IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 2, June 2024, pp. 2354~2363
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i2.pp2354-2363

Chelonia mydas detection and image extraction from field recordings

Khalif Amir Zakry[1], Mohamad Syahiran Soria[1], Irwandi Hipiny[1], Hamimah Ujir[1], Ruhana Hassan[2,3], Richki Hardi[4]
[1]Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Sarawak, Malaysia
[2]Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Sarawak Malaysia
[3]Centre for Pre-University Studies, Universiti Malaysia Sarawak, Sarawak, Malaysia
[4]Faculty of Computer Science, Universitas Mulia (UM), Kalimantan Timur, Indonesia

Article Info

Article history:

Received Jun 22, 2023
Revised Oct 31, 2023
Accepted Dec 3, 2023

Keywords:

Artificial intelligence
Artificial intelligence in ecology
Data filtration
Image recognition
Machine learning
Task automation
Wildlife videography

ABSTRACT

Wildlife videography is an essential data collection method for conducting. The video recording process of an animal like the Chelonia mydas sea turtles in its habitat requires setting up special camera or by performing complex camera movement whilst the camera operator maneuvers over its complicated habitat. The result is hours of footage that contains only some good data that can be used for further animal research but still requires human input in filtering it out This presents a problem that artificial intelligence models can assist, especially to automate extracting any good data. This paper proposes usage of machine learning models to crop images of endangered Chelonia mydas turtles to help prune through hundreds and thousands of frames from several video footages. By human supervision, we extracted and curated a dataset of 1,426 good data from our video dataset and used it to perform transfer learning on a you only look once (YOLO)v7 pre-trained model. Our paper shows that the retrained YOLOv7 model when run through our remaining video dataset with various confidence scores can crop images in the field video recordings of Chelonia mydas turtles with up to 99.89% of output correctly cropped thus automating the data extraction process.

Corresponding Author:

Khalif Amir Zakry
Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak
Datuk Mohammad Musa Road, Samarahan City, Sarawak, 94300, Malaysia
Email: 21020463@siswa.unimas.my

1. INTRODUCTION

Chelonia mydas are one of the animals that have been significantly impacted by global warming, a side effect from humanities production of unnatural greenhouse gases such as chlorofluorocarbons, chlorine, and bromine which has caused our climate to change with an increase in global temperature [1], and rising sea levels [2]. Unlike some species of animals that are able to migrate to newer habitats with more tolerable temperatures to survive [3]; the endangered Chelonia mydas sea turtles depend exclusively on their specific nesting beach habitats that cannot be moved and is being threatened with loss due to rising sea levels [4]. Therefore, it is important for wildlife researchers to monitor these animals to obtain relevant data that can be used to assist in the conservation efforts of the species. The monitoring process for wildlife surveying may involve gathering data such as live sampling by methods such as bait and trapping or via non-trapping wildlife videography techniques such as installing automatic camera traps [5].

Wildlife videography is a complicated and arduous task. As such, analyzing and extracting usable data from amateur recordings of wildlife videography is a labor-intensive process that requires human experts to identify the animal and determine whether the image from the video is usable for research. This hinders the extraction of important information from the video footages to be used for wildlife research and conservation efforts for the protection of the subject species. There are several examples of hindrances when extracting data from wildlife videography. The first is the presence of various other species of animals in the video footage when using sensor-based motion trap cameras [6]. This videographic method of recording wildlife footage utilizes trap cameras that begins and stops recording when the motion sensor is triggered by a crossing animal or object; has been reliably and effectively used by researchers for animal monitoring and conservation studies [7]–[10]. It has been shown that utilizing trap cameras are very effective for large number of detections of an individual species of animals as well as for documenting species richness in an ecosystem [11]. However, one of the main issues with this form of wildlife videography is that after several weeks of collecting video data from a field camera, hours of footage will need to be analyzed through to find and extract data images of the animal that is being researched [12]. The process becomes more time consuming as more and more images are extracted from the videos or photos [13]. Other than that, some images of wildlife captured using motion-based sensors contains background elements that are insignificant to the research and are beyond the region of interest (ROI) needed [14]. The reliability of the trap cameras to operationally collect only the subject animal may also be reduced because of the trap cameras local environment such as the presence of snowfall or wind which can trigger the trap cameras automatic sensors, thus generating false positives [15].

Besides that, another videographic method of wildlife videography is one that is recorded by a cameraman outdoors in the field where the animals are free roaming. This method of wildlife videography requires that a cameraman operate a camera to capture a video or photograph of the roaming animal in its natural habitat of which its natural habitat is an area where the animal lives and grows unforced by humans [16]. This type of wildlife videography can generate noisy frames where a moving animal or camera movement causes the frames to be blurry or the image frame of the subject animal unclear for periods of a time. This is a problem for data extraction processes later as human input is required to prune through the footage to find clear images of the subject animal that can be used among hundreds of blurry frames in the footage.

This paper introduces using neural networks for processing and detecting animal subjects in video field recordings for potential use in data extraction of images from wildlife videography data to create an image dataset. Such neural network can alleviate issues of both videographic methods as mentioned previously. We propose using you only look once v7 model, also known as YOLOv7. Automated methods of detecting animals in camera trap images have been conducted in several works such as [17]–[20]. This paper improves upon the previous works by adapting a different neural networking model or in the case of Yu *et al.* [17], a more a newer and improved version of the YOLO, model than they have used. We hypothesise that by re-training the newer YOLOv7 model that V7 Labs has initially trained with the microsoft common objects in context (MSCOCO) dataset, by doing supervised transfer learning with 1,426 human curated good image data from a small number of our field video recordings; we can produce an automated croping tool with that model to automate the cropping process of good image data from the remainder of the field video recordings of which the accuracy of the good cropped data being outputed may be increased by tweaking the confidence score threshold.

2. METHOD

In this paper, we propose the use of the deep learning model, YOLOv7 to help in the process of identifying and extracting images from a noisy wildlife videography dataset. YOLO is a neural network model developed by Redmon *et al.* [21] that has the capabilities to predict bounding boxes and class probabilities of objects in an image. YOLO itself is not unique in its use for animal detection as other competing neural network models such as faster region-based convolutional network (R-CNN) models have also been used [22]. However, YOLO is opted as YOLO has been demonstrated to perform much faster and with reliable accuracy compared to faster R-CNN in various studies such as works by [23], [24]. For our method, we will be using our own dataset of field video recordings from a data collection session of Chelonia mydas turtles that we have gathered.

2.1. Research dataset

We have collected exactly 49 videos on 35 different female individuals of Chelonia mydas turtles with a total video duration 1 hour 38 minutes and 34 seconds for a wildlife research effort on the Talang Satang Islands off the coast of Sarawak. These videos are to be used for our other wildlife research study on Chelonia mydas turtles but can also be used to conduct our research for this paper. The dataset of Chelonia mydas videos are obtained by our camera operators who accompanied the Sarawak Forestry Corporations' park rangers during animal tagging operations on the nesting female Chelonia mydas individuals with their permission, assistance and guidance while following the park's strict ethic rules. Due to the nature of our recording environment which takes place at the windy and sandy beaches of Talang Satang Island, we utilized an off-the-shelf CCTV camera with infrared capabilities powered by an external high capacity (50,000 milliamp hour) power bank which are both mounted on a man-portable extendable 5-7 foot pole that can be hoisted up to 3 meters above the nesting Chelonia mydas individuals; this allows us to record field recordings of the Chelonia mydas turtle during the rangers' tagging operations at night which is a routine of the park while also not interfering with their task by being too close to the rangers or the turtle. However, the videographic data obtained from this data collection process proves to be very noisy and has multiple instances of capturing other object and subjects in the video frame besides the Chelonia mydas turtles such as rocks, people, and the camera set up. Some frames also feature extreme blurring of the Chelonia mydas turtle. Examples of such objects and noise are shown in Figure 1.

Because of the camera and process used to collect the videographic data of the Chelonia mydas turtles, the videos and image frames obtained are grayscale due to the limitations of the CMOS sensor in our camera equipment and can be significantly noisy since no image stabilization is used. The images shown above are examples of 'bad' images that cannot be used in our wildlife research study; specifically, we cannot directly train a model with the presence of these objects in frame as they can impact the learning process of the machine learning model for this paper as well as in our ongoing research later. As such we consider these as noise that ideally a human must analyze through and ensure is not within the cropped-out section of the images which contains a clear and good representation of the Chelonia mydas individuals for creating a viable dataset in our ongoing research. Object A for example is a rubber pipe that is almost always present in our field video recordings as it is used by the park rangers to mark the exact nest location of a specifc Chelonia mydas individual. Object B on the other hand is an overblured image of a Chelonia mydas invididual that we do not want to be trained on our machine learning models as the object posesses very little image characteristics that represent a Chelonia mydas turtle. Object C and D are similar in nature to object A as they are present significantly in our field video recordings due to the presence of human camera operators and park rangers for the context of Object C; while for Object D which is a stick washed ashore, is a common feature of the beach habitat itself.

Ideally a 'good' cropped image that we would like to extract from our noisy field video recordings are like as shown in Figures 2 and 3 that can be used to create a viable dataset for training machine learning models. A viable dataset for training should contain images that are accurately labeled and closely represent the characteristics of the dataset's classes, and this introduces us to our following methods.

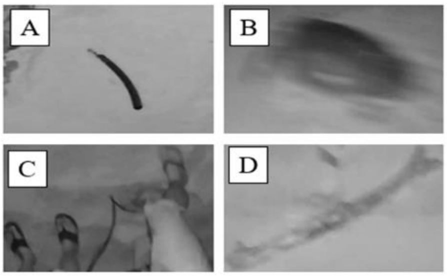

Figure 1. A, B, C, and D are examples of undesired objects or effects that are present in our field video recordings during the tagging operations with the park rangers

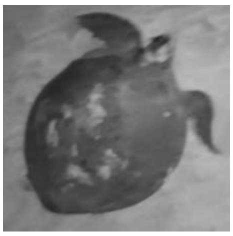

Figure 2. A Chelonia mydas turtle resting

Figure 3. A Chelonia mydas turtle moving

2.2. Training our own custom model with YOLOv7

From our gathered data, we will have to first utilize a human reviewer to extract and identify 'good' images from the noisy field recordings and crop them manually from the numerous number of frames present in our videos. For this part of the method, only 5 videos were used with a total duration of 15 mintes and 26 seconds showcasing 3 different Chelonia mydas female individuals. FFmpeg library in Python was used to extract the image frames from the 5 videos which are then human filtered to bring the total number of frames which contains a clear depiction of the 3 individuals down to only 1426 image frames. The next process is also labor intensive as human reviewers were then required to identify the Chelonia mydas individuals within the 1426 frames remaining and draw a bounding box around only the 3 Chelonia mydas individuals. This must be done thoroughly as 'bad' mislabels can negatively impact the accuracy of our trained custom YOLOv7 model. The impact of low-quality images such as blurred images can significantly reduce the accuracy of deep learning models [25], [26]. The dataset of 1426 images of Chelonia mydas turtles that were extracted, and labeled using Roboflow Annotate platform are then defined into a single class (Chelonia mydas) and split into three separate sets for training detailed in Table 1. This process approximately took the human reviewer 6-7 hours of work across 2 days. Then, utilizing YOLOv7 running on the Jupyter notebook environment in Python powered with a consumer grade RTX 3060 8 GB graphics card and a 6-core processing unit; several different training epochs are tested to determine which epoch can produce a model with the best result suitable for our use case. The epochs used to train each model and its results are shown in Table 2.

Based on the results of training four different sets of models, model C is selected to be adopted for further testing. It is selected as it has reached our desired precision and recall value of 1. Although model C and D both achieved the best precision and recall value, C is selected due to its high mean average precision for intersection over union, also known as IOU, with thresholds of 0.5 and from 0.5 to 0.95 specifically, over model D. The difference in classification capability of Chelonia mydas turtles between model C and the underperforming models such as A or B can be visualized as in Figures 4 to 6.

Table 1. Image distribution used for training our model

Dataset	Number of images
Training set	998
Validation set	285
Testing set	143

Table 2. The different epoch and metrics were used to evaluate three different models

Model	Epoch	Metrics			
		Precision	Recall	mAP@0.5	mAP@0.5:0.95
A	10	0.6901	0.3047	0.3921	0.1439
B	20	0.9927	0.9507	0.9895	0.7129
C	30	1	1	0.995	0.795
D	40	1	1	0.995	0.7806

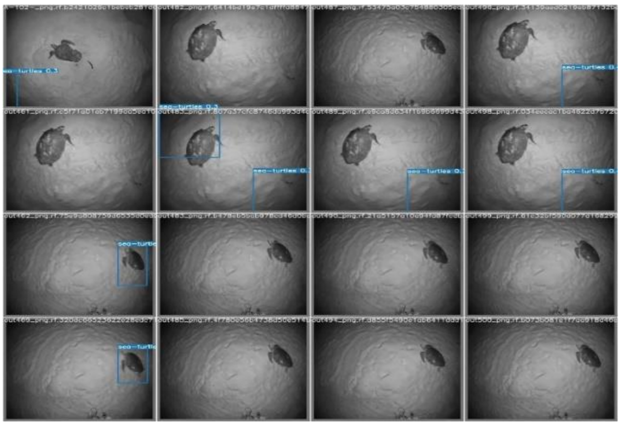

Figure 4. Classification capability of model A while attempting to draw bounding boxes over images in the test set

Note the inaccuracy of prediction of model as shown in Figure 4 where the blue boxes in each image section are the attempts by model A to draw a bounding box around what it thinks is a turtle. We can see that in was only able to accurately draw the bounding box over 2 of the 16 images in this test. Model B as shown in Figure 5 was able to draw a bounding box over every turtle in the above test images. However, it also drew bounding boxes over the feet of the human camera operators as can be seen with the images in the two bottom rows where the human camera operators' feet can be seen and classified in a bounding box as a sea turtle.

Model C on the other hand outperforms both models A and B by recognizing the Chelonia mydas turtle individual in each test frame (the ROI) as well as ignoring other noise elements that are present in each frame such as the human camera operators' feet or the extruding pipes in the test frames by accurately drawing the bounding boxes over the Chelonia mydas turtles only. It is important to restate again that the data we retrieved from our Chelonia mydas turtle field recordings are recorded in grayscale without colorization due to the specifications of the camera used. This can affect the classification capability of the model should this model be used to classify images with color as color can impact the predictive capability of a model [27]. We will be testing the capabilities of model C on various confidence scores thresholds with the remainder of our noisy field video recordings to assess the classification and cropping capability with regards to Chelonia mydas turtles in remaining field video recordings.

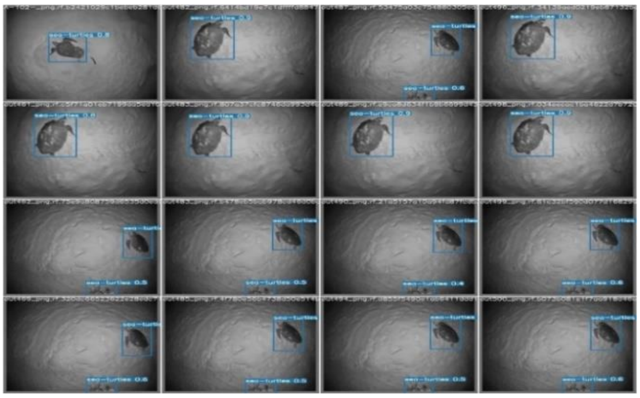

Figure 5. Classification capability of model B while attempting to draw bounding boxes over images in the test set

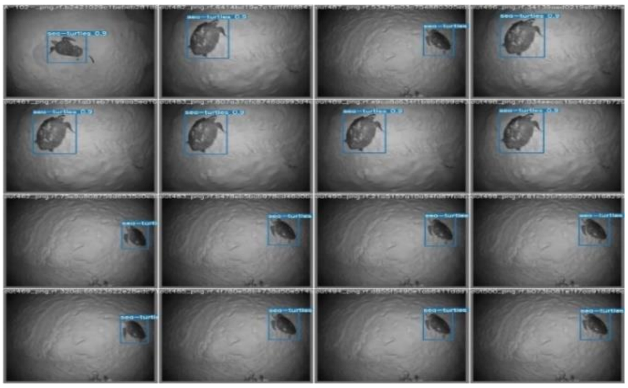

Figure 6. Classification capability of model C while attempting to draw bounding boxes over images in the test set

2.3. Utilizing model C with python in jupyter notebook to perform cropping automation based on classification above various confidence score thresholds

Confidence score is the accuracy level of the model for a classification, and it can be expressed as a percentage of up to a 100% or as a decimal value between 0 and 1. When a high confidence score threshold is set for example 0.95, it means that we only allow the model to classify the object when it is 95% sure that it is able to accurately determine that object is what it thinks it is. Confidence score thresholds have been used by other researchers as a method for filtering false positives while ensuring the predicted bounding box has achieved a desired minimum score [28].

We have the remainder 44 field video recordings with a duration of 1 hour and 23 minutes and 8 seconds that represents a total of 99,479 frames of data to be analyzed. The remaining 44 field recordings are of 32 different Chelonia mydas female individuals. We applied different values of confidence scores in the classification process by creating hardcoded thresholds in our Python program. The Python program will only crop out the bounding box section of the classified object in the frame if it is able to confidently reach or go above the threshold. Doing this on the remaining 99,479 frames of the remaining 44 field video recordings with model C produces varying results corresponding to the confidence score threshold used. The classification and cropping process on the 44 field recordings will yield cropped images with the dimensions of 640x640 resolution as an output. Our example outputs can be seen in Figures 7 and 8. We ran model C with adjustments of 0.05 increments in confidence scores starting from 0.70 to 0.95 with each run going through all 99,479 frames. The output of the 44 field recordings undergoing this process for each confidence score thresold are manually reviwed by a human to verify the true postivie and false positive outputs. This process took about 3 working days for a single human reviewer to verify the output; the exact amount of labour hours taken are however not recorded. The results can be found in the results section.

Figure 7. Example of a 'good' image output. A true positive classification

Figure 8. Example of a 'bad' image output. A false positive classification

3. RESULTS AND DISCUSSION

3.1. Discussion of testing our model C with various confidence score thresholds

Based on the results of the test we can examine that the number of output images decrease proportionally with the increase in confidence score of each test run. Test VI, due to its high confidence score assigned yielded no output images. This is because there is no frame within the 99,479 frames of data capable of reaching the 0.95 confidence score for model C to classify the object as a Chelonia mydas turtle. This means that none of our gathered data can be confidently identified by our model C to be above 95% to be taken classification evaluation in test I, our lowest confidence score assigned to model C for classifying Chelonia mydas turtles, yields the most output images that has been classified as Chelonia mydas turtles. However, among those output images, a significant number of them are misclassified images of noise that have been classified as Chelonia mydas individuals thus has the lowest percentage correctly classified output images among Test I, II, III, IV, and V. A trend can be clearly seen from our 6 different tests with the various confidence score thresholds assigned for each test. The trend follows a few key points:

– The lower the confidence score threshold assigned to the model, the more the number of output images.
– The lower the confidence score threshold assigned to the model, the more the percentage of false positive detections.
– Above the confidence threshold of 95%, our model is not able to classify and label any of our data as no frame of object within any ROI is able to meet the requirement.

As previously stated, by Wenkel *et. al.* [28], there are certain scenarios where a high confidence level score is required as any incorrect detection may cause severe problems. For our use case specifically, we are looking to utilize model C as a method of pruning through our noisy field video recordings to generate a dataset of cropped images that will be used as a dataset to train future models in our next upcoming study. As such, it is vital that the dataset used be 100% correctly labeled as any mislabeled data in that dataset can effectively impact the training of our models to come. Therefore, our scenario dictates that our best course of action is to have the least amount of incorrectly classified cropped output images as possible. We determined that model C with the confidence threshold of 0.90 as in Test V is the best combination for pruning through our noisy field video recordings. Table 3 shows the results of our model C that is ran with the remaining dataset at the various confidence score thresholds.

However, selecting a high confidence score threshold is not without any tradeoffs. As can be seen when comparing with the other tests, even though we ideally selected model C with confidence threshold of 0.90 as our combination of choice, it also provides us with the least amount of output images for our dataset.

This is trivial for us as we prioritize the correctness of our data but may be a concern for certain applications. As such, our recommendation for utilizing our method of pruning noisy field video recordings would be to use at minimum a confidence level of 0.80 and above. The reason behind this is that it will allow the amount of output images to increase while maintaining a tolerable percentage of errors to slip through. It has been validated by Northcutt *et al.* [29] that the 10 most used public datasets contain at least 3.3% mislabeling errors in the dataset while remaining prevalently used such as ImageNet and CIFAR-100. Due to that, our advice is to prioritize selecting the confidence score threshold that would provide the lowest amount of incorrectly classified output while providing the required amount of data as needed.

Table 3. Amount of images output by the automated classification and cropping process and some comparison metrics

Test	Confidence score	Number of output images	No of correctly classified images (TP)	No of misclassified images (FP)	Percentage of images correctly classified (%)	Percentage of images incorrectly classified (%)
I	0.70	67,030	62,494	4,536	93.23	6.77
II	0.75	61,526	58,432	3,094	94.97	5.03
III	0.80	53,271	51,715	1,556	97.08	2.92
IV	0.85	38,609	38,071	538	98.61	1.39
V	0.90	7,618	7,610	8	99.89	0.11
VI	0.95	0	0	0	-	-

3.2. Comparing our results with other similar works

Similar works have been done previously by other authors focusing on utilizing machine learning technology for various animal classification tasks. As such we can attempt to compare their work with our results to see if we are able to improve upon their work and successes. Research by Gray *et al.* [30] for example, whereby they utilized convolutional neural network (CNN) models to help in assessing at-sea densities of Lepidochelys olivacea sea turtles form drone footages. They managed to achieve a detection accuracy of 99.83%. However, they do note that classification accuracy is not the priority metric for their use case. Chen *et al.* [31] has also performed a wildlife surveillance study using deep learning tools, specifically by utilizing CNN models to perform binary classification for Meles meles badgers and with good results. Based on their results, they achieved up to 95.12% of test images correctly classified with only 4.88% of images incorrectly classified. When we compared this result with ours, we managed to achieve an even higher percentage of correctly classified images of our subject animal, the Chelonia mydas turtle with 99.89% images correctly classified at our highest confidence threshold setting. As such we believe that applying different confidence level threshold as a labeling requirement for correct and incorrect predictions can increase the number of true positives while decreasing the number of false negatives.

Besides that, we can also investigate the works done by Kutugata *et al.* [32] where they utilized InceptionV3, which is also a CNN to classify various animal images captured from wildlife trap cameras. We can infer the images taken as a single image frame taken of an animal crossing the camera. Based on their results, we can compare the effectiveness of their classification process on several different animals in their dataset of trap images taken from Rio Grande Valley, in Texas. They managed to achieve above 80% of images being correctly classified on their several different animals such as 91.98% for armadillos, 91.35% for birds, 89.2% for rabbits and up to 100% for tortoises. The results for animals other than tortoises are lower than ours but we believe the amount of correctly classified images can be increased should the confidence threshold for classification also be increased. Based on these comparisons, we encourage the confidence thresholds for classification to be tweaked accordingly based on the application scenario. A high confidence threshold can ensure that more images are correctly labeled and is useful when the number of true negative classifications are less valued because having higher degree of classification correctness is preferred.

4. CONCLUSION

This paper demonstrates that machine learning models can be used in pruning field recordings of wildlife videography to output filtered images of the desired animal subjects such as the Chelonia mydas turtle. We demonstrated, tested, and discussed a method of utilizing a machine learning model that can be trained and may be used to assist in the identification and filtration process of the recordings. We managed to achieve up to 99.89% of output images being correctly labeled by a custom model trained with our own Chelonia mydas turtle dataset. The results obtained has shown that our approach does indeed provide an automated alternative for pruning through noisy field recordings of wildlife and can potentially assist in animal wildlife research by automating the labor needed to perform filtration on noisy field wildlife videography recordings. However, we do acknowledge that our methods may be challenging for ecologists to begin adpotion as it would require time and technical know how in order to set up, program, and train the models such as the ones used in this paper. For future work, we would like to look into the possibility of making a program package that can simplify the methods in this paper as well as explore the probability of utilizing a similar technique but for animal re-identification. Research by other researchers as well has shown that individual animal re-dentification is significantly difficult as animals may have body patterns or features that are only subtly different or insufficiently distinct between individuals thus making individual re-identification much more difficult than species identification. We believe more work can be done and will be looking forward to advances in machine learning technology to experiment various applications in relations to wildlife research and conservation.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support from the Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS) FRGS/1/2021/ICT02/UNIMAS/02/1. We the authors would like to thank the grant providers for supporting us in our research work financially. We would also like to express our thanks to the supportive staff and wardens from Sarawak Forestry Corporation for giving us their permission for our camera operators to conduct field video recordings, approving our permits and for ethical guidance, physical assistance as well as technical assistance in the wildlife fieldwork aspects of this research. Thank you as well to Universiti Malaysia Sarawak, UNIMAS, for the necessary support provided during this research.

REFERENCES

[1] U. Shahzad, "Global warming : causes , effects and solutions," *Durreesamin Journal*, vol. 1, no. 4, 2017.
[2] N. Mimura, "Sea-level rise caused by climate change and its implications for society," *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, vol. 89, no. 7, pp. 281–301, 2013, doi: 10.2183/pjab.89.281.
[3] S. Quratulann, M. Muhammad Ehsan, E. Rabia, and A. Sana, "Review on climate change and its effect on wildlife and ecosystem," *Open Journal of Environmental Biology*, pp. 8–14, 2021, doi: 10.17352/ojeb.000021.
[4] J. A. Seminoff, "Green Turtle-Chelonia mydas" *The IUCN Red List of Threatened Species 2023*: e.T4615A247654386. Southwest Fisheries Science Center, 2023, doi:10.2305/IUCN.UK.2023-1.RLTS.T4615A247654386.en.
[5] A. Hoffmann, J. Decher, F. Rovero, J. Schaer, C. Voigt, and G. Wibbelt, "Field methods and techniques for monitoring mammals," *Manual on field recording techniques and protocols for All Taxa Biodiversity Inventories and Monitoring*, vol. 8, pp. 482–529, 2010.
[6] J. Jumeau, L. Petrod, and Y. Handrich, "A comparison of camera trap and permanent recording video camera efficiency in wildlife underpasses," *Ecology and Evolution*, vol. 7, no. 18, pp. 7399–7407, 2017, doi: 10.1002/ece3.3149.
[7] H. Nguyen *et al.*, "Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring," in *Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017*, IEEE, 2017, pp. 40–49, doi: 10.1109/DSAA.2017.31.
[8] S. L. Krauss, D. G. Roberts, R. D. Phillips, and C. Edwards, "Effectiveness of camera traps for quantifying daytime and nighttime visitation by vertebrate pollinators," *Ecology and Evolution*, vol. 8, no. 18, pp. 9304–9314, 2018, doi: 10.1002/ece3.4438.
[9] T. Battu and D. S. Reddy Lakshmi, "Animal image identification and classification using deep neural networks techniques," *Measurement: Sensors*, vol. 25, pp. 1-9, 2023, doi: 10.1016/j.measen.2022.100611.
[10] 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," *International Journal of Scientific and Technology Research*, vol. 8, no. 10, pp. 2613–2622, 2019.
[11] O. R. Wearn and P. Glover-Kapfer, "Snap happy: camera traps are an effective sampling tool when compared with alternative methods," *Royal Society Open Science*, vol. 6, no. 3, pp. 1-13, 2019, doi: 10.1098/rsos.181748.
[12] M. Tan *et al.*, "Animal detection and classification from camera trap images using different mainstream object detection architectures," *Animals*, vol. 12, no. 15, pp. 1-16, 2022, doi: 10.3390/ani12151976.
[13] V. Miele, G. Dussert, B. Spataro, S. Chamaillé-Jammes, D. Allainé, and C. Bonenfant, "Revisiting animal photo-identification using deep metric learning and network analysis," *Methods in Ecology and Evolution*, vol. 12, no. 5, pp. 863–873, 2021, doi: 10.1111/2041-210X.13577.
[14] I. Hipiny, H. Ujir, A. Mujahid, and N. K. Yahya, "Towards automated biometric identification of sea turtles (Chelonia mydas)," *Journal of ICT Research and Applications*, vol. 12, no. 3, p. 256, 2018, doi: 10.5614/itbj.ict.res.appl.2018.12.3.4.
[15] S. Newey *et al.*, "Limitations of recreational camera traps for wildlife management and conservation research: A practitioner's perspective," *Ambio*, vol. 44, no. S4, pp. 624–635, 2015, doi: 10.1007/s13280-015-0713-1.
[16] D. Randula Podduwage, P. Ratnayake, D. Randula Podduwage, and Rmp. Ratnayake, "Wildlife photography over nature photography," *International Journal of Humanities and Social Science Invention*, vol. 9, no. 9, pp. 49–52, 2020, doi: 10.35629/7722-0909024952.
[17] 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, Sep. 2013, doi: 10.1186/1687-5281-2013-52.
[18] S. Leorna and T. Brinkman, "Human vs. machine: detecting wildlife in camera trap images," *Ecological Informatics*, vol. 72, pp. 1-12, 2022, doi: 10.1016/j.ecoinf.2022.101876.
[19] M. S. Norouzzadeh *et al.*, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 25, pp. E5716–E5725, 2018, doi: 10.1073/pnas.1719367115.
[20] M. A. Tabak *et al.*, "CameraTrapDetectoR: automatically detect, classify, and count animals in camera trap images using artificial intelligence," *bioRxiv*, pp. 1-22, 2022, doi: 10.1101/2022.02.07.479461.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
[22] J. Cowton, I. Kyriazakis, and J. Bacardit, "Automated individual pig localisation, tracking and behaviour metric extraction using deep learning," *IEEE Access*, vol. 7, pp. 108049–108060, 2019, doi: 10.1109/ACCESS.2019.2933060.
[23] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of YOLO v3, faster R-CNN, and SSD for real-time pill identification," *Research Square*, pp. 1-28, 2021, doi: 10.21203/rs.3.rs-668895/v1.
[24] J. A. Kim, J. Y. Sung, and S. H. Park, "Comparison of Faster-RCNN, YOLO, and SSD for Real-Time Vehicle Type Recognition," in *2020 IEEE International Conference on Consumer Electronics - Asia, ICCE-Asia 2020*, IEEE, Nov. 2020, pp. 1-4, doi: 10.1109/ICCE-Asia49877.2020.9277040.
[25] M. Koziarski and B. Cyganek, "Impact of low resolution on image recognition with deep neural networks: An experimental study," *International Journal of Applied Mathematics and Computer Science*, vol. 28, no. 4, pp. 735–744, 2018, doi: 10.2478/amcs-2018-0056.
[26] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *2016 8th International Conference on Quality of Multimedia Experience, QoMEX 2016*, IEEE, 2016, pp. 1-6, doi: 10.1109/QoMEX.2016.7498955.
[27] B. Funt and L. Zhu, "Does colour really matter? evaluation via object classification," *Final Program and Proceedings - IS and T/SID Color Imaging Conference*, vol. 2018, no. 1, pp. 268–271, Nov. 2018, doi: 10.2352/issn.2169-2629.2018.26.268.
[28] S. Wenkel, K. Alhazmi, T. Liiv, S. Alrshoud, and M. Simon, "Confidence score: the forgotten dimension of object detection performance evaluation," *Sensors*, vol. 21, no. 13, pp. 1-21, 2021, doi: 10.3390/s21134350.
[29] C. G. Northcutt, A. Athalye, and J. Mueller, "Pervasive label errors in test sets destabilize machine learning benchmarks," *arXiv-Statistics*, pp. 1-24, 2021.
[30] P. Gray, M. Mckown, A. Fleishman, and V. S. Bézy, "A convolutional neural network for detecting sea turtles in drone imagery, " *Methods in Ecology and Evolution*, vol. 10, no. 3, pp. 345-355, 2019, doi: 10.1111/2041-210X.13132.
[31] R. Chen, R. Little, L. Mihaylova, R. Delahay, and R. Cox, "Wildlife surveillance using deep learning methods," *Ecology and Evolution*, vol. 9, no. 17, pp. 9453–9466, 2019, doi: 10.1002/ece3.5410.
[32] M. Kutugata, J. Baumgardt, J. A. Goolsby, and A. E. Racelis, "Automatic camera-trap classification using wildlife-specific deep learning in nilgai management," *Journal of Fish and Wildlife Management*, vol. 12, no. 2, pp. 412–421, 2021, doi: 10.3996/JFWM-20-076.

BIOGRAPHIES OF AUTHORS

Khalif Amir Zakry holds a B.Sc. (2020) in Computer Science (Software Engineering) from Universiti Malaysia Sarawak (UNIMAS) with his final year thesis titled "Contract distribution system framework with case study on pc building market". He was a system engineer for Sarawak Information Systems, working specifically in the field of audiovisual engineering. He is currently pursuing his master education at the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak (UNIMAS). His research areas of interest include artificial intelligence and computer vision. He can be contacted at email: 21020463@siswa.unimas.my.

Mohamad Syahiran Soria holds a B.Sc. (2021) in Computer Science (Software Engineering) from Universiti Malaysia Sarawak (UNIMAS) with his final year thesis titled "KamekTravel application system using hybrid mobile app development, ionic framework". He is currently pursuing his master education at the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak (UNIMAS). His research areas of interest include artificial intelligence and computer vision. He can be contacted at email: 22020014@siswa.unimas.my.

Irwandi Hipni Mohamad Hipiny is currently a senior lecturer (DS52) at Universiti Malaysia Sarawak. He holds a B.Sc. (2003) and an M.Sc. (2007) in Computer Science from Universiti Teknologi Malaysia; later graduated with a Ph.D. (2014) in Computer Vision from University of Bristol, UK. Prior to his current appointment, Irwandi was a recipient of Ekpress UTM-MARA scholarship (1999–2003), and a MOSTI's National Science fellow (2003-2007). His current research interests are, but not limited to, computer vision, pattern recognition, and animal Re-ID. He can be contacted at email: mhihipni@unimas.my.

Hamimah Ujir is an associate professor at the Faculty of Computer Science and Information Technology (FCSIT), Universiti Malaysia Sarawak (UNIMAS). In 2003, she received a National Science Fellowship scholarship award from the Ministry of Science, Technology, and Innovation (MOSTI) and completed her master's degree (by research) at the Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, Skudai, Johor. In March 2009, she pursued her Ph.D. in the School of Electronic, Electrical and Computer Engineering, University of Birmingham, United Kingdom. Her previous and current works include 3D physical simulation and 3D static and dynamic facial expression analysis. She can be contacted at email: uhamimah@unimas.my.

Ruhana Hassan Graduated with a Bachelor Degree in Aquatic Biology from Universiti Sains Malaysia in 1993 and attained her Master Degree of Science in Biology from the same university several years later. She received her Doctorate in Biological Sciences from the University of Birmingham in 2004. She has been a Ph.D. supervisor for several students (5) with one student currently ongoing. She has won medals in various research expos such as UNIMAS INTEX, Malaysian Technology Expo (MTE) and Pre-University Malaysia Innovation Competition (PIITRAM). She can be contacted at email: hruhana@unimas.my.

Richki Hardi is a researcher and lecturer at the Department of Informatics, Faculty of Computer Science, Universitas Mulia (UM). He pursued his education in Informatics at Ahmad Dahlan University, Electrical Engineering at Gadjah Mada University, and Information and Communication Technology at the Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka. He has previously received research awards from the Ministry of Education and obtained grants for smart classrooms for institutions. He is active as an organizer of renowned international scientific conferences and serves as a reviewer both domestically and internationally. His past and current works encompass artificial intelligence, data science, and information systems. He can be contacted at email: richki@universitasmulia.ac.id.