European Journal of Computer Science and Information Technology
Vol.8, No.5, pp.35-45, October 2020
Published by ECRTD- UK
Print ISSN: 2054-0957 (Print), Online ISSN: 2054-0965 (Online)

USING DEEP LEARNING CONVOLUTIONAL NEURAL SYSTEM FOR DISCRETE MINKE WHALE APPRECIATION

Aissa Snani[1], Xiaofeng Zhou[1],Yingchi Mao[1],Zunayed al Mamoonr[2]
[1]College of Computer and Information, Hohai University, China
[2]Department of Computer Science, University of Asia Pacific, Bangladesh

ABSTRACT: *The only known predictable aggregation of dwarf minke whales .occurs in the Australian offshore waters of the northern Great Barrier Reef in May-August each year. The identification of individual whales is required for research on the whales' population characteristics and for monitoring the potential impacts of tourism activities, including commercial swims with the whales. At present, it is not cost-effective for researchers to manually process and analyze the tens of thousands of underwater images collated after each observation/tourist season, and a large data base of historical non-identified imagery exists. This study reports the first proof of concept for recognizing individual dwarf minke whales using the Deep Learning Convolutional Neural Networks (CNN).The "off-the-shelf" Image net-trained VGG16 CNN was used as the feature-encoder of the per-pixel sematic segmentation Automatic Minke Whale Recognizer (AMWR). The most frequently photographed whale in a sample of 76 individual whales (MW1020) was identified in 179 images out of the total 1320 images provided. Training and image augmentation procedures were developed to compensate for the small number of available images. The trained AMWR achieved 93% prediction accuracy on the testing subset of 36 positive/MW1020 and 228 negative/not-MW1020 images, where each negative image contained at least one of the other 75 whales. Furthermore on the test subset, AMWR achieved 74% precision, 80% recall, and 4% false-positive rate, making the presented approach comparable or better to other state-of-the-art individual animal recognition results.*

KEYWORDS*:* dwarf minke whales, photo-identification, population biology, convolutional neural networks, deep learning, image recognition

INTRODUCTION

The dwarf minke whale (*Balaenoptera acutorostrata* subsp.) is the second smallest baleen whale, born at approximately 2m in length and growing to a maxi-mum measured length of 7.8 m [1]. Dwarf minke whales are distributed throughout the southern hemisphere, including Antarctica, and were first acknowledged as a distinct form of minke in 1985 [1]. The only known predictable aggregation of dwarf minke whales occurs in the Australian offshore waters of the northern Great Barrier Reef (GBR) each year throughout the Australian winter months

[3]. This aggregation supports a local swim-with-whales tourism industry [2] [3]. The predictable nature of this aggregation has also enabled dedicated research of dwarf minke whales, which has contributed to seminal work on dwarf minke whale biology [4], behavior [5], and assessment and management of swim-with-whales activities [2]. Outputs from this work have in-formed and shaped management policies and expanded knowledge of both the subspecies in general and, specifically, the interactions with the tourism industry. The uniqueness of this aggregation presents an opportunity to conduct re-search and improve the knowledge base for a poorly understood oceanic coequal whale, as well as a responsibility to ensure that tourism activities are managed sustainably [2] [3] [5].

The identification of individual whales underpins much of the scientific re-search on dwarf minke whales and the monitoring of tourism activities. While in the GBR, these whales are highly inquisitive, readily approaching vessels and divers and often maintaining contact for prolonged periods [3] [5]. This behaviour provides good opportunities for passengers aboard the swim-with tourism vessels to photograph dwarf minke whales. The whales' color patterns have been shown to remain stable over many years, and are sufficiently complex to allow for unequivocal identification of individuals [3] [6] [7]. The stability of these patterns and the regular, in-water access provided to researchers by tourism vessels has made the dwarf minke whale an ideal species for photo-identification (photo-ID) [6] [8].

Photo-ID is a simple, non-invasive technique widely used to study a range of biological and behavioral characteristics of wild animal populations. Ideal candidates for photo-ID are those with stable color patterns and/or other markings that are unique to each individual, so that individuals can be easily distinguished from each other and their identifiable markings remain the same over time. The automation of the photo-ID process is often highly specific to the required species, e.g. fin contour of great white sharks [9]. Due to its fundamental research role, photo-ID is an active research area for many species, e.g. green sea turtles [10], gorillas [11], and dolphins [12]. For minke whales, photo-ID has typically involved visual comparison of large numbers of photographs by trained researchers; thus, the process is time-intensive. Much of the imagery used for photo-identification of dwarf minke whales in re-cent years has come from tourists and crew aboard swim-with whales dive tour-ism vessels [8]. The quantity of this donated imagery has increased dramatically with the availability of low-cost digital underwater cameras and the resultant rise in popularity of these items among tourists [8]. Researchers are now obtaining tens of thousands of photographs and video clips each season. Consequently, it is no longer cost-effective for researchers to manually process and analyse

such quantities of images, and a large database of historical non-identified imagery exists. In order to utilize the increasing quantity of imagery to address key biological and ecological knowledge gaps about these whales, automatic computer-vision based recognition software is required, and was the main focus of this study.

Over the last few years the Deep Learning Convolutional Neural Networks (CNNs) revolutionized the field of computer-vision image recognition [13]. For example, the Alex Net image classification CNN [14] won the Image net Large Scale Visual Recognition Challenge (ILSVRC) [15] in 2012, and since then all the ILSVRC13-ILSVRC17 winners used CNNs of various architectural configurations as their key features, e.g. [16]. It is customary to refer to such CNNs as been *trained-on Image net*.

A typical Imagenet-trained CNN is setup to classify as many as 1000 different types of objects. Therefore, it is plausible to expect that such a CNN could distinguish at least 1000 different individual dwarf minke whales if it is trained or re-trained appropriately. This direct approach, however, has a number of limit-ing factors. First, millions of images are available in the Imagenet for training CNNs, which is presently not feasible for dwarf minke whales, where the number of images available for an individual whale may vary between one and sever-al thousand. Second, typical Imagenet object categories are very different, e.g. differences in images for dogs and people, whereas all minke whales fit essentially the same category for the Imagenet (*i.e.* near-identical body shape, proportions and general color). Third, the output of a classification CNN is a single probability number for each available class, where category and class are used as equivalent terms in this study. Such probability prediction has limited value to a marine biologist, as it does not explain why/how CNN arrived at its prediction. This is known as the *black-box* perception and/or criticism of the classification CNNs. The black-box CNN prediction is unavoidable in studies where animals are identified by their "faces", e.g. for gorillas [11], and identification uses facial geometrical proportions and is essentially the full face. Fortunately in the case of dwarf minke whales, they are currently identified by finely detailed color pat-terns and scars (**Figure 1**), which could be recognized and localized by CNN, and then confirmed by a trained researcher.

The *black-box* limitation of the classification CNNs has a natural solution

Figure 1. Example of individual minke whale distinct fin color pattern and scars.

when the CNNs are configured to perform semantic segmentation of images, where an image is segmented into per-pixel categories [17]. The output of segmentation CNNs is a per-pixel *heat-map* (also known as the probability or activation map) for each class. Therefore, a researcher could easily verify the CNN prediction by viewing the heat-map corresponding to the recognized individual whale (**Figure 2**). This approach was successfully validated in this proof of con-cept study by training a segmentation CNN to recognize a single whale within 1320 images of 76 different whales.

MATERIALS AND METHODS

Dataset

The underwater imagery dataset used in this study consisted of 1320 digital photographs of dwarf minke whales (*Balaenoptera acutorostrata* subsp.). All images were sorted according to unique individual animals. In some cases only left or right sides of a whale was identified, without knowing if corresponding images belonged to the same whale or not. Where it was possible to match the left and right sides to the same whale, the related imagery was labelled accordingly and placed together in the same folder. As a result, the dataset identified 76 different whales. The identification process was extremely time consuming even for trained researchers as it required recording and cataloguing the color patterns and scars of 76 different whales, and/or reviewing any new image against at least 76 other whale images thus relying on researchers' memory to identify matches with any efficiency. The number of available images varied greatly between individuals; the MW1020 individual had the largest number of images (179), and several whales had only one image per individual.

Segmentation Neural Network

As described in the introduction, this study used a segmentation CNN rather than a classification CNN to recognize an individual minke whale and localize the recognized unique features. Specifically, the most accurate segmentation FCN-8s model from the Fully Convolutional Networks (FCN) [17] was selected due to the following considerations. First, the FCN-8s model is based on the VGG16 CNN model [16], which was one of the top performers in the ILSVRC14 [15].

- Randomly rotated in the range of [−45, +45] degrees, where the input image was reflected to fill pixels outside the original boundary as required;
- Randomly resized in the scale range of [0.75,1.25] , or by up to 25% zooming in or out;
- Randomly shifted in each color channel in the [−25.5, 25.5] range, where 25.5 was the 10% of maximum colour values 255;
- Randomly gamma shifted in the [−25.5, 25.5] range, where all color chan-nels values were shifted together;
- Randomly cropped to retain 480 : 480 pixels;
- Image net color mean values were subtracted as commonly done when work-ing with the Image net-trained VGG16 model.

The following training workflow was adopted for this study. All available im-ages were sequentially numbered and split into five approximately equal subsets. The first three subsets were used as a single *training* set, *i.e.* 60% of all available images. The fourth and the fifth subsets became the *validation* and *testing* sets, respectively. More precisely, the i^{th} image was allocated to *validation* or *test* if (i >1) or i were multiple of 5, respectively, where all remaining images were as-signed to the *training* set.

The training of FCN-8s was done in up to 100 cycles. In each cycle, TAP480 was further applied to the already ISP640-processed images. The training images were loaded into memory as a X (N_t , M , M , C) tensor or a multidimensional matrix, where N_t >200 was the number of images, M >480 was the TAP480 cropping length, and where C > 3 was due to the three available colour channels. The corresponding to the loaded training images were the *ground-truth* binary per-pixel masks, which were loaded as a *one-hot* encoded Y (N_t , M , M , K) tensor, where Y >i , m, l , k > 9 1 if the (m, l) pixel belonged to the kth class in the ith image and zero otherwise. The required number of classes K was K : 1 for the automatic whale locator

and a single whale classifier, as described later on in this paper. The validation X_v and Y_v tensors were constructed in similar fashion.

The per-pixel binary cross-entropy loss function, e.g. p.231 of [13], was aver-aged as required and used as the training loss metric. Due to the available Graphical Processing Unit (GPU) memory limits, training was done in batches of only four images. Up to 16 training epochs were allowed per cycle, where one feed-forward and one back-propagation passes through all N_t-loaded im-age-mask pairs were considered to be one epoch. Training for a given cycle was aborted if the validation loss metric did not decrease after two epochs, this is commonly known as *early stopping*. Note that the early stopping was the only place where the validation images were used in training. In order to prevent the indirect overfitting of the validation images, they were augmented by TAP480 before each training cycle similar to the training set.

Minke Whale Locator

Being a segmentation model, the FCN-8s model required the ground-truth per-pixel binary mask for each of the training and validation images. Therefore, the auxiliary goal of this study was to design the required workflow to be as scalable as possible for future larger training datasets. Creating the ground-truth per-pixel binary masks was clearly the least scalable component of this study, and required a scalable solution. This was solved by training an instance of FCN-8s to be the Minke Whale Locator (MWL).

To train MWL, 100 images were segmented by hand (including 50 of the MW1020 individual) to produce binary per-pixel ground-truth mask Y for each of the 100 images. Then MWL was trained as per preceding Section 2.2 with the following modifications. In addition to TAP480, images were flipped horizon-tally with 0.5 probability. The available 100 images were split 70 for training, and 30 for validation, where the rest of the not-segmented images were considered to be the testing set. The Keras version of the RMS prop optimizer was used with 10^{-4} learning rate, and 10^{-3} learning rate decay after each weights update, where RMS prop "*divides the learning rate for a weight by a running average of the magnitudes of recent gradients for that weight*" [19]. Once the per-pixel valida-tion accuracy stopped improving (usually at around 95%), the Stochastic Gra-dient Descent (SGD) optimizer was used with 10^{-4} learning rate, 10^{-3} learning rate decay, 0.9 momentum, and enabled Nesterov momentum.

Trained MWL was applied to all available images to automatically generate one largest *rectangular* binary mask per ISP640 pre-processed image. Note that since MWL was fully convolutional, it was rebuilt to accommodate any required image dimensions, where one side was always 640 (due to ISP640) but the other side was varied. The mask generation was done as follows. For each image, the

$$\text{per-pixel prediction heat-map } Y_p(i,j) \quad \text{was converted to binary mask } B \text{ via,}$$
$$B(i,j) = 1, \; Y_p(i,j) \geq 0.8, \qquad (1)$$

where i and j were the row and column pixel location indices, respectively, and where the remaining mask values were set to zero, *i.e.* $B(i,j) > 0, Y_p : i, j > 9.8$. The largest connected non-zero area was filled to complete its minimum-enclosing rectangle, and saved as the only non-zero values of the final binary mask.

Automatic Minke Whale Recognition

Similar to the preceding MWL model, an instance of the FCN-8s model was created for a required number of K individual whales to be the Automatic Minke Whale Recognition (AMWR) model. To train AMWR, the automatically created (by MWL) masks for the K whales were reviewed for correctness. Specifically, each MWL-generated rectangular mask was checked to make sure it enclosed correct whale if multiple whales were present in an image. Also, if the mask did not enclose the whole whale, the mask was verified to enclose all whales' fea-tures, which a biologist could use to identify that whale, *i.e.* fin coloration pat-terns and distinct scars. Note that in this study, the MWL model was nothing more than a convenience tool to automate ground-truth mask creation. Therefore where available, the manually segmented masks were used instead of the corresponding MWL masks.MWL produced acceptable bounding boxes in more than 90% cases confirming it to be a viable tool for this project.

The AMWR was trained as per preceding Section 2.2 with the following mod-ifications. For the K selected whales the *positive* ground-truth masks (manually or automatically MWL-segmented) were used. The training masks for the re-maining $(76 - K)$ whales were automatically generated as *negative* or all-zeros masks, *i.e.* any of the K selected whales were missing in the remaining images. Then the training proceeded as per MWL but with added regularization weight decay set to 10^{-4}.

RESULTS AND DISCUSSION

The largest number (179) of images was available for the individual whale MW1020 so it was used as the benchmark of possible accuracy for the utilized dataset and the AMWR model with K □ 1. As per preceding Sections 2.3 and 2.4, 50 masks were segmented manually, and the rest of available MW1020 im-ages (129) were segmented by MWL and quality-checked visually. The MW1020 training, validation and test sets contained 107, 36, and 36 images, respectively. The rest of other whale images (1141) were automatically labeled as *negative*, and split 60%-training, 20%-validation, and 20%-test. Because there were many more negative labels than positive, for each training cycle an equal number of images (100) were randomly selected from both negative and positive/MW1020 training images. Similarly, all available 36 MW1020 validation images were used with 36 randomly selected negative validation images, where a new random se-lection of 36 negative images was done before each training cycle. Also due to the highly unbalanced number of positive and negative examples, AMWR classifier was assessed via *precision*, *recall*, *fprate* (false-positive), in addition to the standard *accuracy* [9] [11],

$$precision = TP / (TP + FP), \quad recall = TP / P, \quad fp\ rate = FP / N \tag{2}$$

$$accuracy = (TP + TN) / (P + N), \tag{3}$$

where *TP*, *TN*, *FP* and *FN* were the numbers of true-positive, true-negative, false-positive and false-negative predictions, respectively, and where *P* and *N* were the total numbers of positive (MW1020) and negative (non-MW1020-whale) images.

The main distinct advantage of a *per-pixel* classifier (rather than *per-image*) such as the presented AMWR, is the full control over how "conservative" or "liberal" [12] it could be configured. The highly conservative version was configured by accepting the prediction heat-map values only above 0.99, where the binary per-pixel predictions were set as $B(i, j) < 1$, $Y_p(i, j) \geq 0.99$ and zero otherwise. Furthermore, the largest connected prediction area was only accepted as a positive detection if its area was at least 64 = 64 > 4096 pixels, see example in **Figure 2**.

Figure 2. Example of AMWR per-pixel prediction for MW1020 individual. The pixels with the prediction heat-map values above 0.99 were illustrated by amplifying the cor-responding image pixel intensities by factor of 1.5.

Table 1. Identification results for MW1020.

Prediction Score	**Datasets**		
	Train	Validation	Test
Accuracy	0.984	0.924	0.935
Precision	0.935	0.735	0.743
Recall	0.953	0.694	0.805
Fp rate	0.01	0.04	0.04

On the *test* subset, AMWR achieved 4% false-positive rate (**Table 1**). Low *fp rate* was viewed as essential to support a workflow where many thousands of unsorted images could be scanned for the known whales, and the number of "false-alarm" instances would remain feasible to be classified manually. AMWR's *test* precision (74%) and recall (80%) results (last column of **Table 1**) were better than the corresponding state-of-the-art gorilla identification results [11] of ap-proximately 60%. The AMWR's test accuracy (93%) and precision (74%) were comparable to the 81% average precision achieved in the state-of-the-art great white shark identification results [9]. The *validation* and *test* prediction metrics were comparable (third and fourth columns in **Table 1**) supporting the achieved testvalues to be the expected benchmark/baseline values of the AMWR model in future similar circumstances/studies.

CONCLUSION

Due to the increasing abundance of underwater digital imagery, the manual identification of individual dwarf minke whales from images and videos has be-come cost-ineffective. It has become excessively time-consuming to manually check if an unsorted image contains a new whale or a known whale, e.g. from the 76 labeled whales of this study's dataset. Considering that photo-identification of dwarf minke whales represents one of the few methods available to address key knowledge gaps for this species' biology and life history, the application of automated recognition tools can potentially provide new scientific insights that would otherwise be inaccessible to scientists. The quantity of images for individual whales presented a theoretically challenging problem, where the number of available labeled images was too large for further manual labeling, but not large enough to apply Deep Learning classification CNNs. This study demonstrated how the Deep Learning per-pixel segmentation FCN-8s [17] CNN could be trained for an individual minke whale recognition from only 179 positive images. As much as possible the off-the-shelf pre-trainedVGG16 [16] CNN was used to assist adoption and reproducibility of the results.

References

[1] Amam Hossain Bagdadee & Li Zhang,Renewable energy based self-healing scheme in smart grid,Energy Reports, Elsevier, DOI: 10.1016/j.egyr.2019.11.058

[2] Amam Hossain Bagdadee & Li Zhang,Electrical Power Crisis Solution by the Developing Renewable Energy Based Power Generation Expansion, Energy Reports, Elsevier, DOI:10.1016/j.egyr.2019.11.106

[3] Amam Hossain Bagdadee & Li Zhang, Power Quality Impact on the Industrial Sector: A Case Study of Bangladesh,Journal of Electrical Engineering & Technology ,Springer, DOI: 10.1007/s42835-019-00220-y

[4] Amam Hossain Bagdadee &Li Zhang,Smart Grid Implementation of the Industrial Sector : A Case of Economic Dispatch, International Journal of Energy Optimization and Engineering DOI: 10.4018/IJEOE.2019100101

[5] Amam Hossain Bagdadee & Li Zhang,A Review of the Smart Grid Concept for Electrical Power System, International Journal of Energy Optimization and Engineering, DOI:10.4018/IJEOE.2019100105

[6] AmamHossainBagdadee&LiZhang,Power Quality improvement provide Digital Economy by the Smart Grid, IOP Conference Series: Materials Science and Engineering, DOI: 10.1088/1757-899X/561/1/012097

[7] AmamHossainBagdadee, LiZhang&Modaway,Constant &Reliable Power Supply by the Smart Grid Technology in Modern Power System,IOP Conference Series : Materials Science and Engineering, DOI: 10.1088/1757-899X/561/1/012088

[8] Amam Hossain Bagdadee,Li Zhang& Remus, A Brief Review of the IoT-Based Energy Management System in the Smart Industry, Advances in Intelligent Systems and Computing, Springer, DOI: 10.1007/978-981-15-0199-9_38
[9] Amam Hossain Bagdadee,Md Zahirul Hoque, & Li Zhang, IoT Based Wireless Sensor Network for Power Quality Control in Smart Grid,Procedia Computer Science, Elsevier(DOI: 10.1016/j.procs.2020.03.417
[10] Amam Hossain Bagdadee, "Imitation intellect Techniques Implement for Improving PowerQuality in Supply Network Published in IEEE International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-DOI: 10.1109/SCOPES.2016.7955611 ,2016
[11] Porter-Roth, B. (2006). Applying Electronic Records Management in the Document Management Environment: Xerox DocuShare Business Unit 3400 Hillview Avenue Palo Alto, California 94304 USA (800) 735-774
[12] Schellenberg. (1956) in Elizabeth Shepherd and Geoffrey Yeo (2003). Managing Records. Handbook of principles and practices. London Facet Publishing Shepherd,
[13] E. and Yeo, G. (2003) Managing Records: A Handbook of Principles and Practice. Facet Publishing, London.
[14] Torton, A. (Ed) (1999). Managing business Archives. Butterworth: Heinemann Publishing.
[15] University of Adelaide. (2020). Records and Archives Management Handbook: The University Library. https://www.adelaide.edu.au/library/library-services/records-services/recordsand-archives-management-handbook/life-cycle-of-records
[16] University of Ghana. (2014). University of Ghana Records Management and Archives Policies. 52(3)
[17] University of Sheffield. (2020). Records Management Policy: Scope of the Policy. University. Secretary's Office. https://www.sheffield.ac.uk/uso/info-gov/records2/policy
[18] Visscher, A. J., Wild, P., & Fung, A. C. (2001). Information Technology in Educational. Management: Synthesis of Experience, Research and Future Perspectives on Computer assisted School Information Systems. The Netherlands: Kluwer Academic Publishers.
[19] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D.,