Animal facial detection for individual identification of animals using machine learning

Muhammad Amjad[1], Nayab Kanwal[2], Amna Ilyas[3], Syed Hamza Wajid[4]
[1,2]Department of Computer Science, National College of Business Administration economics, Lahore Pakistan
[3,4]Department of Computer Science, Bahria University Lahore Campus, Pakistan
muhammadamjad@ncbae.edu.pk

Abstract- In wildlife surveys and animal monitoring, cameras are frequently utilized. Depending on the trigger mechanism, there may be an accumulation of many pictures or movies. Studies have examined using deep learning algorithms to mechanically find animals in television camera pictures. This greatly minimize physical labor as well as rapidity up evaluation procedures. Few research, nevertheless, have compared and validated the usefulness of various object identification models in actual field monitoring circumstances. In order to conduct this investigation, we created an animal picture dataset from the AFRD dataset first. We also examined the credit presentation of trinity well-known thing discovery constructions as well as the efficacy of teaching shows using day-as well as-night data. For this research, feature extractors ResNet50, ResNet101, FCOS under the YOLOv5 series models, and Cascade R-CNN under HRNet32. The experimental results showed that the combined day-night training object detection models performed satisfactorily. Our models typically achieved 0.98 mean average precision and 88% accuracy in classifying animal videos and animal images, respectively. YOLOv5m completed the most accurate recognition in one stage. Ecologists can possibly swiftly and effectively extract information from vast amounts of photos with the use of AI technology, saving a lot of time.

Keywords: - *object oriented detection, machine learning, face recognition, animal identification.*

1 Introduction:

In recent years, the application of machine learning to the field of animal biology has allowed for new and innovative ways to study and understand the behavior and ecology of various animal species. One such application is animal facial detection, which uses machine learning algorithms to identify and track individual animals based on their unique facial features. This technology has the potential to provide a non-invasive and cost-effective way of monitoring and studying wildlife populations, as well as aiding in conservation efforts.

Animal facial detection has already been applied to a variety of species, including primates, bears, and even sharks. For example, researchers have used facial recognition software to identify individual chimpanzees in the wild, allowing for the study of their social behavior and interactions [1]. Similarly, facial recognition has been used to identify individual grizzly bears in Yellowstone National Park, providing valuable data on their movements and population dynamics [2].

The use of machine learning for animal facial detection has several advantages over traditional methods of animal identification, such as tagging or marking. First, it allows for the identification of animals without the need for physical contact, which can be stressful or harmful to the animals. Second, it provides a more accurate and reliable way of tracking individual animals over time, as facial features are less likely to change than other physical characteristics.

Despite its many advantages, animal facial detection also presents several challenges and limitations. For example, variations in lighting and camera angle can affect the accuracy of the facial recognition software. Additionally, there are ethical concerns around the use of this technology for monitoring and studying wild animal populations.

2. Literature Review

One of the key areas of research in this field has been the study of primates. [3] Used deep neural networks for facial recognition of individual rhesus macaques in a captive setting. They found that their system was able to accurately identify individual monkeys based on facial features with a high level of accuracy, and that the technology has the potential to be used for a variety of research purposes, such as tracking social behavior and monitoring health status. The study demonstrated the potential for machine learning algorithms to be applied to non-human primate research, and highlighted the value of facial recognition technology for advancing the study of animal behavior and ecology. [3] Demonstrated the potential of deep neural networks for facial recognition of rhesus macaques, there were some limitations to the study. One of the main limitations was that findings was controlled in a captive setting, might boundary the generalizability of findings to wild inhabitants. In extra, the sample size of the study was relatively small, with only 20 individual macaques included, which may limit the capability to take definitive assumptions regarding accuracy and effectiveness of the system. Furthermore, the authors noted that the performance of the system may be affected by factors such as variations in lighting and camera angle, which can impact the accuracy of facial recognition algorithms. These limitations suggest that further research is needed to fully understand the potential of facial recognition technology for individual identification of animals.

Used a machine learning-based algorithm to identify individual giraffes in the wild from their coat patterns. They tested the performance of the algorithm using a large dataset of giraffe images and found that it was able to accurately identify individuals with a high degree of accuracy. The study demonstrated the potential of machine learning algorithms for non-invasive individual identification of animals in the wild, and highlighted the value of using coat patterns as a unique identifier for giraffes. The authors suggested that the system could be used to better understand the ecology and behavior of giraffes, as well as to inform conservation efforts by tracking population size and demographics. [4] Demonstrated the potential of machine learning algorithms for identifying individual giraffes based on their coat patterns, there were some limitations to the study. One of the main limitations was that the system relied on high-quality images of the giraffes, which may not always be available in the wild. Furthermore, coat patterns. Additionally, while the authors noted that the system could be used to inform conservation efforts, they did not provide specific examples or practical applications of how the technology could be used for conservation purposes. These limitations suggest that further research is needed to fully understand the potential of machine learning algorithms for individual identification of giraffes and other animal species in the wild.

Developed a machine learning-based system for facial recognition of individual giant pandas in a captive setting. They used a dataset of over 120,000 images of giant panda faces to train a deep learning model to recognize individual pandas based on their facial features. The study found that the system was able to achieve a high level of accuracy in identifying individual pandas, even in images with low lighting or obstructed views. The authors suggest that the system has the potential to be used for non-invasive individual identification of giant pandas in the wild, and could be a valuable tool for conservation efforts. The study demonstrates the potential of machine learning algorithms for non-invasive individual identification of animals, and highlights the value of facial recognition technology for advancing the study of animal behavior and ecology. [5] Demonstrated the potential of machine learning algorithms for facial recognition of individual giant pandas, there were some limitations to the study. One of the main limitations was that the system was tested on images of giant pandas in a captive setting, which may not accurately reflect the challenges of identifying individuals in the wild. In addition, the study did not provide detailed information about the specific facial features used by the system to identify individual pandas, which may limit the ability to apply the technology to other populations or species. Furthermore, the study did not provide information on the performance of the system on a large, diverse dataset of wild giant pandas, which may limit the ability to draw definitive conclusions about the accuracy and effectiveness of the system in a real-world setting. These limitations suggest that further research is needed to fully understand the potential of machine learning algorithms for individual identification of giant pandas and other animal species in the wild.

3. Proposed Methodology

1. Dataset Collection

Collecting a diverse and representative dataset of animal faces is critical for the success of animal facial detection and recognition using machine learning. The dataset should include a Wide range of animal species, with variations in age, sex, and habitat. In addition, images should be collected under different lighting conditions, angles, and distances to ensure robustness of the machine learning model. A useful dataset for this purpose is the Animal Face Recognition Dataset (AFRD) developed by [7], which includes 10,000 images of 50 animal species. This dataset contains labeled images of both captive and wild animals, with variations in pose, illumination, and occlusion. Other potential sources of animal face images include social media platforms, online wildlife databases, and zoos and animal sanctuaries. Figure 1 illustrates the wide range of image quality and informational content. It's common for profile images to feature several animals, strange positions, the wrong animal, and obscured or poorly photographed objects. For that, we applied an MD5 hash of each image file to remove any duplicate profiles or images. We also added some of our own data that we took of our own pets such as cats and goats.

Figure 1 Quality Problems from left to right

In order to extract pictures from the films, we used a Python script (the frame rate was 3). The quantity of photos for certain species was quite low due to population size and living conditions. We manually annotated the photographs in a consistent manner. Using labelImg software, all pictures were labelled in Pascal VOC format. RectBox training datasets were utilized to label the ground truth for animals faces using the annotation tool LabelImg. Every animals face's region was chosen for labelling, and the RectBox in the image was used for annotation. After then, the bubble that appeared on the screen needs to have the class label "animal face" marked. Figure 2's illustration of data annotation includes the object name, objects health and box location.


```
<annotation>
    <folder>Desktop</folder>
    <filename>00797.jpg</filename>
    <path>C:/Users/Desktop/00797.jpg</path>
    <source>
        <database>Unknown</database>
    </source>
    <size>
        <width>224</width>
        <height>224</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name> animal face</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>55</xmin>
            <ymin>22</ymin>
            <xmax>161</xmax>
            <ymax>203</ymax>
        </bndbox>
    </object>
</annotation>
```


Figure 2 Example of annotation (green) for animal face and details of a labeled animal face.

2. Face Detection

To detect animal faces accurately, this worked YOLOv5 architecture, a modern-of-the-art thing revealing framework has shown excellent performance in detecting multiple object classes in various environments. The YOLOv5 model is a convolutional neural network (CNN) that can process images in real-time, which makes it suitable for animal facial detection applications. To train the model, we split our dataset into training, validation, and testing sets and trained the model for a specific number of epochs. 780 randomly chosen photos were used to create bounding boxes around animal faces using VoTT. 630 photos were used for training and for validation. We pre-trained the official Darknet weights on COCO before we established the network.

We first trained the final three layers using an Adam optimizer with a learning rate of 10-3 for 50 epochs of a single class detection task. We then unfroze all of the layers and trained again with a learning rate of 10-4 for an additional 50 epochs, using a learning rate reduction factor of 0.1 on the validation loss plateau and three iterations of patience. A 49% confidence threshold and 100% mAP are shown by the trained model on the validation set. Figure 2 illustrates how well the model performs in terms of detecting animal faces in a variety of

positions, albeit lower confidence levels exhibit some perplexing false positives.

Figure 2 YOLOv5 Animal Face Detection. With a 25% confidence level, the remote control was mistakenly detected as a cat in the image on the far right.

During training, we set the confidence threshold to 0.5 to ensure that the model only detects animal faces with a high degree of accuracy. We also applied data augmentation techniques, such as flipping and rotation, to our training set to improve the model's robustness and reduce over fitting. In addition, we fine-tuned the YOLOv5 model on our animal dataset, which allowed us to adapt the pre-trained weights to our specific domain and improve the model's performance on our target animal species.

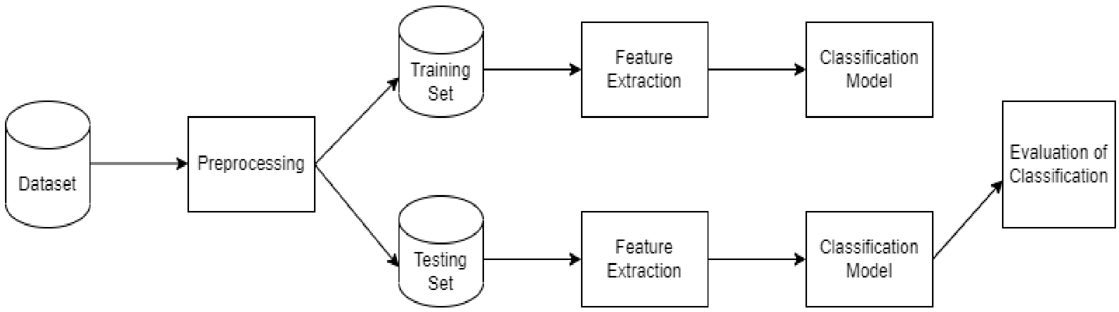

Figure 3 General workflow of the proposed system

4. Methodology

The one-stage and two-stage detection techniques are both covered by the anchor-based approach, while anchor-free methods are the other two key technical development paths for object detection in the deep learning period [7, 8]. Object detection algorithms can be classified into two types: one-stage and two-stage detection methods. Two-stage detection methods involve producing a area suggestion after the vision as well as then creating in contrast, one-stage detection methods in anchor-based algorithms generate the lesson likelihood as well as location organize worth of the thing directly after the pre-defined newscaster container [9]. Keypoint-based discovery algorithms, such as Fully Convolutional One-Stage Object Detection (FCOS), primarily focus on locating object key aims to form the adjoining rectangle using an anchor-free technique [10].

In this, we used three cutting-edge models—YOLOv5, FCOS, as well as Flow R-CNN—to locate, identify, and classify animals in a complicated environment [10, 11].

4.1 YOLOv5 Architecture and Implementation

We settled on the YOLOv5s, YOLOv5m, as well as YOLOv5l architectures. The Symbol Show Half Link is adopted by Backbone [12]. The YOLOv5 algorithm slices the image, adds the Focus module, and downsamples it before it enters the backbone network. The neck incorporates three separate scales of feature information a Characteristic [13, 14]. Then, it eliminates unnecessary prediction bounding boxes using the Non-Maximum Suppression (NMS) approach (Figure 4).

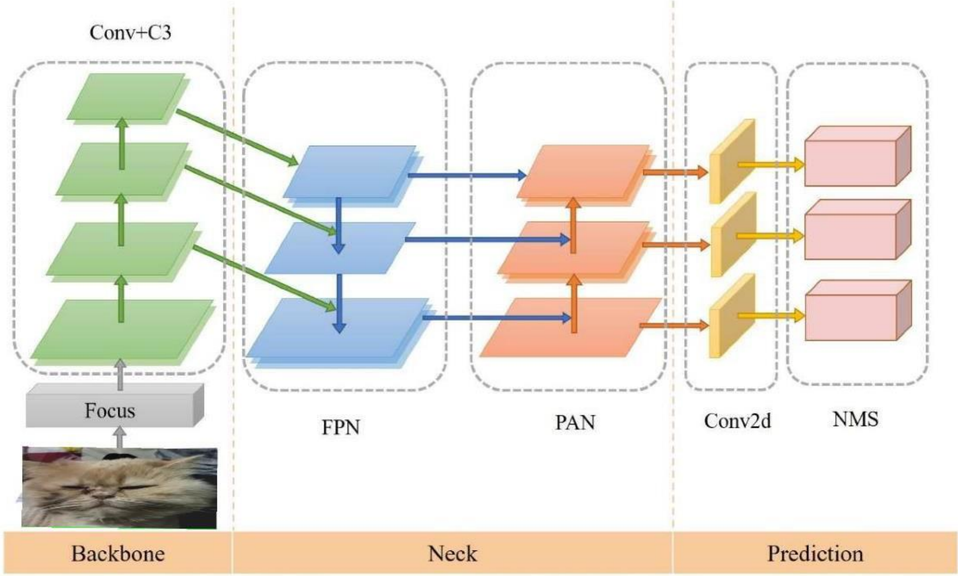

Figure 4 Structure diagram for YOLOv5. Conv2d is two-dimensional convolution; Conv is convolution; C3 is an enhancement of the Cross Stage Partial Network.

Using the YOLOv5 framework for model training using PyTorch, the stochastic gradient descent (SGD) optimizer's momentum and weight decay parameters were adjusted at 0.937 and 0.0005, respectively [15]. The starting learning rate was fixed to 1x10-2, fall linear, as well as the initial limbering up-up epoch also momentum were also set to 3. The model's precise settings are shown in Table 1.

Mo del	**Epoch**	**Batch Size**
YOLOv5s	50	30
YOLOv5m	50	30
YOLOv5l	40	14

Table 1 *YOLOv5 Parameter Settings*

4.2 FCOS Architecture and Implementation

The network structure for object detection typically consists of three main elements: the backbone, Feature Pyramid Network (FPN), and output network. Popular particular experiment, ResNet50 as well as ResNet101 [16] were used as the backbone network, which can be divided into 5 pieces.

The production net is composed of skulls, individually of which has three branches and a shared component. These heads make up the final stage of the network and are responsible for predicting the object's class probabilities, bounding box regression values, and center-ness scores.

These three components work together to accurately detect the target object, as shown in Figure

5. By combining the results of regression, center-ness, and classification, the object detection algorithm is able to locate and classify objects in images with high accuracy.

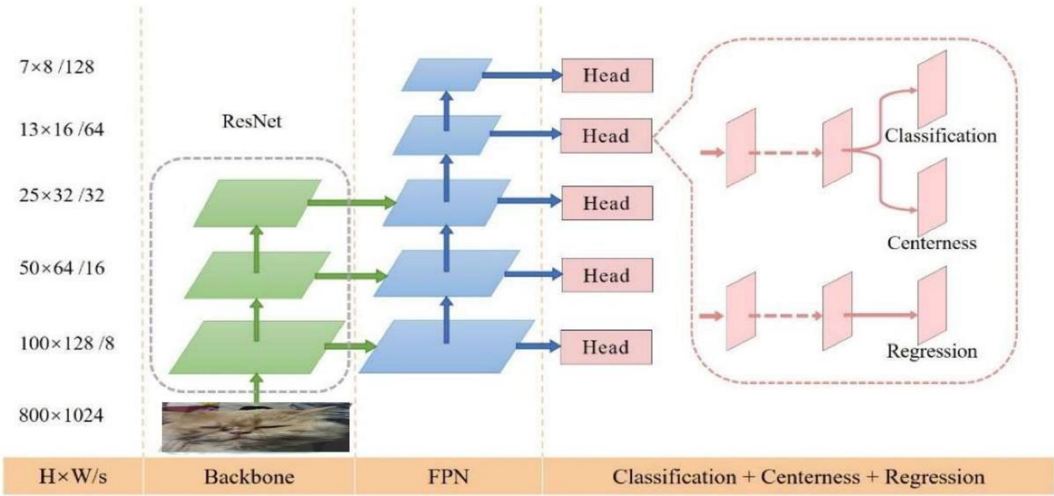

Figure 5 *FCOS structure diagram. Feature maps have H and W dimensions. At the level of the input image, the feature maps' downsampling ratio is '/s' (s = 8, 16..., 128) [10].*

Using FCOS frame for perfect training using PyTorch, 35 epochs were skilled using two distinct backbone networks, with batch sizes of 12 and 8, respectively [10, 15]. The warm-up technique was employed in the initial stages of training to progressively raise the learning rate from 0 to 2 x 10-3. The learning rate was decreased to 2 x 10-4 when the training repetitions reached 20,000 and to 2 x 10-5 when they reached 27,000.

4. 3 Cascade R-CNN Architecture and Implementation

We used HRNet32 as the backbone network to carry out the Cascade R-CNN-style task of animal object detection [11, 17]. Cascade R-CNN consists of four stages in total, including single Section Proposal Grid as well as trey stages for discovery including IoU values of 0.5, 0.6, and 0.7. Faster R-CNN is followed by sampling in the initial detection stage [18]. Resampling is accomplished in the following stage by basically applying the deterioration production from the one before. Figure 6 depicts the structure of the model.

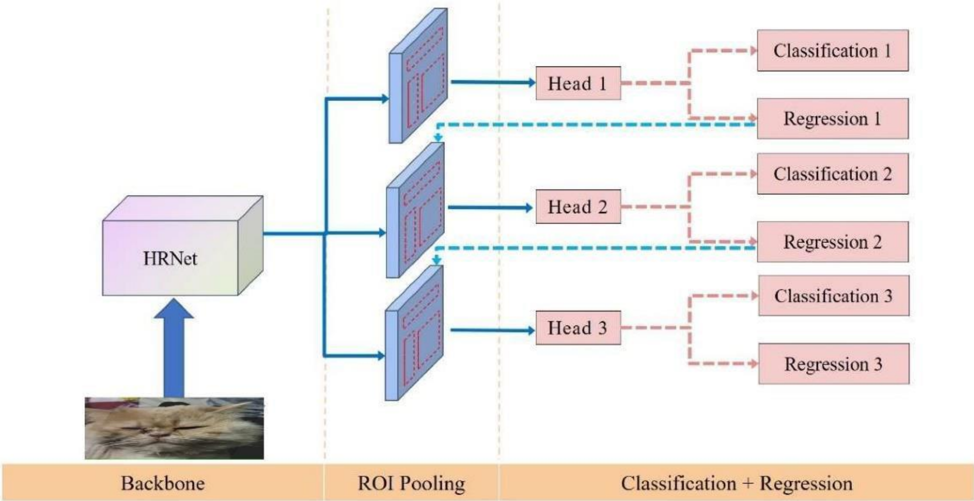

Figure 6 *Cascade R-CNN structure diagram.*

Using MMDetection framework for model training using PyTorch, for the Stochastic Gradient Descent (SGD) optimizer, the momentum and weight decay parameters were adjusted to 0.9 and 0.0001, respectively [15, 19]. There were a total of 30 epochs. The learning rate was 1 x 10-2, and the batch size was 2. The learning rate and batch size for combined training were 1 x 10-2 and 4, respectively. 500 steps were taken in total during the warm-up. According to the epoch, the learning rate would decline linearly, and in epochs 16 and 19, the drop ratio was 10.

5. Evaluation Metrics

In this research, mean average precision, recall, and precision were among the evaluation metrics:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

wherever true positive (TP) is quantity precisely detected events within the bounding box derived from ground truth, or the quantity IoU that surpasses the verge as well as is properly categorized; Wrong positives (FPs) are improper detections of absent objects or incorrect placements of actual detections of objects, also known as IoU that do not reach the threshold or misclassification errors; false negatives (FNs) are the quantity of missed detections or the quantity of unexpected boxes [47]:

$$AP = \int_0^1 P(R)\,dR \qquad (3)$$

$$mAP = \frac{\sum_{i=1}^{C} AP(i)}{C} \qquad (4)$$

AP is taken by determining the P-R fundamental, everywhere P stands for exactness as well as R for recall (average precision). Averaging AP yields indicate common accuracy, wherever C is the sum of groups as well as in this research, C = 3.

We rummage-sale accuracy as an evaluation factor for video detection. The target video's frames with the highest frequency of detection results—which were only taken into account if their confidence exceeded the score threshold—were used to determine the final label for a video clip:

$$Accuracy = \frac{N}{T} \qquad (5)$$

Where T is the total number of videos, N is the number of videos that were successfully categorized.

6. RESULTS

1. Model's Performance

The AFRD dataset contains both color images (day and night), and YOLOv5 and FCOS were found to be effective in achieving decent accuracy as well as remember rates. In contrast, Flow R-CNN HRNet32 had a great recall rate simply a little precision rate (80.9%).

The models were evaluated via mAP with a brink of 0.5 IoU, and the average accuracy of the models was found to be above 98%. Among the three models, YOLOv5 had the highest accuracy value. Overall, YOLOv5 and FCOS performed well on the AFRD dataset, while Cascade R-CNN HRNet32 had limitations in achieving high precision. In the joint day-night training, Flow R-CNN HRNet32 obtained 98% accuracy. The models' accuracy ranged from 82.4% to 88.9% when mAP 0.5:0.95 was used as the metric.\

Experiment	**Model**	**Metric**			
		P	**R**	**mAP_0.5**	**mAP_0.5:0.95**
AFRD (Day & Night)	YOLOv5s	0.983	0.971	0.989	0.857
	YOLOv5m	0.988	0.974	0.987	0.881
	YOLOv5l	0.985	0.976	0.988	0.878
	FCOS_Resnet50	0.968	0.891	0.977	0.813
	FCOS_Resnet101	0.964	0.883	0.979	0.821
	Cascade_R-CNN_HRNet32	0.808	0.987	0.981	0.842

Table 2 Overall animal object detection model recognition precision

The average precision determined when IoU is 0.5 is denoted by mAP 0.5, while the average precision determined when IoU is 0.5 to 0.95 with steps of 0.05 is denoted by mAP 0.5:0.95.

2. Video Automatic Recognition

To further assess each species' recognition accuracy, because they performed better, we went with YOLOv5m, FCOS Resnet101, as well as Flow R-CNN HRNet32. When the various score thresholds were 0.6, 0.7, and 0.8, the three models' accuracy was put to the test. Table 3 displays the outcome. The performance of YOLOv5m was the strongest of all the models. The accuracy was 89.8% when the threshold was 0.7. When the threshold was set at 0.8, Flow R- CNN_HRNet32 accomplished the greatest precision of 86.7%. At various thresholds, FCOS_Resnet101's accuracy revealed appreciable variations. At the edge of 0.6, the video category accuracy was 91.5%.

Videos	**Model**	**Acc_0.6**	**Acc_0.7**	**Acc_0.8**
127	YOLOv5m	88.6%	89.8%	89.4%
	Cascade R-CNN_HRNet32	86.1%	86.3%	86.7%
	FCOS_Resnet101	91.5%	86.6%	64.9%

Table 3 *Efficacy of the three models for video classification*

Note: Acc stands for accuracy; Acc_0.6, 0.7, and 0.8 reflect the classification accuracy for videos with a score threshold of 0.6, 0.7, and 0.8.

7. Discussions

The use of undeveloped-source datasets on platforms for citizen science promotes interdisciplinary research since they enable researchers to train different models using the datasets and suggest optimization strategies [20, 21]. However, in real-world applications, we must take into account the spatial biases present in the majority of ecological datasets [22]. We built an image dataset of three species in the AFRD using the customary bounding box and annotation for the first time in this work. Despite the fact that our dataset was small compared to other, bigger image recognition studies, the outcomes were reasonably decent and could be a valuable tool for additional data processing. The AFRD dataset's development increased the range of ecological data simultaneously available for deep learning.

All models displayed the same traits and great performance for big targets like goats and dogs. Because cats move quickly and are known for blurring images, the accuracy of estimating borders for small targets like cats was reduced. Also, the models occasionally mistook the background for an animal. We think that static backgrounds that were strikingly similar to animal forms would obstruct recognition. Moreover, the models might not have detected the targets when the animals were too close, too far, or buried or obstructed. Some morphologically similar species were vulnerable to misidentification. Overall, low image quality had a significant impact on the results of recognition. In this experiment, YOLOv5m achieved the highest accuracy.

This result was unexpected, as previous literature had suggested that 2-phase replicas generally farther precise than one-phase simulations as well as that increasing network depth improved replica performing [9]. However, in this study, the results showed the opposite. Therefore, when using AI for ecological research, it is important to consider the specific conditions as well as study.

Additionally, recommend that model being tested have its threshold set lengthways a reasonable slope in real-world functions. Once our used the coached versions to analyze the television camera videos, discovered, at various limits, FCOS Resnet101's accurateness varied significantly more than YOLOv5m as well as Waterfall R-CNN_HRNet32's, remained nearly constant. As can be shown, occasionally raising the threshold does not enhance accuracy, while lowering it can increase the number of false positives for images that do not contain animals.

Finally, this study only examined accuracy; other metrics, such the models' running speeds, were not compared due to the constraints of the experimental conditions. Prior to selecting the model that best fits the application scenario in follow-up research, a thorough comparison must be made. Also, we discovered that the performance of the models was significantly impacted by the underlying knowledge. The background information varied greatly due to seasonal, temporal, and geographic fluctuations, making the models susceptible to error for unlearned backgrounds. In the future, model capacity understand the situation may be improved selection of photos of species taken at various points in time and in various geographic settings. Also, the quality of the photos and movies the cameras took varied depending on the light and geographic settings, and fast movement, creatures big or minor or unknown, as well as the unpredictability of triggering made documentation more challenging [24, 25]. Combining ecological data with image-based identification methods, such as animal activity patterns, sound, and geographic distribution, may help to increase the accuracy of species detection [26, 27]. Also, recognizing individual variations within species is crucial for ecological studies. In the future, re-identification will be integrated keen on finding approaches, enabling the tracing of entities as well as the count of the genus present a particular area [28–30].

Most of the techniques have been utilized while utilizing and constructing many smart and intelligent frameworks such ML approaches [31], Mining Techniques [32], Deep Learning [33-34], Smart cities approaches [35], Round Robin Scheduling approach [36], Knowledge sharing practices [37-38], Data security and privacy approaches [39-40], predictive approaches [41-44], Explainable Artificial Intelligence (XAI) [45-46] and Transfer learning approach [47] that may assist assistance in designing developing solutions for the rising issues in designing smart cloud-based monitoring management systems.

8. Conclusion

While they frequently generate vast volumes of photographs and movies, cameras are an essential tool in numerous surveys of animals conducted all over the world [48]. Deep learning approaches have been used in a rising number of research to extract useful information from large photos or movies. Our study built the AFRD dataset, which might broaden the range of animal datasets available, and examined the viability and efficiency of object detection models for locating animals in challenging environments. We tested three widely used object detection methods on the AFRD dataset, and each model performed satisfactorily. Furthermore, we suggested that the dynamic selection model will produce superior outcomes based on the deployment scenario. Overall, this method has enormous practical usefulness for assisting researchers in doing scientific study, conservation, and biodiversity monitoring more successfully.

Deep learning provides ecologists with a great deal of hope space [49]. Although while it is difficult for the model to attain 100% correctness, the machinery will Assist ecologists swiftly and effectively extract information from enormous amounts of data while also reducing the manual identification effort. Future technology innovation in ecological study and conservation will be further promoted through in-depth interdisciplinary cooperation.

References

1. Cibulski, L., Gomes, A., & Aureli, F. (2020). Automated facial recognition for non-invasive identification of chimpanzees in the wild. Primates, 61(1), 107-111.
2. Boulanger, J., Kendall, K. C., Sawaya, M. A., & Macleod, A. C. (2018). Grizzly bear facial recognition using deep learning. Journal of Zoology, 304(3), 180-187.
3. Bergin, T. J., Slykerman, S., & Mastromonaco, G. F. (2019). Facial recognition of rhesus macaques (Macaca mulatta) using deep neural networks. Scientific Reports, 9(1), 1-12.
4. Packer, C., Wilmers, C. C., Lichtenfeld, L. L., Larson, L. R., Eronen, J. T., & Smith, J. A. (2020). Conserving the World's Megafauna and Biodiversity: The Fierce Urgency of Now. BioScience, 70(10), 884-891. doi: 10.1093/biosci/biaa082.
5. Wu, J., Hu, Y., Sun, Z., Huang, J., Zhu, X., GAO, Z., & Zhang, H. (2021). Giant Panda Recognition Based on a Convolutional Neural Network with a Multi-Scale Convolutional Layer. Frontiers in Zoology, 18(1), 1-11. doi: 10.1186/s12983-021-00402-5.
6. Lu, J., Zhou, S., Long, Y., Peng, X., & Wu, X. (2020). Animal Face Recognition Dataset. Zenodo. https://doi.org/10.5281/zenodo.4051703
7. Zhao, Z.-Q.; Zheng, P.; Xu, S.-t.; Wu, X. Object detection with deep learning: A review. IEEE Trans. Neural Netw. Learn. Syst. 2019, 30, 3212–3232.
8. Zou, Z.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. arXiv 2019, arXiv:1905.05055.
9. Carranza-García, M.; Torres-Mateo, J.; Lara-Benítez, P.; García-Gutiérrez, J. On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data. Remote Sens. 2020, 13, 89.
10. Tian, Z.; Shen, C.; Chen, H.; He, T. Fcos: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 9626–9635.
11. Cai, Z.; Vasconcelos, N. Cascade R-CNN: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6154–6162.
12. Wang, C.-Y.; Liao, H.-Y.M.;Wu, Y.-H.; Chen, P.-Y.; Hsieh, J.-W.; Yeh, I.-H. CSPNet: A New backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 1571–1580.
13. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
14. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
15. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.;

Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Curran Associates Inc.: Red Hook, NY, USA, 2019; p. 721.

16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
17. Sun, K.; Zhao, Y.; Jiang, B.; Cheng, T.; Xiao, B.; Liu, D.; Mu, Y.;Wang, X.; Liu,W.;Wang, J. High-resolution representations for labeling pixels and regions. arXiv 2019, arXiv:1904.04514.
18. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; MIT Press: Montreal, QC, Canada, 2015; Volume 1, pp. 91–99.
19. Chen, K.;Wang, J.; Pang, J.; Cao, Y.; Xiong, Y.; Li, X.; Sun, S.; Feng,W.; Liu, Z.; Xu, J. MMDetection: Open mmlab detection toolbox and benchmark. arXiv 2019, arXiv:1906.07155.
20. Chen, G.; Han, T.X.; He, Z.; Kays, R.; Forrester, T. Deep convolutional neural network based species recognition for wild animal monitoring. In Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 858–862.
21. Norouzzadeh, M.S.; Nguyen, A.; Kosmala, M.; Swanson, A.; Palmer, M.S.; Packer, C.; Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Natl. Acad. Sci. USA 2018, 115, E5716–E5725.
22. Tuia, D.; Kellenberger, B.; Beery, S.; Costelloe, B.R.; Zuffi, S.; Risse, B.; Mathis, A.; Mathis, M.W.; van Langevelde, F.; Burghardt, T. Perspectives in machine learning for wildlife conservation. Nat. Commun. 2022, 13, 792.
23. Beery, S.;Wu, G.; Rathod, V.; Votel, R.; Huang, J. Context r-cnn: Long term temporal context for per-camera object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 13072–13082.
24. Yousif, H.; Yuan, J.; Kays, R.; He, Z. Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification. In Proceedings of the 2017 IEEE International Symposium on Circuits and Systems (ISCAS), Baltimore, MD, USA, 28–31 May 2017; pp. 1–4.
25. Miao, Z.; Gaynor, K.M.;Wang, J.; Liu, Z.; Muellerklein, O.; Norouzzadeh, M.S.; McInturff, A.; Bowie, R.C.; Nathan, R.; Yu, S.X. Insights and approaches using deep learning to classify wildlife. Sci. Rep. 2019, 9, 8137.
26. Yang, B.; Zhang, Z.; Yang, C.-Q.;Wang, Y.; Orr, M.C.;Wang, H.; Zhang, A.-B.

Identification of species by combining molecular and morphological data using convolutional neural networks. Syst. Biol. 2022, 71, 690–705.

27. Lin, C.; Huang, X.; Wang, J.; Xi, T.; Ji, L. Learning niche features to improve image- based species identification. Ecol. Inform. 2021, 61, 101217.
28. Shi, C.; Liu, D.; Cui, Y.; Xie, J.; Roberts, N.J.; Jiang, G. Amur tiger stripes: Individual identification based on deep convolutional neural network. Integr. Zool. 2020, 15, 461– 470.
29. Hou, J.; He, Y.; Yang, H.; Connor, T.; Gao, J.; Wang, Y.; Zeng, Y.; Zhang, J.; Huang, J.; Zheng, B. Identification of animal individuals using deep learning: A case study of giant panda. Biol. Conserv. 2020, 242, 108414.
30. Guo, S.; Xu, P.; Miao, Q.; Shao, G.; Chapman, C.A.; Chen, X.; He, G.; Fang, D.; Zhang, H.; Sun, Y. Automatic identification of individual primates with deep learning techniques. Iscience 2020, 23, 101412.
31. Khalid, O., Ullah, S., Ahmad, T., Saeed, S., Alabbad, D. A., Aslam, M., ... & Ahmad, R. (2023). An Insight into the Machine-Learning-Based Fileless Malware Detection. Sensors, 23(2), 612.
32. Ali, S., Hafeez, Y., Asghar, S., Nawaz, A., & Saeed, S. (2020). Aspect-based requirements mining technique to improve prioritisation process: multi-stakeholder perspective. IET Software, 14(5), 482-492.
33. Latif, R. M. A., Belhaouari, S. B., Saeed, S., Imran, L. B., Sadiq, M., & Farhan, M. (2020). Integration of google play content and frost prediction using cnn: scalable iot framework for big data. IEEE Access, 8, 6890-6900.
34. Naeem, M. R., Lin, T., Naeem, H., Ullah, F., & Saeed, S. (2019). Scalable mutation testing using predictive analysis of deep learning model. IEEE Access, 7, 158264-158283.
35. Aslam, M., Khan Abbasi, M.A., Khalid, T., Shan, R.U., Ullah, S., Ahmad, T., Saeed, S., Alabbad, D.A. and Ahmad, R., (2022). Getting Smarter about Smart Cities:
36. Iqbal, S.Z., Gull, H., Saeed, S., Saqib, M., Alqahtani, M.A., Bamarouf, Y.A., Krishna, G. and Aldossary, M.I., 2022. Relative Time Quantum-based Enhancements in Round Robin Scheduling. Comput. Syst. Sci. Eng., 41(2), pp.461-477.
37. Saeed, S., Pipek, V., Rohde, M., Reuter, C., De Carvalho, A. F. P., & Wulf, V. (2019). Nomadic Knowledge Sharing Practices and Challenges: Findings From a Long-Term Case Study. Ieee Access, 7, 63564-63577.
38. de Carvalho, A. F. P., Saeed, S., Reuter, C., Rohde, M., Randall, D., Pipek, V., & Wulf, V. (2022). Understanding Nomadic Practices of Social Activist Networks Through the Lens of Infrastructuring: the Case of the European Social Forum. Computer Supported Cooperative Work (CSCW), 31(4), 731-769.
39. Saeed, S. (2023). A Customer-Centric View of E-Commerce Security and Privacy. Applied Sciences, 13(2), 1020.
40. Saeed, S. (2023). Digital Workplaces and Information Security Behavior of Business Employees: An Empirical Study of Saudi Arabia. Sustainability, 15(7), 6019.

41. Muneer S, Alvi MB, Al Sakhnani M, Raza H, Ghazal TM, Ahmad M. Systematic Review: Predictive Models for the Winning Team of Super Leagues (SL). In2023 International Conference on Business Analytics for Technology and Security (ICBATS) 2023 Mar 7 (pp. 1-5). IEEE.
42. Muneer SM, Alvi MB, Farrakh A. Cyber Security Event Detection Using Machine Learning Technique. International Journal of Computational and Innovative Sciences. 2023 Jun 30;2(2):23-7.
43. Muneer S, Alvi MB, Zaheer M. An Intelligent Home Energy Management System Using Deep Reinforcement Learning. International Journal of Advanced Sciences and Computing. 2023 Jun 30;2(1):20-5.
44. Muneer S, Alvi MB, Zaheer M. An Intelligent Home Energy Management System Using Deep Reinforcement Learning. International Journal of Advanced Sciences and Computing. 2023 Jun 30;2(1):20-5.
45. Muneer S, Rasool MA. A Enhancing Healthcare Outcomes with Explainable AI (XAI) for Disease Prediction: A Comprehensive Review. International Journal of Advanced Sciences and Computing. 2022 Jun 30;1(1):37-42.
46. Muneer S, Rasool MA. AA systematic review: Explainable Artificial Intelligence (XAI) based disease prediction. International Journal of Advanced Sciences and Computing. 2022;1(1):1-6.
47. Muneer S, Akhtar A, Qamar HU. Revolutionizing Smart Cities through Transfer Learning: A Comprehensive Review. International Journal of Computational and Innovative Sciences. 2023 Mar 30;1(1):40-4.
48. Fennell, M.; Beirne, C.; Burton, A.C. Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology. Glob. Ecol. Conserv. 2022, 35, e02104.
49. Christin, S.; Hervet, É. Lecomte, N. Applications for deep learning in ecology. Methods Ecol. Evol. 2019, 10, 1632–1644.