: www.tensorflow.org.

[44] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model." Accessed: Jan. 02, 2021. [Online]. Available: http://pose.mpi-inf.mpg.de.

[45] "(PDF) Darwin's Camera: Art and Photography in the Theory of Evolution (review)." https://www.researchgate.net/publication/236771907_Darwin's_Camera_Art_and_Photography_in_the_Theory_of_Evolution_re view (accessed Jan. 02, 2021).

[46] G. S. Cheema and S. Anand, "Automatic Detection and Recognition of Individuals in Patterned Species," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10536 LNAI, pp. 27–38, 2017, doi: 10.1007/978-3-319-71273-4_3.

[47] J. P. Crall, C. V Stewart, T. Y. Berger-Wolf, D. I. Rubenstein, and S. R. Sundaresan, "HotSpotter-Patterned Species Instance Recognition."

[48] D. T. Bolger, T. A. Morrison, B. Vance, D. Lee, and H. Farid, "A computer-assisted system for photographic mark-recapture analysis," Methods Ecol. Evol., vol. 3, no. 5, pp. 813–822, Oct. 2012, doi: 10.1111/j.2041-210X.2012.00212.x.

[49] M. Lahiri, C. Tantipathananandh, R. Warungu, D. I. Rubenstein, and T. Y. Berger-Wolf, Biometric Animal Databases from Field Photographs: Identification of Individual Zebra in the Wild. .

[50] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Accessed: Jan. 02, 2021. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/results.