AUTOMATED DETECTION OF DOLPHINS IN IMAGERY FROM UNMANNED AERIAL VEHICLES AND PERFORMANCE OPTIMIZATION - DEEP-LEARNING IN ANIMAL ABUNDANCE SURVEYS

AUTOMATED DETECTION OF DOLPHINS IN IMAGERY FROM UNMANNED AERIAL VEHICLES AND PERFORMANCE OPTIMISATION

Deep-learning in animal abundance surveys

Eyal Bigal[a,*], Opher Bar Natan[b], Asaf Levy[b], Ori Galili[a], Massimiliano Rosso[c], Dan Tchernov[a], Tali Treibitz[c], Aviad P. Scheinin[a,b]

[a] *Morris Kahn Marine Research Station, Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa 3498838, Israel*

[b] *Department of Marine Technologies, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa 3498838, Israel*

[c] *CIMA Research Foundation, Savona 17100, Italy*

* Correspondence author: ebigal@staff.haifa.ac.il

ABSTRACT

1. Research into the integration potential of unmanned aerial vehicles with conventional abundance surveys has highlighted the importance of deep-learning applications to data processing and analysis.

2. Here, we employed the *RetinaNet* convolutional neural network for the automated detection of dolphins in imagery from unmanned aerial vehicles and assessed the optimal performance trade-off between coverage and resolution.

3. We first established an analytical limit for detection based on the minimal pixel size that is used by *RetinaNet* to generate region proposals, or 'anchor boxes.' Then, we empirically validated the effect of object size representation in the data on the accuracy of dolphin detection.

4. The minimal anchor size that is required to facilitate detection was found to be 32^2 pixels. Similarly, a high F-measure score was obtained when objects that are smaller than 32^2 pixels were excluded from the training process; dolphins occupying more than 45^2 pixels yielded recall rates above 0.8.

4. Our results demonstrate that the pixel size of both anchors and objects contribute to the optimal trade-off between coverage and resolution and influence network performance. We demonstrate the applicability of our findings to survey design in large-scale monitoring programmes as well as designated data-acquisition missions.

KEYWORDS

abundance, cetaceans, convolutional neural network, deep-learning, drones, *RetinaNet*

1. | INTRODUCTION

Unmanned aerial vehicles (UAVs; or drones) are avowed as a revolutionary tool in environmental monitoring, management and conservation (Anderson and Gaston, 2013; Gonzalez et al., 2016; Fiori et al., 2017; Jiménez López and Mulero-Pázmány, 2019). They offer a cost-effective, low-risk and less invasive alternative to conventional survey techniques as well as easier operation logistics (Christie et al., 2016). Moreover, they exhibit high spatial and temporal resolution, and the footage they produce constitutes systemic and permanent data which can later be reviewed by a high number of experts (Linchant et al., 2015; E. Bigal, *in press*). Finally, recent technological developments, such as the miniaturisation of electronic components, the improvement of payloads, the increase in computational power for central processing units (CPUs), and the development of intelligent flight control subsystems, have made UAVs more versatile and affordable in the civilian market (Colefax et al., 2018; Wang et al., 2019). Indeed, the burgeoning application of UAVs in ecological and wildlife studies demonstrates that a growing number of scientists are embracing these platforms (Christie et al., 2016). Achieved applications include the optical sampling of animal behaviour and health (e.g., Durban et al., 2015, 2016; Christiansen et al., 2016; Cheney et al., 2017; Pirotta et al., 2017; Torres et al., 2018), autonomous telemetry tracking (e.g., Selby et al., 2011; Raoult et al., 2018), and habitat monitoring and research (Crocker et al., 2011; Mancini et al., 2013; Tang and Shao, 2015).

UAVs have proven useful as an unobtrusive means of surveying species in hard-to-access environments (Chabot and Bird, 2015). This sentiment is particularly pertinent to marine animals which are notoriously difficult to study and are usually surveyed from occupied aircraft owing to the vast and often remote areas of their occurrence (Linchant et al., 2015; Wang et al., 2019). In this respect, two categories of UAV operations can be distinguished: ‘close-up’ missions, of which objective is to observe or sample animals at known locations and in

relatively small airfields, therefore typically employing rotary-wing UAVs, and ‘overflight’ missions, which use fixed-wing platforms to survey the abundance and distribution of animals across larger, pre-delineated areas (Chabot and Bird, 2015). However, considering the typically low densities of marine megafauna populations as well as their limited availability for detection, the possibilities of spotting animals in images have been assessed for a limited number of species, e.g., sea turtles and sharks (Sykora-Bodie et al., 2017; Hensel et al., 2018), and only a few studies have employed UAVs for the monitoring of wildlife across large areas (e.g., Hodgson et al., 2013, 2017; Aniceto et al., 2018). Furthermore, overflights generate substantial volumes of data as a high overlap between pictures is required to ensure that all animals on the transect are detected and counted. The analysis of such data may be time-consuming, expensive and prone to human error, and the lack of resources required for the manual reviewing of images remains a significant bottleneck to UAV integration with conventional monitoring programmes (Jiménez López and Mulero-Pázmány, 2019; Saqib et al., 2019). Therefore, there is a strong need to automate the detection, localisation and enumeration of objects in UAV data (van Gemert et al., 2015; Karnowski et al., 2016). In recent years, a variety of image analysis and machine-learning techniques have been applied to assess wildlife populations in aerial imagery (Gururatsakul et al., 2010; Mejias et al., 2013; Raoult and Gaston, 2018), and interest in deep-learning methods, particularly, has increased (Gray et al., 2019).

Unlike conventional machine-learning applications, deep-learning includes automated feature extraction, i.e., ‘self-taught’ pre-processing of raw input data, creating high-level abstractions for analytical tasks including object classification and detection (Gray et al., 2018). Among deep-learning methods, convolutional neural networks (CNNs) are most optimally-designed for image processing and analysis (Hong et al., 2019). CNNs represent semantic cues in the data while maintaining generalisation capacity, i.e., transferability to other object classes such

as different biological taxa, marine debris and vessels (Fallati et al., 2019; Lorencin et al., 2019). Their applicability to UAV-based surveys in the marine environment has been demonstrated for a variety of megafauna species including dugongs (Maire et al., 2015), sharks (Levy et al., 2018; Sharma et al., 2018), whales (Gray et al., 2019), dolphins and rays (Saqib et al., 2019), and sea turtles (Gray et al., 2018). However, only Gray et al. (2018) employed a CNN for automated object detection in a purely UAV-based, animal abundance survey. An often-cited challenge, in this respect, is that abundance estimates are based on the area of search which is directly related to the flight altitude and the resultant ground sampling distance (GSD; or image resolution). Several authors reported partial or complete loss of detections due to severe class imbalance, i.e., differing object-to-background pixel representation as image altitude increased (Levy et al., 2018; Eikelboom et al., 2019). In turn, this may result in a high number of false training samples and, thus, reduced network performance. Hence, UAVs must strike a balance between image resolution and strip width while also considering the size of the target species (Colefax et al., 2019; Wang et al., 2019).

In this study, we assessed the influence of GSD on the performance of a state-of-the-art, fully convolutional, one stage neural network, namely *RetinaNet* (Lin et al., 2017), trained for the automated detection of dolphins in imagery from UAVs. This CNN has been demonstrated to complement manual counts during animal abundance surveys and proposed to potentially outperform humans in detection from the aerial perspective (Eikelboom et al., 2019). It uses entire images as input and constructs feature maps to generate ‘anchors’, i.e., region proposals for detections. Classification and bounding box regression for each anchor is then performed to predict the presence and location of objects. Moreover, by using the *Focal Loss* function, which concentrates the training process on difficult classification instances, it is able to address the class imbalance issue (Levy et al., 2018). Thus, given the optimal trade-off between coverage and resolution, it may prove more suitable for animal abundance surveys than other

CNNs which, in a general context, were found to be preferable for the processing of UAV imagery (Hong et al., 2019). Indeed, these authors mentioned that, for certain applications, the *RetinaNet* CNN occasionally outperforms supposedly superior models. Our analysis considers the smallest object representation area required for the automated detection of dolphins and relates to two aspects of the network's training process: minimal anchor size and dataset composition. We posit that both parameters contribute to the optimal trade-off between coverage and resolution and influence network performance. Finally, we demonstrate the calculation of the maximal flight altitude per minimal object size, providing for survey design in close-up as well as overflight missions.

2. | MATERIALS AND METHODS

In 2018, the Agreement for the Conservation of Cetaceans of the Black Sea, Mediterranean Sea and contiguous Atlantic Area (ACCOBAMS) launched a region-wide monitoring effort employing conventional survey platforms – the ACCOBAMS Survey Initiative (ASI). Inspired by this programme, the current study was aimed to identify potential hindrances to the widespread adoption of UAVs by large-scale efforts of marine mammal monitoring. We, therefore, concentrated our empirical work on the three smallest species of cetaceans occurring in the Mediterranean Sea, namely striped dolphin (*Stenella coeruleoalba*), short-beaked common dolphin (*Delphinus delphis*) and common bottlenose dolphin (*Tursiops truncatus*). Due to their small size, compared to most cetaceans, those species are likely to require the highest GSD and, hence, the lowest flight altitude per any given sensor. Nevertheless, the general approach described here can also be applied to aerial surveys for smaller species likely to be observed at the surface, such as sea turtles, fish and birds. Finally, animal detectability in UAV imagery may be influenced by intrinsic (e.g., group size and diving synchrony) as well as extrinsic factors (e.g., image quality and environmental conditions). In order to demonstrate

the robustness of our work, data collection was carried out in a variety of geographical and environmental settings and included multiple UAV models, as described in the following section.

2.1. | Data collection:

We utilised shipboard platforms to perform close-up UAV flights around observed groups of dolphins. Surveys for *S. coeruleoalba* were conducted in August and September 2017 off the coast of Italy, in the northern Ligurian Sea, using a DJI Phantom 1 quadcopter (DJI Co., Shenzehn, China) with a mounted GoPro Hero3+ Black edition camera (GoPro Inc., Ca, USA) Footage of *D. delphis* and *T. truncatus* was obtained between September 2017 and May 2020, off the coast of Israel, in the eastern Levantine Sea, using four quadcopter models: DJI Mavic Pro; DJI Phantom 3 Advanced; DJI Phantom 4 Advanced; and DJI Phantom 4 Pro. The flights were carried out in a range of sea conditions and included focal follows in both video and still modes. Once launched, the UAV was positioned 100 m directly above the animals and then gradually lowered to an altitude of 20 m. Descents were paused in 10 m intervals to provide for the UAV's stability as the imagery was obtained. We implemented best practice to pilot the UAVs around marine megafauna following Hodgson and Koh (2016) and Ramos et al. (2018). A total of 1125 images containing 6075 objects, i.e., non-unique dolphins, were included in the analysis, representing 20 different encounters (*S. coeruleoalba*, n = 8; *D. Delphis*, n = 6; *T. truncatus*, n = 6). Three separate datasets were generated to train, validate and test the CNN (Table 1). Image annotation was performed on Labelbox [https://labelbox.com/], an online training data platform.

Table 1: Data partitioning according to object size.

Dataset	All Images		> 512 pixels		> 1024 pixels	
	Images	Objects	Images	Objects	Images	Objects
Training	788 (70%)	4094 (67%)	777 (70%)	4385 (71%)	717 (70%)	3488 (70%)
Validation	169 (15%)	966 (16%)	164 (15%)	850 (14%)	150 (15%)	698 (14%)
Testing	168 (15%)	1060 (17%)	168 (15%)	977 (16%)	156 (15%)	773 (16%)

2.2. | Data analysis:

During the training process, the detector generates anchors to be classified as either objects or background based on their degree of pixel overlap with manually annotated, ground truth bounding boxes. In order for a given anchor to qualify as true positive (TP), the intersect-over-union (IoU) ratio of their pixel representation areas must be greater than or equal to a pre-set threshold value. Each anchor and bounding box are available for one use only while high-IoU pairings are prioritised by the detector. Unpaired anchors and bounding boxes are considered false positives (FP) and false negatives (FN), respectively. We used the standard IoU threshold of 0.5 for training. By default, the *RetinaNet* detector employs an area of 32^2 pixels as the minimum anchor size, as smaller objects usually do not have enough details to facilitate detection. Hence, the smallest object size yielding a TP is $0.5 \cdot 32^2 = 512$ pixels.

In order to assess the potential of improving network performance by using smaller anchors, we compared the above dimensions with a configuration of 16^2 pixels, which yields a minimal detectable area of $0.5 \cdot 16^2 = 128$ pixels. The detection performance for each trained network was evaluated using the precision and recall metrics. The former is defined as the proportion of correct detections from the total number of proposals, i.e., the number of TP divided by the sum of TP and FP, whereas the latter is the proportion of correct detections from all ground truth bounding boxes, i.e., TP divided by the sum of TP and FN. In every system, there is an inherent trade-off between those two measures, where 100% recall can be gained with 0% precision and vice-versa. In our case, a higher recall would potentially translate into a greater

number of ‘small’ dolphins being detected by the network. However, this could also be obtained by adjusting the UAV’s flight or imaging parameters to increase the GSD while maintaining a higher quality of detections. Therefore, the analyses described below, for the smallest object size required to facilitate detection and its corresponding maximal flight altitude, employed the anchor dimensions yielding the highest value of precision.

In order to determine the smallest dolphin representation area, i.e., object size, that can be detected by the network, we repeated the training process a total of three times using different dataset compositions: 1. All objects, including instances that were smaller than the multiplication product of the anchor dimensions and IoU threshold; 2. Excluding those small objects; 3. Objects greater than or equal to the minimal anchor size with an IoU value of 1, i.e., full overlap. Again, the detection performance was evaluated using the trade-off between recall and precision. This trade-off is controlled by varying a single parameter, the score confidence threshold, which indicates the class prediction probability given to each bounding box. In each run, we automatically chose the threshold resulting in the highest overall accuracy according to the F-measure, as defined in Equation 1:

$$\text{F-measure} = 2 \cdot \frac{\text{recall} \cdot \text{precision}}{\text{recall} + \text{precision}}$$

For network testing, the number of epochs was 70, and we used a lower IoU threshold of 0.4, because the accurate positions of animals in the image is often irrelevant for animal abundance surveys (Eikelboom et al., 2019), and the multiplication product of this value and the minimal anchor size would enable to detect potentially smaller objects.

Finally, given the smallest detection area in pixels as well as the sensor’s specifications, the properties of perspective projection were employed to calculate the maximal flight altitude per body size of target species, $h_{max}(S)$, as described in Equation 2:

$$h_{\max}(S) = f \cdot \sqrt{\frac{S_{\text{organism}}}{p^2 \cdot S_{\min}}},$$

where f is the camera's focal length in millimetres; $S_{organism}$ is a multiplication product of the animal's width and length in metres; p is pixel pitch in millimetres; S_{min} is the minimal object size, in pixels, required by the network for detection. Notably, this equation assumes the camera sensor plane is parallel to the sea surface.

3. | RESULTS

The analysis of network performance per minimal anchor size, i.e., 32^2 and 16^2 pixels, returned a higher recall for the former yet improved precision for the latter (Figure 1). Again, for any anchor size, a high proportion of TP from the total number of proposals entails lower network sensitivity and, thus, more FP and FN. An example of the trade-off between recall and precision in our data is shown in Figure 2, where a high number of small dolphin detections was obtained using the 16^2-pixel minimal anchor size (top right) yet, in this configuration, sun glitter was also proposed as objects (middle right) and dolphins of a larger size were sometimes not detected (bottom right). Therefore, in the subsequent performance analyses, we used a minimal anchor size of 32^2 pixels.

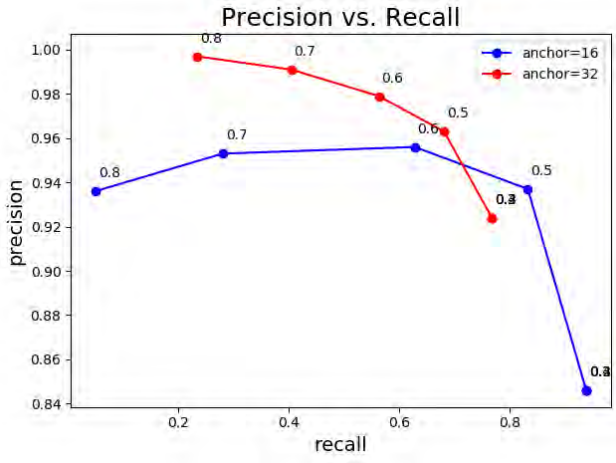

Figure 1: Receiver operating characteristic (ROC) curves of the recall and precision gained from two networks trained with a minimal region proposal area of 162 and 32 pixels (blue and red, respectively).

Given the above results, we repeated the network's training using three object size representations: 1. All objects; 2. Excluding objects smaller than 512 pixels; 3. Objects greater than or equal to 1024 pixels. Figure 3 depicts the average network performance for all object

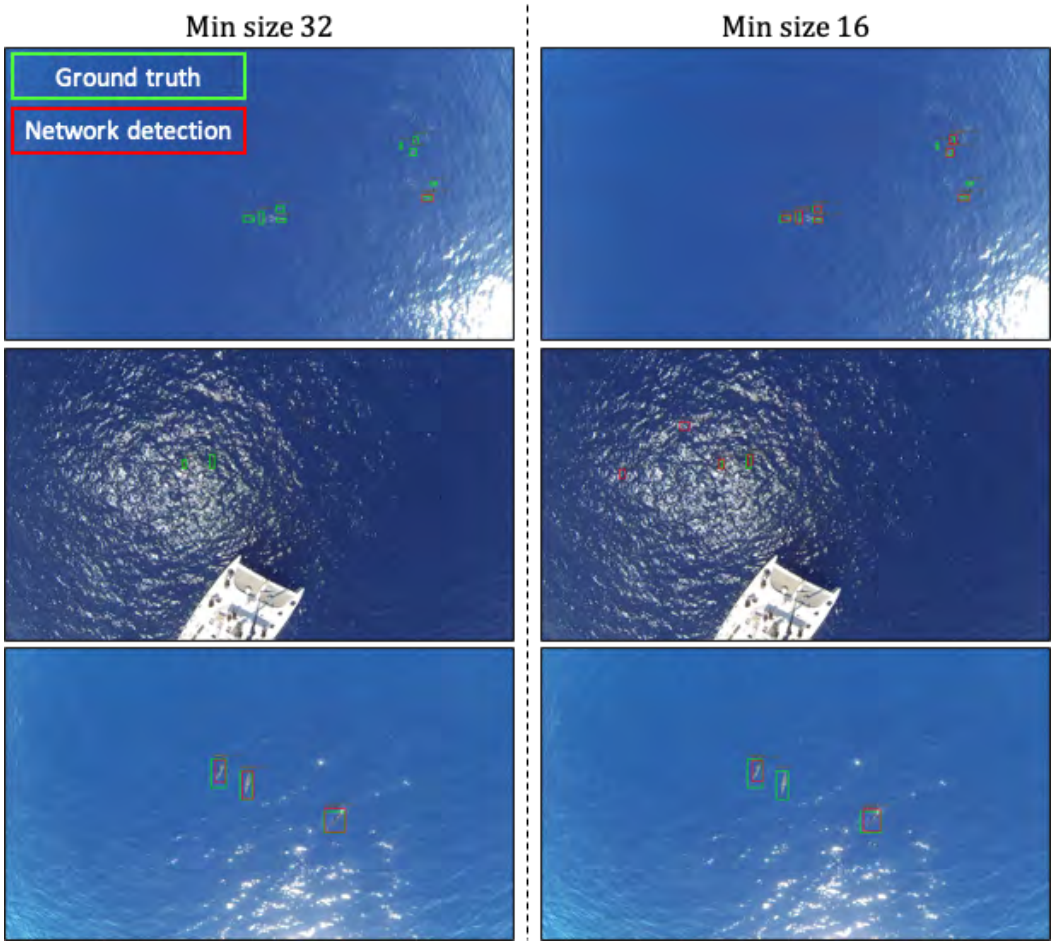

Figure 2: Compared performance of networks trained with a small (left) and standard (right) anchor size. Network output and manual ground truth annotations are represented by the red and green rectangles, respectively.

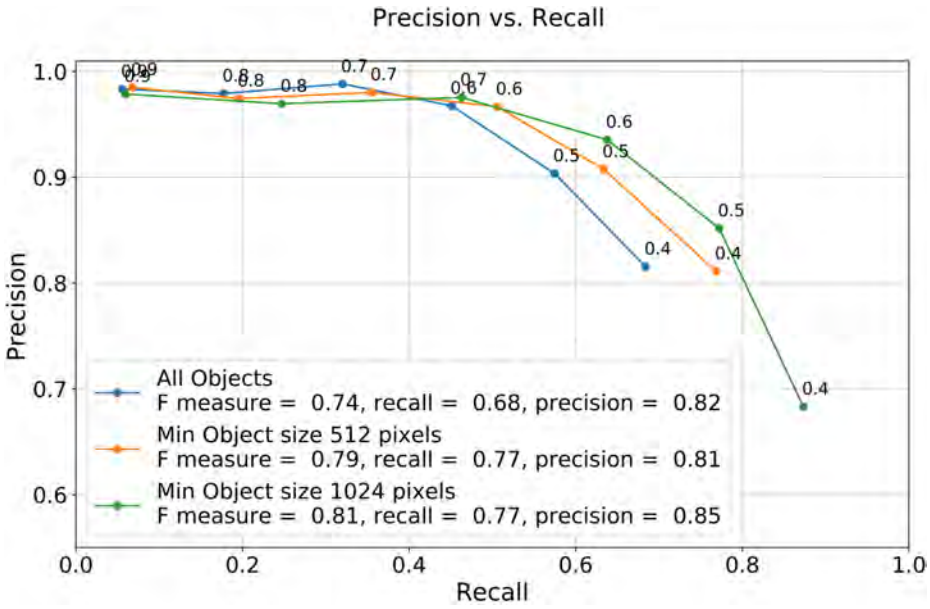

Figure 3: Effect of object size representation in the data on the maximum F-measure score, recall rate and precision. The numbers along the curve indicate the corresponding threshold values.

sizes in each of the training sets. The overall accuracy of the network, represented by the maximum F-measure, improved from 0.74 to 0.79 and 0.81 in training sets 1, 2 and 3, correspondingly; the average recall increased from 0.68 to 0.77 and 0.77, respectively, and the average precision decreased from 0.82 for the entire dataset to 0.81 when small objects were excluded, and then improved to 0.85 where only large objects were used. In order to determine the minimal representation area in pixels that is required to facilitate detection, we assessed the recall rate as a function of object size for each of the three networks (Figure 4). While the recall metric may be misleading without its corresponding precision, the latter incorporates FN of which size is of irrelevant. Therefore, we excluded it from the analysis.

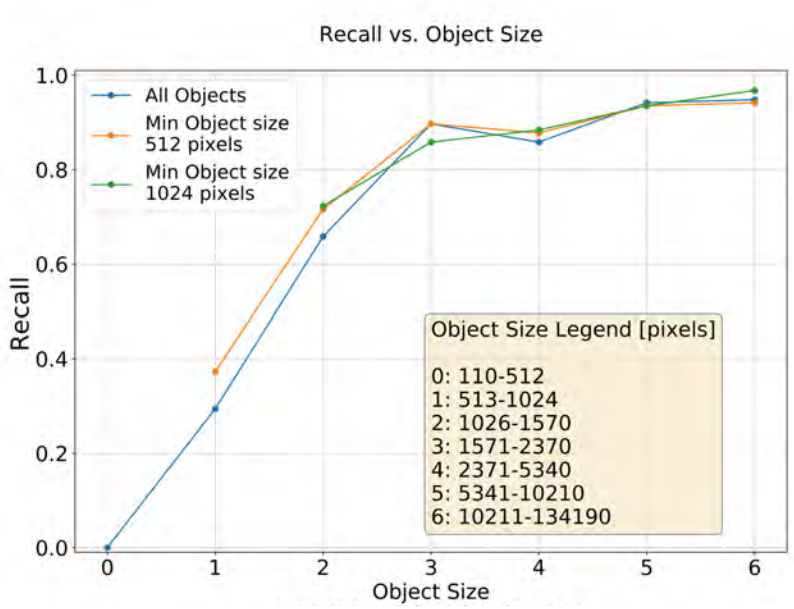

Figure 4: Recall as a function of object size.

The plot depicts a marked increase in recall rates as object size increases up to 1000 pixels. Objects of approximately 2000 pixels returned recall rates of 85-95% with a more moderate slope for instances of larger sizes. Note, that the three networks exhibited similar curves, yet, as mentioned with regard to Figure 3, precision was higher where small objects were excluded. Figure 5 depicts two examples of differences in overall accuracy between the three trained networks. In the first example (a-c), where objects smaller than 512 pixels were included in the training set (a), a higher number of FN is observed compared to the two other networks (b, c); in the second example (d-f), high performance is obtained by only including objects larger than 1024 pixels (f).

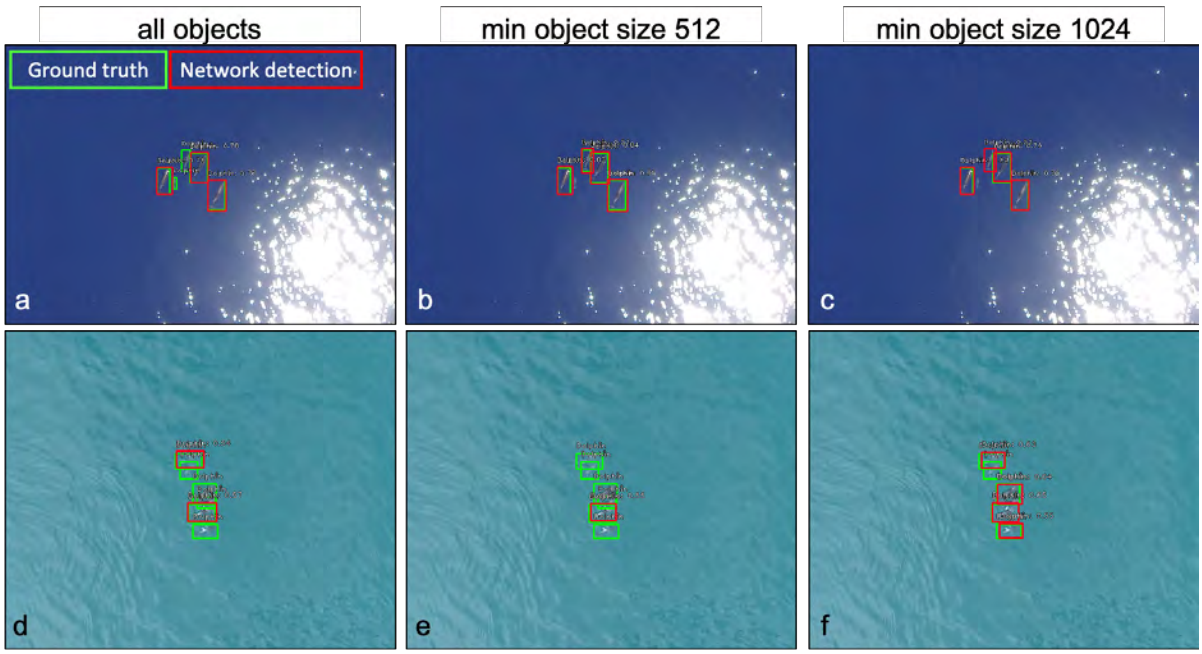

Figure 5: Effect of object size representation in the data (*left*, small; *centre*, medium; *right*, large) on detection performance. Two different examples are provided (*top*, a-c; *bottom*, d-f). Network output and manual ground truth annotations are represented by the red and green rectangles, respectively.

Finally, concerning data acquisition, the minimum object size of 1024 pixels may be used to calculate the maximal flight altitude per body size of target species. For example, during our surveys of *S. coeruleoalba*, we used a GoPro Hero3+ Black edition camera (2.1). Given the camera specifications as well as a focal length of 2.9 mm and pixel pitch of 0.0015 mm, a generic dolphin of 2.5 m in length and 1.5 m in width would be detected from an altitude of up to 113 m, based on Equation 2. As regards *D. delphis* and *T. truncatus*, the primary UAV employed was a DJI Phantom 4 Pro quadcopter and the calculated focal length and pixel pitch were 7.5 mm and 0.0088 mm, respectively, yielding a maximal flight altitude of 51.5 m for the above dolphin dimensions.

4. | DISCUSSION

In this study, we employed the *RetinaNet* CNN for the detection of dolphins in imagery from UAVs. We assessed the potential of optimising survey design in aerial abundance surveys by establishing the optimal performance trade-off between coverage and resolution.

First, we examined the effect of utilising the detector's standard dimensions of minimal region proposals, or 'anchors,' compared to a smaller size configuration, i.e., 32^2 and 16^2 pixels, respectively. Higher precision, which we considered the most relevant metric (2.2), was obtained by the standard anchor dimensions. In the second stage, we assessed the effect of object size representation in the data on the overall accuracy of detections. High average performance was obtained when all objects that are smaller than 32^2 pixels were excluded from the training process, which improved the F-measure score from 0.74 to 0.79 and 0.8 in networks 1, 2 and 3, respectively. Concerning the direct effect of object size on performance, dolphins occupying less than 22.5^2 pixels were not detected, whereas those larger than 45^2 pixels yielded recall rates above 0.8. For reference, Eikelboom et al. (2019) used the *RetinaNet* CNN to detect elephants, giraffes and zebras in aerial images from a semi-automated, animal abundance survey, and reported F-measure scores of 0.76, 0.78 and 071, respectively, for larger object sizes, of 50^2 pixels.

Considering the lack of publicly-available, annotated aerial imagery for the study of marine species (Saqib et al., 2019), ecologists are advised to avoid the collection of unusable data during animal encounters which, in the case of some cetaceans, may be particularly rare. Hence, species-specific protocols referring to the flight and imaging parameters should be developed based on the required accuracy and transect width (Wang et al., 2019), and not vice-versa. Indeed, our results demonstrate the effect of object size representation in the data on the CNN's detection performance. This example may be particularly useful to studies that rely on pre-existing training sets, where animals are photographed in a variety of GSDs. In cases where

the *RetinaNet* detector's accuracy is used to compare between different counting methods or alternative CNNs, and imagery has been collected from a constant GSD, the minimal anchor size can also be adjusted to meet recall or precision requirements. Notably, such surveys may not always generate sufficient volumes of data to ensure model robustness. Therefore, training set augmentation methods are usually used as a pre-processing method as well as complimentary, species-specific close-up flights (e.g., Hong et al., 2019). Thus, the results of this study apply to operations employing both rotary- and fixed-wing UAVs.

Finally, our study emphasises the importance of uploading annotated aerial imagery for future studies employing CNNs. Data availability may be particularly valuable with regard to low-density populations of highly-mobile species in the offshore marine environment, such as small cetaceans. To that end, we make our image data and annotations available for public use.

We suggest that future research into the potential of UAV integration with conventional cetacean surveys, and the use of CNNs for automated object detection, should be focused on inference data obtained during large-scale, animal abundance surveys employing fixed-wing UAVs. However, training sets should be pre-established and represent a range of object sizes.

ACKNOWLEDGEMENTS

This study was funded by the Agreement for the Conservation of Cetaceans of the Black Sea, Mediterranean Sea and contiguous Atlantic Area (ACCOBAMS). We are grateful to Mr Dror Vardimon and Mr Guy Lavian for their contribution to the collection of data in the field as well as the image annotators.

AUTHORS' CONTRIBUTIONS

E.B., A.P.S. and T.T. conceived the ideas and designed the methodology; E.B. coordinated the project; A.S.C. and D.T. supervised the project; E.B., A.P.S., O.G., and M.R. collected the

data; T.T., O.B., and A.L. performed the analyses; E.B. led the writing of the manuscript. All authors contributed critically to this work and gave final approval for publication.

DATA AVAILABILITY STATEMENT

The annotated image data is publicly available at the University of Haifa Data Science Research Center.

REFERENCES

Anderson, K., Gaston, K.J., 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Front. Ecol. Environ. 11, 138–146. doi:10.1890/120150

Aniceto, A.S., Biuw, M., Lindstrøm, U., Solbø, S.A., Broms, F., Carroll, J., 2018. Monitoring marine mammals using unmanned aerial vehicles: Quantifying detection certainty. Ecosphere 9. doi:10.1002/ecs2.2122

Chabot, D., Bird, D.M., 2015. Wildlife research and management methods in the 21st century: Where do unmanned aircraft fit in? J. Unmanned Veh. Syst. 3, 137–155. doi:10.1139/juvs-2015-0021

Cheney, B., Wells, R.S., Barton, T.R., Thompson, P.M., 2017. Laser photogrammetry reveals variation in growth and early survival in free-ranging bottlenose dolphins. Anim. Conserv. 1–10. doi:10.1111/acv.12384

Christiansen, F., Dujon, A.M., Sprogis, K.R., Arnould, J.P.Y., Bejder, L., 2016. Noninvasive unmanned aerial vehicle provides estimates of the energetic cost of reproduction in humpback whales 7.

Christie, K.S., Gilbert, S.L., Brown, C.L., Hatfield, M., Hanson, L., 2016. Unmanned aircraft systems in wildlife research: Current and future applications of a transformative technology. Front. Ecol. Environ. 14, 241–251. doi:10.1002/fee.1281

Colefax, A.P., Butcher, P.A., Kelaher, B.P., 2018. The potential for unmanned aerial vehicles

(UAVs) to conduct marine fauna surveys in place of manned aircraft. ICES J. Mar. Sci. 75, 1–8. doi:10.1093/icesjms/fsx100

Colefax, A.P., Butcher, P.A., Pagendam, D.E., Kelaher, B.P., 2019. Reliability of marine faunal detections in drone-based monitoring. Ocean Coast. Manag. 174, 108–115. doi:10.1016/j.ocecoaman.2019.03.008

Crocker, R.I., Maslanik, J.A., Adler, J.J., Palo, S.E., Herzfeld, U.C., Emery, W.J., 2011. A sensor package for ice surface observations using small unmanned aircraft systems. IEEE Trans. Geosci. Remote Sens. 50, 1033–1047.

Durban, J.W., Fearnbach, H., Perryman, W.L., Leroi, D.J., 2015. Photogrammetry of killer whales using a small hexacopter launched at sea. J. Unmanned Veh. Syst. 3, 1–5. doi:dx.doi.org/10.1139/juvs-2015-0020

Durban, J.W., Moore, M.J., Chiang, G., Hickmott, L.S., Bocconcelli, A., Howes, G., Bahamonde, P.A., Perryman, W.L., LeRoi, D.J., 2016. Photogrammetry of blue whales with an unmanned hexacopter. Mar. Mammal Sci. 32, 1510–1515. doi:10.1111/mms.12328

Eikelboom, J.A.J., Wind, J., van de Ven, E., Kenana, L.M., Schroder, B., de Knegt, H.J., van Langevelde, F., Prins, H.H.T., 2019. Improving the precision and accuracy of animal population estimates with aerial image object detection. Methods Ecol. Evol. 10, 1875–1887. doi:10.1111/2041-210X.13277

Fallati, L., Polidori, A., Salvatore, C., Saponari, L., Savini, A., Galli, P., 2019. Anthropogenic Marine Debris assessment with Unmanned Aerial Vehicle imagery and deep learning: A case study along the beaches of the Republic of Maldives. Sci. Total Environ. 693, 133581. doi:10.1016/j.scitotenv.2019.133581

Fiori, L., Doshi, A., Martinez, E., Orams, M.B., Bollard-Breen, B., 2017. The use of unmanned aerial systems in marine mammal research. Remote Sens. 9, 11–17.

doi:10.3390/rs9060543

Gonzalez, L.F., Montes, G.A., Puig, E., Johnson, S., Mengersen, K., Gaston, K.J., 2016. Unmanned aerial vehicles (UAVs) and artificial intelligence revolutionizing wildlife monitoring and conservation. Sensors (Switzerland) 16. doi:10.3390/s16010097

Gray, P.C., Bierlich, K.C., Mantell, S.A., Friedlaender, A.S., Goldbogen, J.A., Johnston, D.W., 2019. Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry. Methods Ecol. Evol. 10, 1490–1500. doi:10.1111/2041-210X.13246

Gray, P.C., Fleishman, A.B., Klein, D.J., McKown, M.W., Bézy, V.S., Lohmann, K.J., Johnston, D.W., 2018. A convolutional neural network for detecting sea turtles in drone imagery. Methods Ecol. Evol. 10, 345–355. doi:10.1111/2041-210X.13132

Gururatsakul, S., Gibbins, D., Kearney, D., Lee, I., 2010. Shark detection using optical image data from a mobile aerial platform. Int. Conf. Image Vis. Comput. New Zeal. doi:10.1109/IVCNZ.2010.6148828

Hensel, E., Wenclawski, S., Layman, C.A., 2018. Using a small, consumer-grade drone to identify and count marine megafauna in shallow habitats. Lat. Am. J. Aquat. Res. 46, 1025–1033. doi:10.3856/vol46-issue5-fulltext-15

Hodgson, A., Kelly, N., Peel, D., 2013. Unmanned aerial vehicles (UAVs) for surveying Marine Fauna: A dugong case study. PLoS One 8, 1–15. doi:10.1371/journal.pone.0079556

Hodgson, A., Peel, D., Kelly, N., 2017. Unmanned aerial vehicles for surveying marine fauna: Assessing detection probability. Ecol. Appl. 27, 1253–1267. doi:10.1002/eap.1519

Hodgson, J.C., Koh, L.P., 2016. Best practice for minimising unmanned aerial vehicle disturbance to wildlife in biological field research. Curr. Biol. 26, R404–R405.

doi:10.1016/j.cub.2016.04.001

Hong, S.J., Han, Y., Kim, S.Y., Lee, A.Y., Kim, G., 2019. Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. Sensors (Switzerland) 19, 1–16. doi:10.3390/s19071651

Jiménez López, J., Mulero-Pázmány, M., 2019. Drones for Conservation in Protected Areas: Present and Future. Drones 3, 10. doi:10.3390/drones3010010

Karnowski, J., Johnson, C., Hutchins, E., 2016. Automated Video Surveillance for the Study of Marine Mammal Behavior and Cognition. Anim. Behav. Cogn. 3, 255–264. doi:10.12966/abc.05.11.2016

Levy, D., Belfer, Y., Osherov, E., Bigal, E., Scheinin, A.P., Nativ, H., Tchernov, D., Treibitz, T., King, A., Bhandarkar, S.M., 2018. Automated Analysis of Marine Video With Limited Data, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 1385–1393.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision. pp. 2980–2988.

Linchant, J., Lisein, J., Semeki, J., Lejeune, P., Vermeulen, C., 2015. Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges. Mamm. Rev. 45, 239–252. doi:10.1111/mam.12046

Lorencin, I., Anđelić, N., Mrzljak, V., Car, Z., 2019. Marine objects recognition using convolutional neural networks. Nase More 66, 112–119. doi:10.17818/NM/2019/3.3

Maire, F., Alvarez, L.M., Hodgson, A., 2015. Automating marine mammal detection in aerial images captured during wildlife surveys: a deep learning approach, in: Australasian Joint Conference on Artificial Intelligence. Springer, pp. 379–385.

Mancini, F., Dubbini, M., Gattelli, M., Stecchi, F., Fabbri, S., Gabbianelli, G., 2013. Using

unmanned aerial vehicles (UAV) for high-resolution reconstruction of topography: The structure from motion approach on coastal environments. Remote Sens. 5, 6880–6898.

Mejias, L., Duclos, G., Hodgson, A., Maire, F., 2013. Automated marine mammal detection from aerial imagery. Ocean. 2013 MTS/IEEE - San Diego An Ocean Common.

Pirotta, V., Smith, A., Ostrowski, M., Russell, D., Jonsen, I.D., Grech, A., Harcourt, R., 2017. An economical Custom-Built drone for assessing whale health. Front. Mar. Sci. 4, 1–12. doi:10.3389/fmars.2017.00425

Ramos, E.A., Maloney, B., Magnasco, M.O., Reiss, D., 2018. Bottlenose dolphins and Antillean manatees respond to small multi-rotor unmanned aerial systems. Front. Mar. Sci. 5, 1–15. doi:10.3389/fmars.2018.00316

Raoult, V., Gaston, T.F., 2018. Rapid biomass and size-frequency estimates of edible jellyfish populations using drones. Fish. Res. 207, 160–164. doi:10.1016/j.fishres.2018.06.010

Raoult, V., Tosetto, L., Williamson, J., 2018. Drone-Based High-Resolution Tracking of Aquatic Vertebrates. Drones 2, 37. doi:10.3390/drones2040037

Saqib, M., Daud Khan, S., Sharma, N., Scully-Power, P., Butcher, P., Colefax, A., Blumenstein, M., 2019. Real-Time Drone Surveillance and Population Estimation of Marine Animals from Aerial Imagery. Int. Conf. Image Vis. Comput. New Zeal. 2018-Novem, 1–6. doi:10.1109/IVCNZ.2018.8634661

Selby, W., Corke, P., Rus, D., 2011. Autonomous aerial navigation and tracking of marine animals. Proc. 2011 Australas. Conf. Robot. Autom. 7–9.

Sharma, N., Scully-Power, P., Blumenstein, M., 2018. Shark Detection from Aerial Imagery Using Region-Based CNN, a Study, in: Australasian Joint Conference on Artificial Intelligence. Springer, pp. 224–236.

Sykora-Bodie, S.T., Bezy, V., Johnston, D.W., Newton, E., Lohmann, K.J., 2017.

Quantifying nearshore sea turtle densities: applications of unmanned aerial systems for population assessments. Sci. Rep. 7, 17690.

Tang, L., Shao, G., 2015. Drone remote sensing for forestry research and practices. J. For. Res. 26, 791–797.

Torres, L.G., Nieukirk, S.L., Lemos, L., Chandler, T.E., 2018. Drone up! Quantifying whale behavior from a new perspective improves observational capacity. Front. Mar. Sci. 5, 1–14. doi:10.3389/fmars.2018.00319

van Gemert, J.C., Verschoor, C.R., Mettes, P., Epema, K., Koh, L.P., Wich, S., 2015. Nature Conservation Drones for Automatic Localization and Counting of Animals, in: Agapito, L., Bronstein, M.M., Rother, C. (Eds.), Computer Vision - ECCV 2014 Workshops. Springer International Publishing, Cham, pp. 255–270.

Wang, D., Shao, Q., Yue, H., 2019. Surveying wild animals from satellites, manned aircraft and unmanned aerial systems (UASs): A review. Remote Sens. 11. doi:10.3390/rs11111308