Deep Neural Networks based Multiclass Animal Detection and Classification in Drone Imagery

*

1st Changrong Chen
Department of Computer Science
Loughborough University
Loughborough, UK
C.Chen2@lboro.ac.uk

2th E.A. Edirisinghe
School of Computing and Mathematics
Keele University
Keele, UK
E.Edirisinghe@keele.ac.uk

3th Andrew Leonce
College of Technology Innovation
Zayed University
United Arab Emirates
Andrew.Leonce@zu.ac.ae

4rd Gregory Simkins
Shaybah Wildlife Sanctuary
Kingdom of Saudi Arabia
greg.simkins@gmail.com

5th Tamer Khafaga
King Salman Royal Nature Reserve
Kingdom of Saudi Arabia
t.khafaga@ksrnr.gov.sa

6th Moayyed Sher Shah
King Salman Royal Nature Reserve
Kingdom of Saudi Arabia
m.shah@ksrnr.gov.sa

7th Umar Yahya
Department of Computer Science
lslamic University
Kampala, Uganda
umar.yahya@ieee.org

2023 International Symposium on Networks, Computers and Communications (ISNCC) | 979-8-3503-3559-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISNCC58260.2023.10323685

***Abstract*—There is a growing interest among the research community in the search for possible technology-driven strategies for the conservation of the much-needed, historically rich and culturally important, desert life. In this work, we investigate the use of one of the best available Deep Neural Networks, YOLO Version-5 (v5), to enable offline detection, identification and classification of three popular desert animals (i.e Camels, Oryxes, and Gazelles) in a Drone Imagery Dataset captured by the Dubai Desert Conservation Reserve (DDCR), United Arab Emirates. The dataset contains over 1200 images, which were partitioned into training, validation, and testing data sub-sets in a 8:1:1 ratio, respectively. We trained three multi-class models, animal classification models, based on YOLO v5 Small(S), Medium(M) and Large(L), representing increasingly deep and complex architectures, to simultaneously detect and label the 3 kinds of animals. Models' performance was compared on the basis of classification accuracy (F1-Measure). The multi-class detector models generated were also compared with the single animal detector models created using the same network architectures, to assess the trained network's robustness against detecting more than one class of object. YOLO v5 L achieved the highest multi-class average classification accuracy of 96.71 percent (95.39 – 98.98). In comparison with the single animal detector models, the multi-class models exhibited the ability to correctly detect the target objects even for cases where the objects are located close to each other. We show that the promising results achieved in this work provide a promising foundation for the development of real-time multiclass identification and classification applications utilizing UAV imagery, to aid in the conservation efforts of fauna, particularly in the urbanized modern-day deserts and semi-desert places, such as the DDCR. We provide comprehensive test results and an analysis of results to demonstrate the effectiveness of the proposed models.**

***Index Terms*—YOLO v5, Multi-class animal detector, performance comparison, confusion matrix, drone images analytics, drone imagery, desert ecology**

979-8-3503-3559-0/23/$31.00 2023 IEEE

I. Introduction

As the world continues to experience several ambitious urbanization and infrastructural developments, the importance of conserving the fauna cannot be overemphasized. Particularly crucial to conservation, is the continuous monitoring and observation of wildlife in deserts and semi-desert regions, as this consequently facilitates their protection. In the year 2000, the government of Dubai designated a 225-square-kilometer land as the Dubai Desert Conservation Reserve (DDCR) to protect the different animal and plant species native to the Dubai Desert [1]. Famous among the desert animals of the United Arab Emirates (UAE) are the Oryxes, Camels, and Gazelles [2]. In order to better observe and thus monitor the activities of animals in the DDCR, the use of drone-based surveillance has been adopted recently. However, the analysis of the image content has been restricted to manual means, i.e. human observers conducting the analysis based on the captured drone footage [2]. Based on images captured from UAVs and drones that give an advantageous view direction for video / image analysis, the success of manual approaches to ecological image analysis using RGB cameras and basic automated approaches to image analysis based on captured multi-spectral images, have been well demonstrated in literature [3–17]. In addition to this the use of the more

traditional machine learning approaches [17] and the more recent deep learning based approaches for done based animal detection [4], remote sensing [16] and analysing camera trap data (i.e. cameras set at ground-level) [18–21] for automatic, computer based, wildlife/ecological/livestock analysis, have been demonstrated in more recent literature. Drone-based surveillance offers several benefits including low noise image capture, long-range and fast coverage, being able to maximize the observation of animals in a reserve without affecting their activities, providing a unique view direction that is not available to capture based on cameras set up at ground level (e.g. camera traps) and real-time, high-resolution data capture at a relatively reasonable cost [14] . Additionally, drones also effectively allow animals to be observed while moving along with them, thereby enabling additional opportunity to collect more useful data such as grazing patterns and other habitat aspects relevant to the conservation of the monitored animals. However, a common challenge with drone-based surveillance is the considerable amount of time and effort required to manually do visual inspection of the obtained drone imagery [15] . These challenges have consequently inspired the recent growing interest within the research communities of ecological remote sensing, wildlife protection, machine learning, to attempt to use computer-based, object detection or/and classification algorithms/models for automated drone image analysis [4, 16, 17]. Further, recently, there have been several attempts of using the more recent advances in Deep Neural Networks for detecting multiple animal classes in Camera Trap Images [18–21] and in drone footage [4] .

Although Deep Neural Networks are superior in their ability to detect and recognise objects under challenging environmental (night-time, dust, fog, etc.), lighting (shadows, illumination intensity, colour constancy etc.), and spatial (i.e., orientation, occlusion etc.) conditions, their effective use is not yet demonstrated widely in many application areas. Effectively training Deep Neural Network models to detect and recognise objects under challenging conditions require, domain knowledge of the practical problem being addressed, knowledge of how the various architectures of Deep Neural Networks can be optimised and tailored for use in solving a given problem, substantial amount of data for training and following careful approaches to data labelling, which is an essential, but is often ineffectively carried out. Therefore, a truly inter-disciplinary approach to research needs to be followed, with careful consideration given to the selection of the right network, optimally selecting its parameters, ensuring the right amount and type of data is used and the data is appropriately labelled, for training.

In this paper, we utilize a comprehensive drone imagery dataset generated by the DDCR, Dubai drone-team (captured in collaboration and with guidance from the authors) to effectively train a Deep Neural Network based on the latest version of the popular CNN architecture, You Only Look Once (YOLO) – Version 5 [22] , to automate the detection and classification of the three types of desert animals, namely Oryxes, Camels, and Gazelles. YOLOv5 is one of the most effective one-stage CNN architectures used in many practical object detection and recognition applications at present [22, 23] , but it's use in animal / wild-life detection and recognition, based on drone footage, has not yet been fully exploited to-date. In our previous work [24] we demonstrated the use of YOLOv5 for the detection of single type of animals, i.e. Oryxes, but using one model that can detect multiple types of animals calls for more careful design, solving the challenges around class-imbalance, between the animal types. A further popular one-stage CNN architecture is the Single Shor Detector (SSD) [25]. However, our previous research [24] demonstrated that YOLOv5 is much more accurate that SSD in detecting objects/animals as seen from drones. Alternatively, two-stage CNN architectures such as, R-CNN [26] , and Fast R-CNN [27] have better accuracy for specific tasks, but need substantial time for training and requires a large computing power to train. They are also not designed in a way that they can locate the detected objects within frames. Therefore, a better speed-accuracy tradeoff would be to utilize one-stage methods, such as YOLO as opted for in this current work. The ultimate aim of this research was to support the team of ecological experts within DDCR by automating the detection and classification of dessert wildlife and other protected animals, based on footage captured by surveillance drones, allowing the team to monitor the whereabouts, movements and bahaviour of such herds. In the proposed research we develop three different animal detection and recognition CNN models from YOLO-V5 sub-version architectures, namely, S(small), M(medium) and L(large) and compare their performance and suitability.

For clarity of presentation, this paper is divided into several sub-sections. Section-2 proposes the research background, including a brief introduction to the YOLOv5, it's sub-version architectures and defines the network anetwork architecture, it's associated methodology adopted for creating the CNN models for animal detection and recognition. Section-3 provides experimental results and a comprehensive analysis of the results. Finally section-4 concludes with an insight to potential future research.

II. Proposed Methodology

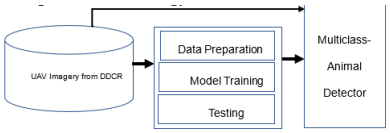


Fig. 1. The flow diagram of the project framework

The YOLO architecture has evolved over time and is currently in it's 5th version, YOLOv5[22] . Though having a similar backbone architecture, the YOLOv5 architecture can further be subdivided into sub-version architectures, namely, YOLOv5S, YOLOv5M and YOLOv5L, corresponding to smallest to highest depth of network and the size of feature map used, respectively. i.e YOLOv5L has the highest network depth as well as the largest feature map, while YOLOv5S

has the smallest network depth and the smallest feature map size. Due to this it is the least complex network, and hence training is faster and deployment of models created requires lets computational capacity. However , YOLOv5S architecture often produces the lowest average precision (AP) accuracy. It is the preferred network of the three if the target object that needs detection is sufficiently large enough. For the M and L YOLOv5 architectures, the network becomes deeper and more complex, and therefore the AP accuracy increases accordingly, but at the expense of speed (i.e more time is needed for training and testing). It is noted that a fourth sub-version of YOLOv5 exists, V5X (extra-large) and due to the added complexity of its architecture we have excluded its use in this paper.

The complexity of a Deep Neural Network architecture is typically defined by the network depth and width, which are often represented by the depth-multiple and the width-multiple respectively, as per previous implementation of YOLO-V5 [23] . The Table I indicates the values utilized for the two network parameters, depth and width, for the three different YOLOv5 architectures, trained and tested in this work.

TABLE I
THE DEPTH AND WIDTH FOR YOLO-V5 S, M, AND L FROM THE CODE OF BACKBONE. [23]

YOLOv5 Backbone	Depth	width	CSP units
YOLOv5S	0.33	0.5	128
YOLOv5M	0.67	0.75	256
YOLOv5L	1.0	1.0	512

Table I illustrates that moving from S, M to L, both the network depth and width increases. When the network becomes deeper, the ability of feature extraction and integration will become better. On the other hand, when the width increases, during the convolution process, the so-called Focus structure [28] of the neural network, uses more convolution kernels, making the feature map wider. For example, in S, the network uses 32 convolution kernels and the feature map is 304*304*32, whereas M uses 48 convolution kernels and the feature map thus used is 304*304*48. With the network becoming wider, it gains a better learning ability due to the extraction of more features.

During training, it is important to appropriately set the batch-size, and to change the model weights accordingly. The three YOLOv5 networks we have used, i.e., S, M and L, have different weight files. If a batch-size that is too large is selected [29], the reduction of training loss will become hard due to the generalization gap becoming large. On the other hand, a large batch-size will need more memory, and a small batch-size will need more time for training. Given the above reasoning, we have chosen a batch-size of 16 for v5S and batch-size of 8 for v5M and v5L.

The dataset available for training and testing comprised of approximately 1200 images of oryxes, gazelles and camels captured/photographed from a drone. However, due to the need of maintaining class-balance in training, we were only able to use 300 images of each type of animal for training. 40 drone captured images were used for testing, having variable numbers of each of the three types of animals . The training and test mages were randomly selected from the available images of the specific animal type from within the original 1200 image set. It is vital that class balance is maintained to optimize the accuracy of detecting all three types of animals.

The LabelImg [30] tool was used in labelling the samples of each animal type. The tool provides a simple way to draw a rectangle around an identified animal and giving the labelled area a name, i.e. oryx, gazelle or camel. The tool automatically picks up the coordinates of the four corners of the rectangle drawn (from which the model training process can capture a sub-image for training the deep neural network) and assigns a label for the selected rectangle. The coordinates and the label data is saved in a text file to be later used in training.

$$Accuracy(allcorrect/all) = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

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

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1 = \frac{2 * Precision * Reacll}{Precision + Recall} \quad (4)$$

After training the three YOLOv5 architectures, the trained models were evaluated on a reserved (i.e. a dataset not used for training) test data subset and their performance were compared based on the observed confusion matrices. The confusion matrix is defined as per Table below, given prediction X, the following terms hold:

Predicted Class	True Class	
Positive	True Positive (TP): X prediction is positive and X is true.	False Positive(FP): X Prediction is positive and X is false.
Negative	False Negative (FN): X prediction is negative and X is true.	True Negative(TN): X Prediction is negative and X is false.

III. EXPERIMENTAL RESULTS

The following sections present a rigorous comparison of experimental results obtained, when the three Deep Neural Network models created by training the three sub-versions of YOLOv5, S, M and L, were used for the detection and classification of Oryxes, Gazelles and Camels in drone footage.

A. Multiclass Classification Results for each of the three YOLOv5 configurations (YOLOv5 S, M, L)

1) YOLO-V5s: Table II tabulates the prediction results obtained when using YOLO-V5S. It indicates that out of the 603 Camels, 540 were correctly classified as Camels (True Positives), 2 were classified wrongly as Oryxes, one wrongly as a Gazelle, and the remaining 60 were not detected, i.e., a total of 63 False Negatives. Out of the 174 Oryxes, 161 were correctly classified as Oryxes, none of them were wrongly classified as a Camel, one was wrongly classified as a Gazelle

and 12 were not detected. Finally, out of the 239 Gazelles, 236 were correctly classified as Gazelles, one was wrongly classified as a Oryx and 2 were not detected. Table also lists that 127 unrelated objects such as trees, white stones, etc., were wrongly detected and classifies as Camels (i.e. FP = 127 + 2 + 1 = 130), 13 such objects were wrongly detected and classified as Oryxes and 63 were wrongly classified as Gazelles.

TABLE II
THE CONFUSION MATRIX FOR YOLO-V5S

Yolo V5S		Predicted Class(1134labels)		
		Camel	Oryx	Gazelle
Actual Class	Camel(603)	540	2	1
	Oryx(174)	0	161	1
	Gazelle(239)	0	1	236
	Others	127	13	63

2) YOLO-V5M: Table III tabulates the prediction results obtained when using YOLO-V5m. It indicates that out of the 603 Camels, 574 were correctly classified as Camels (True Positives), none were classified wrongly as Oryxes or as Gazelles, and the remaining camels were not detected, i.e., a total of 27 False Negatives. Out of the 174 Oryxes, 159 were correctly classified as Oryxes, one of them were wrongly classified as a Camel, none was wrongly classified as a Gazelle and 14 were not detected. Finally, out of the 239 Gazelles, 238 were correctly classified as Gazelles, one was wrongly classified as a Oryx and none were identified as Camels. Table also lists that 83 unrelated objects such as trees, white stones, etc., were wrongly detected and classifies as Camels (i.e. FP = 83 + 0 + 1 = 84), 16 such objects were wrongly detected and classified as Oryxes and 76 were wrongly classified as Gazelles.

TABLE III
THE CONFUSION MATRIX FOR YOLO-V5M

Yolo V5M		Predicted Class(1150labels)		
		Camel	Oryx	Gazelle
Actual Class	Camel(603)	574	0	0
	Oryx(174)	1	159	0
	Gazelle(239)	0	1	238
	Others	83	16	76

3) YOLO-V5L: Table IV tabulates the prediction results obtained when using YOLO-V5L. It indicates that out of the 603 Camels, 572 were correctly classified as Camels (True Positives), two were classified wrongly as Oryxes and none were wrongly classified as Gazelles, and the remaining camels were not detected, i.e., a total of 27 False Negatives. Out of the 174 Oryxes, 169 were correctly classified as Oryxes, none of them were wrongly classified as Camels nor Gazelles, and 5 were not detected. Finally, out of the 239 Gazelles, none were wrongly classified as Camels or Oryxes. Table also lists that 47 unrelated objects such as trees, white stones, etc., were wrongly detected and classifies as Camels (i.e. FP = 47 + 0 + 2 = 49), 9 such objects were wrongly detected and classified as Oryxes and 45 were wrongly classified as Gazelles.

TABLE IV
THE CONFUSION MATRIX FOR YOLO-V5L

Yolo V5L		Predicted Class(1150 labels)		
		Camel	Oryx	Gazelle
Actual Class	Camel(603)	572	2	0
	Oryx(174)	1	169	0
	Gazelle(239)	0	1	239
	Others	47	9	45

The subjective performance comparison of the performance of the single object detection models developed above, using the three sub versions of YOLO-V5, for detecting Oryx, Camels and Gazelles, are illustrated in Fig 2-4 below.

Fig 2, 3 and 4 demonstrate the subjective performance of the three models on a set of selected images that are representative of test and training image datasets used in the experiments. Although the three images of Fig.2 illustrate similar number of detections, a closer look at the confidence values of objects in the three images indicate that will YOLO-V5L, the detection confidence as a Oryx is much higher than in the case of the other two models. It is noted that the size of the Oryxes are much larger in Fig.2 as compared to the sizes in Camels and Gazelles. Fig.3 demonstrates that YOLO-V5L detects more camels on the bottom right-hand corner of the image, not detected by the other two models. A closer investigation of Fig.4 reveals that the models created by the sub versions S and M results in false detections at the top right hand corner of the images, which are avoided by the model created by YOLO-V5L. It is also noted in Fig.4 that the YOLO-V5M has more false positives than YOLO-V5S. This is likely to be due to the fact that the training data used in the experiments is not sufficient to train the deeper network of YOLO-V5M.

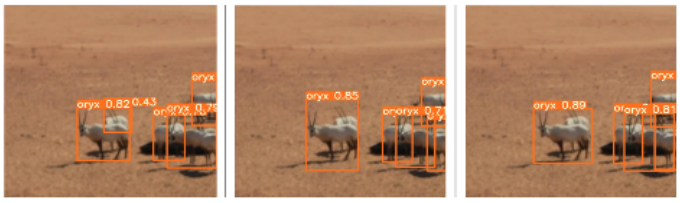


Fig. 2. The oryx detection results from YOLO-V5 S, M, and L

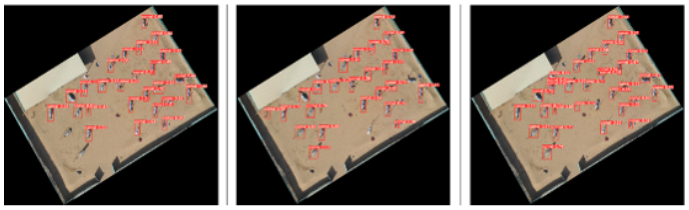


Fig. 3. The camel detection results from YOLO-V5 S, M, and L

Fig. 4. The gazelle detection results from YOLO-V5 S, M, and L

B. Comparison of performance of single-object detection models and the multiclass detector model, when using each of the three YOLO configurations (YOLOv5S, M, L)

The multi-class detector was trained on 300 samples of each type of animals. Although we had 3500 labelled Oryxes and 1600 labelled Gazelles, as we only had 300 Camels, to maintain class-balance during training, 300 samples of Oryxes and 300 samples of Gazelles were randomly picked up and used in training along with the 300 samples of Camels.

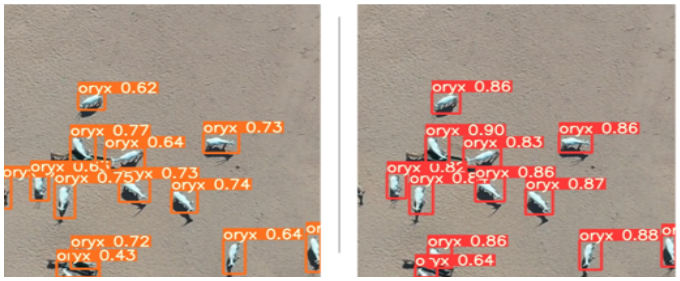
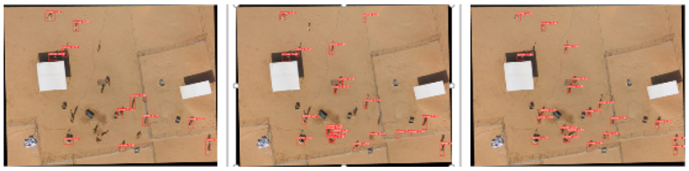


Fig. 5. [Left] The multi-class animal detector (YOLO-V5L) and the [Right] best single animal detector (YOLO-V5L) for oryx detection.

Fig. 6. [Left] The multi-class animal detector (YOLO-V5L) and the [Right] best single animal detector (YOLO-V5L) for gazelle detection.



(a) (b) (c)

Fig. 7. [Left] The multi-class animal detector (YOLO-V5L) and the [Right] best single animal detector (YOLO-V5L) for camel detection.

In this section we compare the performance of the multi-class animal detector with single animal detectors modelled using 3500, 1600 and 300, Oryxes, Gazelles and Camels, respectively. Fig.5 and Fig.6 provides a comparison of results for between the best (out of the sub-versions) YOLO-V5 based model, YOLO-V5L (Right image), and when using the multi-class classifier (left-image). Whilst both models detect all animals (Oryxes and Gazelles), the single-class model indicates a higher level of confidence in detections. This is expected as it only deals with one class, rather than three. The confidence levels indicated by the multi-class detector is still very much comparable with that of the two single-class detectors.

Out of the three types of animals being considered in this research, drove video footage used for observing Camels are the lowest in resolution. This was to restrictions around the minimum altitude a drone is given permission to fly over human occupied camp sites, as against in nature reserves. Unfortunately, this means that the size of camels used in training and testing are significantly smaller than the size of the animals of the other two classes. When we compared the performance of the sub versions of YOLO-V5, within the multi-animal-detection model architecture, we found that YOLO-V5M and 5L performs much better than YOLO-V5S. This is because a deeper architecture is needed to create effective detection models for smaller objects which could otherwise be difficult to differentiate with other smaller objects. Fig.7 (b) and (c) illustrate the outcomes. It is seen that when using YOLO-V5L, the false detections are far less than when using YOLO-V5M. The YOLO-V5L model for multi animal detection picked up all camels and did not pick up any false positives. We illustrate the results of the single-animal detection model for Camels in Fig.7 (a). It is seen that some camels have not been detected at all. However the best single-animal detector model obtained was based in YOLO-V5S, which will find difficult to analyse fine features of objects to contribute to a positive detection result. When using the single-animal detection approach, we found that YOLO-V5M and YOLO-V5L models did not perform effectively. The 300 samples used for training for a single type of obvjects was no sufficient to fine tune the deeper neural network architectures.

In the natural drone footage, we were able to capture, it was very rarely that we were able to find images that included more than one type of animals, when considering Camels, Oryxes and Gazelles. Camels are domesticated animals and Oryxes and Gazelles live in the wild or animal sanctuaries, in herds, and hence the three types are seen together very rarely. Therefore in Fig 8-10, we use Windows Image Capture to create various image mosaics that mimics the collective presence of these animals to test the single-class detectors and the multi-class classifier further in environments that include multiple type of animals.

Fig.8 illustrates the outcomes of applying the multi-animal detector and the single-animal detectors trained for Oryx only or Gazelle only detection on the mosaic image that includes both Oryxes and Gazelles. Fig.8 (a) shows the effective ability of the multi-animal detector to differentiate between Gazelles and Oryxes. The two types of animals have been differentiated

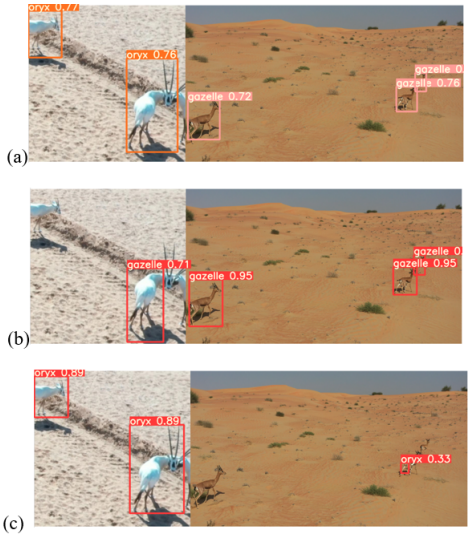


Fig. 8. Application of (a) multi-animal detector, (b) Gazelle-only detector, (c) Oryx-only detector on mosaic images

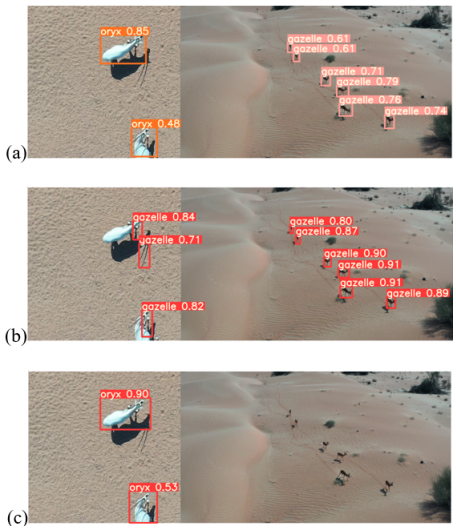


Fig. 9. Application of (a) multi-animal detector, (b) Gazelle-only detector, (c) Oryx-only detector on mosaic images

with confidence of over 0.7 in all cases of detections. None of the animals have been mis-classified or not detected and there are no false positives. Animals part occluded have also been accurately detected and classified. Fig.8 (b) shows the failure of the Gazelle only detector to prevent the detection of Oryxes as Gazelles, although the confidence values indicated in detecting and labelling a Oryx as a Gazelle is marginally lower than the confidence shown in a true-positive Gazelle detection. However in Fig.8 (c) it shows that the Oryx only detector has been able to successfully avoid picking up Gazelles and Oryxes. This is due to the fact that the Oryx only detector has been better trained with more samples or Oryxes taken at various altitudes/sizes, orientations and in different environments. It knows how to identify an Oryx more accurately.

Fig.9 and Fig.10 provides further experimental results on two different mosaic images. The conclusions above can be further justified by the results illustrated.

C. Overall performance comparison between the three YOLO-V5 sub-configurations (YOLO-V5- S, M and L)

Table V and VI bellow tabulates, Accuracy, Precision/Recall and TP/FR/F1-score values respectively when using YOLO-V5 (S, M amf L) models for detecting the three types of animals.

From the results tabulated in Table V and VI, it is seen that YOLO-V5L has the highest accuracy, procession, recall, TP and F1 scores and the lest FPs for all three types of animals. However, there is a significant difference between the performance of YOLO-V5 sub-versions, in detecting smaller sized animals (camels) used in training and test data. As mentioned previously the footage of camels were taken from high altitude drone flights making their size relatively small in the images used for training and testing. It is clear that YOLO-V5S did not perform well in the detections of such small sized

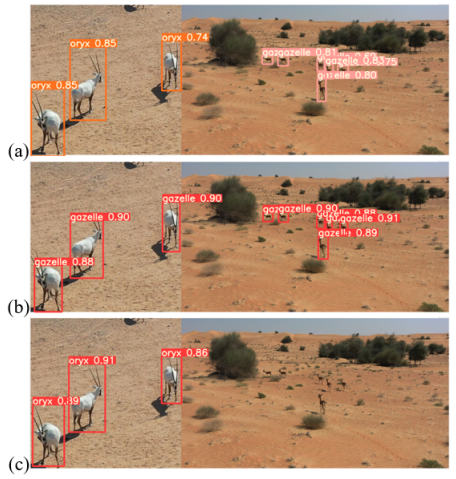


Fig. 10. Application of (a) multi-animal detector, (b) Gazelle-only detector, (c) Oryx-only detector on mosaic images

TABLE V
COMPARISON OF ACCURACY, PRECISION AND RECALL

Accuracy/precision/Recall	Camel	Oryx	Gazelle
S	88.54%/ 80.60%/ 89.55%	98.5%/ 92%/ 92.53%	94.44%/ 78.67%/ 99.16%
M	92.52%/ 87.37%/ 95.20%	98.43%/ 90.34%/ 91.91%	93.13%/ 75.15%/ 99.58%
L	95.39%/ 92.11%/ 94.86%	98.98%/ 94.94%/ 97.13%	95.76%/ 84.75%/ 99.99%

TABLE VI
COMPARISON OF TP/FP/F1-SCORE

TP/FP/F1	Camel	Oryx	Gazelle
S	540/130/84.84%	161/14/92.26%	236/64/87.57%
M	574/83/91.11%	159/17/90.86%	238/79/85.61%
L	572/49/93.46%	169/9/96.02%	239/46/91.22%

animals. The best performance of YOLO-V5 S model was in detecting Oryxes, the largest sized animals used in training and testing. When it comes to analysing the performance of models in detecting Gazelles, the variations of background on top of the smaller size, impacts negatively on precision and recall, more specifically. The lowest recorded precision and recall values are in the detection of Gazelles.

We also investigated the performance of the single animal detection models we created. The results demonstrated that in most cases, the YOLO-V5M models have a similar detection accuracy result as compared with YOLO-V5L models, particularly when the objects to be detected are sufficiently large, e.g, in detecting Oryxes. But when the object is small and not well differentiated from the background (Camels, Gazelles) the YOLO-V5L models demonstrated better performance . However, it was also clear that deeper the network architecture, more data is needed for effective training. Further with the increasement of data used in training, the performance of multi-class detector models exceeded the performance on single class detector models.

The overall results and detailed analysis above leads to the recommendation that multi-class animal detectors that can be created using YOLO-V5 sub-versions, requires significant amount of balanced data for training. The training data should have sufficient variations in resolution, size, angle, changes in background and illumination, occlusions etc., to be able to build the most accurate animal detectors and classifiers that can be applied on drone captured video footage.

IV. Conclusion

This paper investigated the effective use of Deep Neural Network architectures in the detection and classification of multiple animal types, Camels, Oryxes and Gazelles, in images captured from drones. The work presented compared the use of the Single Shot Detection DNN with the use of YOLO-V5 and its sub-versions, S, M and L. The results concluded the superiority of YOLO-V5 over SSD and the ability of its sub versions with a deeper architecture to accurately detect and classify the animals. The importance of adopting effective approaches to training the DNN architectures, the impact of class-balance, low resolution imagery and the size of objects being detected and their relationship to the different sub-versions of YOLO-V5 was investigated, establishing a number of recommendations for their practical use. The research also presented single animal detection models created using the sub-versions of YOLO-V5 and compared their performance with the performance of the multiple-animal-detector models derived. It is concluded that YOLO-V5 provides suitable architectures for the accurate detection and classification of animals via drone footage, under significant variations of image quality, altitude of drone flying, size and inter/intra class variations. Accuracy levels of above 98 percent has been achieved with above 94 percent precision and recall 97 percent recall for detecting Oryxes, for which the best data was available for training.

Acknowledgment

Authors acknowledge the permission and assistance given by the Dubai Desert Conservation Reserve (DDCR) in capturing drone imagery. At the time of conducting this research the authors, Simkins, Khafaga and Shah, were employees of DDCR. The images used in this research were provided to Loughborough University by the DDCR, under a formal agreement. The drone images were captured by authors, Leonce and Shah.

References

[1] D. J. Gallacher and J. P. Hill, "Effects of camel vs oryx and gazelle grazing on the plant ecology of the dubai desert conservation reserve," 2009.

[2] H. E. Alqamy, "Mammals of dubai desert conservation reserve: Initial assessment and baseline data. using camera-traps," 2010.

[3] J. J. López and M. Mulero-Pázmány, "Drones for conservation in protected areas: Present and future," *Drones*, 2019.

[4] B. Kellenberger, D. Marcos, and D. Tuia, "Detecting mammals in uav images: Best practices to address a substantially imbalanced dataset with deep learning," *ArXiv*, vol. abs/1806.11368, 2018.

[5] G. Schofield, K. A. Katselidis, M. K. S. Lilley, R. D. Reina, and G. C. Hays, "Detecting elusive aspects of wildlife ecology using drones: new insights on the mating dynamics and operational sex ratios of sea turtles," *Functional Ecology*, vol. 31, pp. 2310–2319, 2017.

[6] S. L. Pimm, S. K. Alibhai, R. A. Bergl, A. Dehgan, C. P. Giri, Z. C. Jewell, L. N. Joppa, R. W. Kays, and S. Loarie, "Emerging technologies to conserve biodiversity." *Trends in ecology & evolution*, vol. 30 11, pp. 685–696, 2015.

[7] M. Mulero-Pázmány, R. Stolper, L. D. van Essen, J. J. Negro, and T. Sassen, "Remotely piloted aircraft systems as a rhinoceros anti-poaching tool in africa," *PLoS ONE*, vol. 9, 2014.

[8] R. Näsi, E. Honkavaara, P. Lyytikäinen-Saarenmaa, M. Blomqvist, P. Litkey, T. Hakala, N. Viljanen, T. Kantola, T. Tanhuanpää, and M. Holopainen, "Using uav-based photogrammetry and hyperspectral imaging for mapping bark beetle damage at tree-level," *Remote. Sens.*, vol. 7, pp. 15 467–15 493, 2015.

[9] D. Chabot and D. M. Bird, "Wildlife research and management methods in the 21st century: Where do unmanned aircraft fit in?1," 2015.

[10] T. Adão, J. Hruska, L. Pádua, J. E. Bessa, E. Peres, R. Morais, and J. J. Sousa, "Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry," *Remote. Sens.*, vol. 9, p. 1110, 2017.

[11] S. Manfreda, M. F. Mccabe, P. E. Miller, R. M. Lucas, V. P. Madrigal, G. Mallinis, E. Ben-Dor, D. Helman, L. D. Estes, G. Ciraolo, J. Müllerová, F. Tauro, M. I. de Lima, J. L. M. P. de Lima, A. Maltese, F. Francés, K. K. Caylor, M. Kohv, M. T. Perks, G. Ruiz-Pérez,

Z. Su, G. Vico, and B. Tóth, "On the use of unmanned aerial systems for environmental monitoring," *Remote. Sens.*, vol. 10, p. 641, 2018.

[12] L. P. Koh and S. A. Wich, "Dawn of drone ecology: Low-cost autonomous aerial vehicles for conservation," *Tropical Conservation Science*, vol. 5, pp. 121 – 132, 2012.

[13] K. S. Christie, S. L. Gilbert, C. L. Brown, M. Hatfield, and L. Hanson, "Unmanned aircraft systems in wildlife research: current and future applications of a transformative technology," *Frontiers in Ecology and the Environment*, vol. 14, pp. 241–251, 2016.

[14] R. Díaz-Delgado and S. Mücher, "Editorial of special issue "drones for biodiversity conservation and ecological monitoring"," *Drones*, 2019.

[15] J. Linchant, J. Lisein, J. Semeki, P. Lejeune, and C. Vermeulen, "Are unmanned aircraft systems (uass) the future of wildlife monitoring? a review of accomplishments and challenges," *Mammal Review*, vol. 45, pp. 239–252, 2015.

[16] L. P. Osco, J. M. Junior, A. P. M. Ramos, L. Jorge, S. N. Fatholahi, J. de Andrade Silva, E. T. Matsubara, H. Pistori, W. N. Gonçalves, and J. Li, "A review on deep learning in uav remote sensing," *Int. J. Appl. Earth Obs. Geoinformation*, vol. 102, p. 102456, 2021.

[17] A. I. Khan and Y. A. Al-Mulla, "Unmanned aerial vehicle in the machine learning environment," in *EUSPN/ICTH*, 2019.

[18] A. Gomez-Villa, A. Salazar, and J. F. Vargas-Bonilla, "Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks," *ArXiv*, vol. abs/1603.06169, 2016.

[19] S. Schneider, G. W. Taylor, and S. C. Kremer, "Deep learning object detection methods for ecological camera trap data," *2018 15th Conference on Computer and Robot Vision (CRV)*, pp. 321–328, 2018.

[20] M. Willi, R. T. Pitman, A. W. Cardoso, C. M. Locke, A. Swanson, A. Boyer, M. Veldthuis, and L. Fortson, "Identifying animal species in camera trap images using deep learning and citizen science," *Methods in Ecology and Evolution*, vol. 10, pp. 80 – 91, 2018.

[21] M. A. Tabak, M. S. Norouzzadeh, D. W. Wolfson, S. J. Sweeney, K. C. Vercauteren, N. P. Snow, J. M. Halseth, P. A. D. Salvo, J. S. Lewis, M. White, B. S. Teton, J. C. Beasley, P. E. Schlichting, R. K. Boughton, B. Wight, E. S. Newkirk, J. S. Ivan, E. A. Odell, R. K. Brook, P. M. Lukacs, A. K. Moeller, E. G. Mandeville, J. Clune, and R. S. Miller, "Machine learning to classify animal species in camera trap images: applications in ecology," *bioRxiv*, 2018.

[22] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2015.

[23] G. Yang, W. Feng, J. Jin, Q. Lei, X. Li, G. Gui, and W. Wang, "Face mask recognition system with yolov5 based on image recognition," *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 1398–1404, 2020.

[24] T. Jintasuttisak, A. N. J. Leonce, M. S. Shah, T. Khafaga, G. Simkins, and E. A. Edirisinghe, "Deep learning based animal detection and tracking in drone video footage," *Proceedings of the 8th International Conference on Computing and Artificial Intelligence*, 2022.

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. E. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European Conference on Computer Vision*, 2015.

[26] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask r-cnn," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 386–397, 2017.

[27] R. B. Girshick, "Fast r-cnn," *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015.

[28] A. Song, Z. Zhao, Q. X. Xiong, and J. Guo, "Lightweight the focus module in yolov5 by dilated convolution," *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)*, pp. 111–114, 2022.

[29] J. Wu, W. Hu, H. Xiong, J. Huan, V. Braverman, and Z. Zhu, "On the noisy gradient descent that generalizes as sgd," in *International Conference on Machine Learning*, 2019.

[30] H. Li, "Research on wheat ears recognition algorithm based on yolov5s neural network," 2022.